Dynamics at the Lexicon-Syntax Interface

Formelhafte Sprache
Formulaic Language

———

Edited by
Natalia Filatkina, Kathrin Steyer and Sören Stumpf

Volume 6

Dynamics at the Lexicon-Syntax Interface

Creativity and Routine in Word-Formation and
Multi-Word Expressions

Edited by
Sabine Arndt-Lappe and Natalia Filatkina

DE GRUYTER

Contents

C Dynamics of Creativity and Routine in Synchrony

Sabine Arndt-Lappe & Natalia Filatkina

Dynamics at the Lexicon-Syntax Interface: Creativity and Routine in Word-Formation and Multi-Word Expressions

Products of word-formation and multi-word expressions (phraseologisms, phrasemes, set phrases) are well-known to be characterised by both creativity and routine. Despite obvious parallels, however, word-formation and multi-word constructs are often studied independent of each other, by different research communities. Springing from a workshop with the same name that was held at the 45th Annual Conference of the German Linguistic Society at the University of Cologne (7.-10.3.2023), the present volume on *Dynamics at the Lexicon-Syntax Interface: Creativity and Routine in Word-Formation and Multi-Word Expressions* brings together different research communities to discuss empirical evidence on the role of creativity in word-formation and multi-word units.

Current frameworks accounting for word-formation and multi-word expressions are divided over the extent to which they see creativity as a central issue in their domains. The volume focusses on dynamics as a central source of such evidence, asking what role creativity plays in the emergence, usage and propagation of lexical patterns, and in the development of 'routine'. At the same time, the contributions discuss in what sense and to what extent 'routine' is a necessary prerequisite for creative use of word-formation and multi-word units. Whereas routine is generally (but not always) captured in terms of morphological (word-formation) and lexical/syntactic/pragmatic (multi-word expressions) fixedness, 'creativity' is defined in different ways in the field, with definitions ranging from seeing creativity at the heart of human (linguistic) cognition (e.g. Chomsky 1964, 1965; Goldberg 2003) to seeing creativity as precisely beyond 'regular' routines (e.g. Dobrovol'skij and Piirainen 2009; Langlotz 2006; Sialm andBurger 2007: 61–104; Filatkina 2018), for instance as an attention-seeking tool or a tool for achieving a pragmatic effect (Arndt-Lappe et al. 2018; Norrick 2007). In this latter sense, 'creativity' is often sharply distinguished from 'productivity', i.e. rule-governed, grammatical behaviour.

In research on multi-word units, creative strategies are usually discussed as 'modifications'. Diachronically, however, modifications may develop into productive patterns. In word-formation research, creative strategies are often labelled as 'extravagant' (Haspelmath 1999), 'extragrammatical' (Dressler 2000), 'analogy-based' (Mattiello 2017), or simply 'creative' (Benczes 2006); again, such patterns may diachronically develop into regular morphological processes ('debonding',

Norde and Van Goethem 2018). Despite these categorisations, the question on what basis patterns are to be defined as 'creative' still remains subject to debate. Criteria that have figured particularly prominently are (a) the degree of consciousness with which an expression was formed, (b) the expressive function of a pattern, and (c) structural properties of a pattern (like, e.g., non-concatenative properties). None of these criteria is without its problems, however, as all properties mentioned can also be true of processes that are generally not conceived of as 'creative'. The contributions in this volume put precisely these questions and approaches center stage and discuss them from the perspectives of word-formation and multi-word units research.

The structure of the volume is as follows. Section A starts with a series of programmatic papers discussing different approaches towards defining and operationalising creativity and routine (with contributions by Alexandra Bagasheva; Raymond W. Gibbs; Vsevolod Kapatsinski; Hans-Jörg Schmid). Sections B and C present studies of novel empirical evidence on the dynamics of creativity and routine; the focus of Section B is more on diachrony (with contributions by Fabian Fleissner; Regina Ruf & Elena Smirnova; Sören Stumpf; Monica Vasileanu & Anabella Niculescu-Gorpin), the focus of Section C is more on synchrony (with contributions by Maximilian Frankowsky & Barbara Schlücker; Stefan Hartmann & Tobias Ungerer; Pedro Ivorra Ordines & Carmen Mellado Blanco; Muriel Norde, Francesca Masini, Kristel Van Goethem & Daniel Ebner; Martin Schäfer). This structure is deliberately independent of the categorisation of phenomena as word-formation or multi-word constructs. To us both the discussions at the workshop and the reading of the manuscripts showed that, despite differences in terms of how pertinent phenomena might be contextualised in the respective research traditions, such categorisation is only of limited value for the issues that are in the focus of this volume.

References

Arndt-Lappe, Sabine, Angelika Braun, Claudine Moulin & Esme Winter-Froemel (eds.). 2018. *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and Ludicity*. Berlin & New York: de Gruyter Mouton.

Benczes, Réka. 2006. *Creative compounding in English* (Human Cognitive Processing). Vol. 19. Amsterdam & Philadelphia: John Benjamins Pub. Co. http://www.worldcat.org/oclc/937294327.

Chomsky, Noam. 1964. *Current Issues in Linguistic Theory* (Janua Linguarum. Series Minor 38). The Hague: Mouton. https://doi.org/10.1515/9783110867565.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA.: MIT Press.

Dobrovol'skij, Dmitrij Olegovič & Elisabeth Piirainen. 2009. *Zur Theorie der Phraseologie: kognitive und kulturelle Aspekte* (Stauffenburg Linguistik 49). Tübingen: Stauffenburg.

Dressler, Wolfgang U. 2000. Extragrammatical vs. marginal morphology. In Ursula Doleschal (ed.), *Extragrammatical and marginal morphology.[collection of papers presented at a Workshop on "Extragrammatical and Marginal Morphology" held in Vienna in February 1996]* (LINCOM Studies in Theoretical Linguistics 12), 1–10. München: LINCOM Europa.

Filatkina, Natalia. 2018. Expanding the lexicon through formulaic patterns: The emergence of formulaicity in language history and modern language use. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin & Esme Winter-Froemel (eds.), *Expanding the Lexicon*, 15–42. Berlin & Boston: De Gruyter. https://doi.org/10.1515/9783110501933-017.

Goldberg, Adele E. 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences* 7(5). 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068. https://doi.org/10.1515/ling.37.6.1043.

Langlotz, Andreas. 2006. *Idiomatic Creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English* (Human Cognitive Processing). Vol. 17. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/hcp.17.

Mattiello, Elisa. 2017. *Analogy in Word-formation*. Berlin & Boston: de Gruyter. https://doi.org/10.1515/9783110551419.

Norde, Muriel & Kristel Van Goethem. 2018. Debonding and clipping of prefixoids in Germanic: Constructionalization or constructional change? In Geert Booij (ed.), *The Construction of Words* (Studies in Morphology), vol. 4. Cham: Springer. http://www.worldcat.org/oclc/1034540596.

Norrick, Neal R. 2007. Set phrases and humor. In Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick (eds.), *Phraseologie / Phraseology: Ein internationales Handbuch der zeitgenössischen Forschung / An International Handbook of Contemporary Research*, 302–308. Berlin & New York: De Gruyter Mouton. https://doi.org/10.1515/9783110171013.

Sialm, Ambros & Harald Burger. 2007. Grundbegriffe. In Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick (eds.), *Phraseologie / Phraseology: Ein internationales Handbuch der zeitgenössischen Forschung / An International Handbook of Contemporary Research*, 61–104. Berlin & New York: De Gruyter Mouton. https://doi.org/10.1515/9783110171013.

A Theoretical Approaches to Creativity and Routine

Alexandra Bagasheva

Creativity and Routine in Word-Formation

Three Case Studies

Abstract: The venerable tradition of studying creativity (overviews in Jordanous and Keller 2016; Jones 2016, etc.) has led to the recognition that linguistic creativity is a multifactorial phenomenon whose parameters vary along the axes of numerous contributing factors and remains a concept difficult to operationalize for empirical investigation. Creative linguistic behaviors involve at least the following: heightened linguistic awareness, intentionality, deviance from established norms, attention grabbing properties, novelty, effectiveness, successful communication, etc. The chapter discusses three case studies which showcase the interplay of different contributing factors in different communicative settings, emphasizing the gradient spaces between individual, conscious, and effortful creations and societal, incidental bursts of creativity, etc., leading to the recognition of the complex ontology of routine and creativity as two feedback loops greasing the mechanisms of everyday language. The analysis reveals that the techniques of recombinance and rechunking are constant across various situations of creativity enhancement, highly sensitive to the immediate contextually variable values of co-semiosis in individual acts of communication. The degree of creativity results from the overlaying of various axes of cumulative effects of converging forces in individuals negotiating meaning in specific contexts. In view of language being an emergent complex adaptive system (Beckner et al. 2009) and cognition characterized as highly distributed (Hutchins 2010; Sharifian 2015, 2017), the inception of creativity is invariably achieved in micro events (interpersonal communicative acts) whose rippling consequences may become in time fully conventionalized, resulting in language change. The analyses of the case studies, each focusing on a different leading factor contributing to heightened creativity (conscious, purposeful and playful creations, interlingual translation, and analogy-based enhancement of the productivity of a process) corroborate the hypothesis that the nature of the word formation process is of the least significance for creativity (as reanalysis + deaffixation, affixation and blending can be equally creative), whereas context, which brings into the picture shared cognitive routines and markedness of the hypostatized concept, either in terms of novelty or unexpectedness, verging on incongruity, and formal experimentation (involving violation of highly routinized form-meaning mappings) play far more significant roles in creative word formation.

"Creativity is piercing the mundane to find the marvelous".
Bill Moyers

1 Introduction

The ecology and nature of (linguistic) creativity has been a definitional and analytical challenge ever since Croce's (1994 [1902]: 37) proclamation that "language is perpetual creation" and has generated a viable scholarly tradition. Definitions of linguistic creativity range from detailed and comprehensive ones (Munat 2010: 147–148; Zawada 2006: 236–237; for an overview of definitions see Jordanous and Keller 2016; for comprehensive accounts see Jones 2016; Kaufman and Stenberg 2010; among many others) through ones identifying it as a psychological factor in word formation (Körtvélyessy, Štekauer and Kačmár 2021: 1018–1019) to one-line adages such as "[Creativity] is seeing what others see and thinking what no one else ever thought"[1] (Schopenhauer 1851) or "blind variation and selective retention" (Simonton 2011: 158).

The question as to who or what is creativity a property of has also accrued opposing views, deducible from the modifiers used to indicate types thereof, such as linguistic creativity, artistic creativity, human creativity, individual (psychological) creativity, historical creativity (Boden 1990), etc. The multidimensionality, multifactoriality and indeterminacy of immediate causality (agency, contextual environment, degree of effectiveness, degree of intentionality, degree of novelty, etc.) have posed a number of debatable questions regarding creativity in and of language whose answers depend on the understanding of the nature and architecture of language as a semiotic system, the essence of language-mediated communication and the nature of meaning. The latter two inevitably lead to associating creativity with interoperationality between communicators. Just as beauty is in the eye of the beholder, item familiarity and lexical network awareness on the part of the perceiver have a strong influence on the identification and recognition of creativity in language and more specifically in word formation.

No less problematic is the relation between routine and creativity, with the former basically perceived as the background against which creative products are profiled. As Munat (2010: 147) reasons, picking up from Pope XIV, it is not clear whether creativity resides in "initiating things, continuing them or completing them". The nature of this relationship has been interpreted as modification, extension or violation of rules of productivity (or routine) (Bauer 2001). A further complication arises from the understanding of routine, which itself is not less di-

---

1 The adage has been attributed to numerous scholars and authors, among whom Albert Einstein (see https://quoteinvestigator.com/2015/07/04/seen/).

verse, from rule-based understanding of word formation as opposed to extra-grammatical creations (Dressler and Barbaresi 1994; Mattiello 2013), the opposition between word creation and word formation (Ronneberger-Sibold 2006, 2010, 2015), through exemplar-based schema abstraction (Audring 2019) or analogical extension to recognising it as a characteristic of the individual's established linguistic habits (Schmid 2020). Further contentious issues stem from the difficulty of teasing apart and quantifying (if possible) the multiple factors playing a role in recognizing, applying or analyzing creativity. The parameters suggested as minimally necessary and sufficient in the standard definition of creativity (Runco and Jaeger 2012: 92), "originality and effectiveness", pose challenges with their indeterminate, gradient nature. Originality has been identified as co-terminous with novelty or innovativeness (e.g., Bergs 2018; Hoffmann 2018), where novel necessarily dictates that "the creative product did not exist previously in precisely the same form . . . . The extent to which a work is novel depends on the extent to which it deviates from the traditional or the status quo" (Stein 1953: 311). Such definitions highlight the mutual dependency and inherent opposition of the notions of routine and creativity, which as the two extremes of a gradient scale enclose the space of innovation and productivity in word formation.

In the chapter qualitative commentary of three case studies is offered. The case studies can be grouped into ones showcasing individual creativity and those showcasing societal or linguistic creativity (i.e., one not directly contextually attributable to a single individual's mind). In each case, a different leading factor among the relevant ones takes the upper hand in underscoring creativity: conscious effort and heightened language awareness of the creator (author and translator) as the leading ones in two of the case studies, and lay people's extensions and modifications of linguistic routines in the other. The chapter is structured as follows: Section two overviews the parameters deemed essential for the detection and functioning of creativity in word formation; Section three showcases the three parametrically differentiated case studies, Section four provides general commentary based on the case studies and Section five concludes.

2 Creativity in Word Formation

In the middle of a traffic jam the tired and hungry driver says: I'm one oid. "What is oid?" asks the passenger and reflects, "Well, if you are one oid, then you are . . .?" Applying the rules of English grammar, the passenger completes with *an oid* – and the penny drops. This example fully conforms to the definition of creativity as "the ability to produce work that is both *novel* (i.e., original, unexpected) and *appropriate*

(i.e., useful, adaptive concerning task constraints" (Sternberg and Lubart 1999: 3). The form is unexpected and unfamiliar, yet the listener manages to comprehend the message, i.e., the effect is achieved. The mechanism employed here for achieving (and alerting to) linguistic creativity capitalizes on linguistic awareness by employing homonymy and intermixing levels of patterning (word formation and syntax) by substituting the conventionalized deverbal adjective used for naming the concept not with a different form of the same category (a complex, derived word) but by a different construction – a pseudo noun phrase.

More specifically, word-formation creativity "is conceived as the ability of any speaker of a language to approach the naming act in a creative way by selecting one out of a number of possible ways of semiotic representation of an object to be named" (Körtvélyessy, Štekauer and Kačmár 2021: 1017, citing Štekauer 2005). Although according to this definition creativity is at the heart of any naming act, i.e., in any instance of word formation, additional restrictions have been suggested for distinguishing routine word formation and creative word formation. The latter is characterized as "non-rule-governed and therefore intentional" (Bergs 2019: 175; see also Fernández-Domínguez 2010; Lieber 2010; Ronneberger-Sibold 2010). This invites the contrast between productivity, which is identified as rule governed and therefore automatic and unintentional (Körtvélyessy, Štekauer and Kačmár 2021), and creativity. The opposition is neither binary, nor privative. It represents the complex gradient space between F-creativity and E-creativity as defined by Sampson (2016: 19) as, "activities which characteristically produce examples drawn from a fixed and known (even if infinitely large) range as 'F-creative', and activities which characteristically produce examples that enlarge our understanding of the range of possible products of the activity as 'E-creative'". F-creativity is witnessed in the creation of numerous novel lexical items via analogical modeling following an exemplar as when "the suffix -er is used to create new words such as *mansplainer* "somebody who mansplains" (Bergs 2019: 175). Bergs (2019) directly maps F-creativity onto productivity but fails to define E-creativity. He evokes Haspelmath's (2002: 100) distinction between "(mostly unintentional, subconscious) productivity and intentional, creative neologisms that do not follow any major, productive pattern" (Bergs 2019: 175) as in "*va-jay-jay* "vagina"" (Bergs 2019: 175). This understanding significantly problematizes the correlation between creativity and routine in word formation. As the case studies presented below illustrate, all creative word formation products (be they intentional or not) employ the same mechanisms and formal manipulations, albeit the degree of intentionality varies. Despite the fact that routine and creativity have traditionally been conceptualized as opposites, they are not only definitionally tightly connected, but coexist in a dynamic ontology where the former is the background for the recognition of the latter, which is transformed into the former by conventionalization of the pattern

which has led to the initial fabrication of the creative product via schematic abstraction and societal propagation.

Against this theoretical background, the three case studies below provide corroborative evidence for the hypothesis that creativity cannot be ascribed individually to any of the factors mentioned above, let alone be associated with specific processes in word formation. It can only be measured in the interaction among all factors in specific contexts. Amongst this complexity of creativity, it appears that semantic incongruity (conceptual clash or contrast) or unexpectedness of the hypostatized concept is most likely to be recognized as creative in communication.

3 Three Case Studies

In Zawada's (2006: 236–237) definition "[l]inguistic creativity is primarily the activity of making new meaning by a speaker (in the broadest sense of the user of language in all forms and in all mediums), and the recreation and re-interpretation of meaning(s) by a receiver. Linguistic creativity is secondarily observable as a feature or product in a language." This suggests that cognitive / conceptual flexibility (the intended new meaning) is, if not a causal factor for word formation creativity, at least a very strong correlate. When the infinity of conceptual combinability is captured by a recognizable and hypostatizing linguistic form word formation creativity ensues. Each case study underscores the idea that despite the heightened significance of diverse factors, unexpectedness of a newly hypostatized concept or conceptual integration of incompatibles seems to be an indispensable ingredient of creativity. In the first two case studies the pragmatic dimension of the communicative act and the contextual factors are the common denominator. Conscious creative effort on the part of the producer and the pragmatically informed expectations on the part of the comprehender are leading factors in the first. With slight modifications, in the second, again the highly elaborate efforts on the part of the producer (due to the inherent constraints imposed by the situation of interlingual translation), as well as the comprehender's expectations, paly a leading role. The third stands apart in terms of pragmatic and contextual supporting factors, as it focuses on unrestrained spontaneous creative productions in unspecifiable immediate contexts, where the significance of the nature of the hypostatized concept stands out as the generator of creativity.

3.1 Heightened Effort on the Part of the Creator

The first case study focuses on a recognizably conscious effort on the part of the creator as its context is a vocabulary competition. The poem appeared in the Opinion section of the New York Times under Schott's Vocab op-ed portion on 19 April 2010. "Schott's Vocab is a repository of unconsidered lexicographical trifles – some serious, others frivolous, some neologized, others newly newsworthy. Each day, Schott's Vocab explores news sites around the world to find words and phrases that encapsulate the times in which we live or shed light on a story of note" (*New York Times*, 19.04.2010). This particular poem appeared in *Schott's Vocab. A miscellany of modern words and phrases* amidst a weekly competition for examples of and definitions of words among co-vocabularists. The actual author of the poem is not known but Schott ascribes it "to one J. H. Parker". Ben Schott quotes the poem as tribute to the winners of the weekly vocabulary competition. The original context of the composition and its initial appearance in front of the public eye remain unknown. The use of the poem as a tribute to the efforts of voluntary contributors to a vocabulary competition suggests that the poem is considered a worthy achievement of creativity. The pragmatics of praising the creative achievement of the unknown author determines the readers' expectations of something out of the ordinary and noteworthy, i.e. of something with heightened creativity.

 A Very Descript Man, J. H. Parker

I am such a dolent man,	*I travel cognito and make*
I eptly work each day;	*A delible impression:*
My acts are all becilic,	*I overcome a slight chalance,*
I've just ane things to say.	*With gruntled self-possession.*
My nerves are strung, my hair is kempt,	*My dignation would be great*
I'm gusting and I'm span:	*if I should digent be:*
I look with dain on everyone	*I trust my vagance will bring*
And am a pudent man.	*An astrous life for me.*

The tongue-in-cheek poem is most revealing about the influence of three specific factors for the recognition of creativity of word formation products: i) intentional deviation from established routine – reanalysis of synchronically unanalyzable, i.e. considered simplex lexemes, by the lay speaker of the language who lacks detailed diachronic knowledge, followed by removal of presumed prefixes (some of which are diachronically actual prefixes), ii) conscious or purposeful creation on the part of the producer, who demonstrates heightened linguistic awareness and purposeful form manipulation and iii) (genre-specific) expectations on the part of

the comprehender for heightened conscious creativity on the part of the writer due to the nature of the communicative act (consuming poetry).

The poem itself reads as purposeful showcasing of deviations from established affixation routines in English. It experiments quite intentionally with the plasticity of word formation as a cognitive-linguistic routine demonstrating a distorted pattern of deaffixation[2] (Renner 2020) or if we are to use traditional labels it would come closest to the process of backformation, with which it shares movement of lexeme internal boundaries. The mechanics involve reanalysis in its simplest form, i.e., re-segmentation (Burridge and Bergs 2017: 107–109; Gaeta 2010; Haspelmath 1995, among others) and removal of an affix (which is synchronically not easily identifiable without specialized linguistic knowledge of the diachronic development of separate lexemes such as *disdain*, *disgruntled*, etc.). Despite the ease with which an analyst would recognize the process, the removal of elements does not exactly follow the routine affixal combinations in English, since many of the resultant elements can hardly be recognized as bases (or) roots with the potential to realize free standing lexemes. Even though *dolent* can be recognized as *sorrowful* (albeit a recognition dependent on the linguistic acumen of the reader as the word is identified as archaic), neither *eptly*, nor *pudent* can be recognized as legitimate lexical items or roots in English. The author quite systematically employs the novelty and unexpectedness dimension of creativity to invite the comprehender to put some extra effort into understanding both the message and the employed linguistic technique. By removing elements from quite familiar words, where some of the manipulated elements are synchronically homophonous with prototypical prefixes (e.g., *dis-*, *in-*, *non-*, *im-*), while others are quite arbitrary segments (e.g., *-mund* or *-ane*), the author pliably extends the rules of English affixation. Truncation surfaces as an across-the-board routinized process used by the author to sensitize readers of the poem to the malleability of words and the limitless potential for novel word creation by form manipulation. The human propensity for pattern completion, which is activated in decoding/retrieval (Hunsaker and Kesner 2013), and the Gestalt principle of the whole taking precedence over constituent parts (Wangemans et al. 2012) are immediately activated to substantiate the form manipulation with a parallel manipulation of the semantics, thus maintaining the biuniqueness of the linguistic sign, understood as form-meaning isomorphism (Ungerer 1999: 309). The conceptual recalibration is straightforward since the removal of what resembles a negative marker results in a positively

---

2 Renner (2020: 5) uses the term "desuffixation" for "the deletion of a suffix or pseudo-suffix to bring about a change of lexical category". By extension here the term is transformed to deaffixation to include prefixation and to encompass meaning change with a preservation of lexical category.

marked concept. The fully intentional conceptual creativity of the author spans over diverse cognitive/semantic domains but utilizes a single technique – deaffixation. The semantic transparency and consistency – creating a positively marked concept from a presumed negative one, maintains an optimal balance between economy and transparency, with resultant iconicity of reducing form in parallel with deleting a semantic feature. The removed elements (some of which happen to be homonymous with prefixes or are diachronically such) are all treated as if they were synchronically functioning prefixes denoting the semantic feature "negation", a procedure supported by the typically prefixally composed negative forms of adjectives in English. The routinized knowledge of English speakers that the addition of the prefix *im-* to *polite* leads to the negation of the quality, i.e. the semantic feature "lack of", guides the reader of the poem into computing *pudent* and *becilic* as adjectives affirming the possession of a property and helps them interpreter the intended meaning. The paradigmatic semantic relation of oppositeness of meaning (i.e., antonymy/contrast, Murphy 2010: 117–123) eases the conceptual manipulation and helps the comprehender appreciate or at least acknowledge the creativity of the author.

3.2 Constrained Creativity

In the second case study, the intentionality and conscious efforts of the creator are constrained by the task at hand – rendering an English text into Bulgarian. Fully aware of the complexity of literary translation and the impossibility to comprehensively analyze the creativity of translating Shakespeare in any language, the focus is exclusively on the "thumping neological creations" (defined as such by professional translators and commentators of the translations, e.g., Pancheva 2013, 2014) found in the latest rendition of the four great tragedies (*Othello*, *Macbeth*, *Hamlet* and *King Lear*) in Bulgarian by Alexander Shurbanov (2012a). Not downplaying the individual cognitive and linguistic creativity of both the original author and the translator by restricting the data chosen for analysis, the parameters employed for the sampling are: lack of the formations in previous translations of the same text and their analytical singling out as "unfamiliar", "strange", "creative neologisms", "novel stumbling creations" (Pancheva 2013, 2014; Shurbanov 2012b) by literary critics, linguists and general readership. The choice of this narrow perspective on interlingual translation is congruent with its classical definition "as an interpretation of *verbal signs* by means of some other language" (Jakobson 2004: 139 [emphasis mine]) and with the definition of words as "affordances: they afford opportunities for individuals to experience the meaning of things and situations and events" (de Oliveira and de Souza Bittencourt 2008: 25).

Remaining again in the realm of individual (linguistic) creativity (of Shakespeare and the translator), I cherry-picked from the identified neologisms in Bulgarian a sample of the word formation products resulting from the most frequently employed process[3] in these renditions of the tragedies in Bulgarian – (re)affixation[4]. The creativity of the individual as a cognizing agent in translation is multiply constrained (by the original text, the socio-historical and cultural context of the translation process, etc.) in comparison to the freedom of poetic creations. Yet, the translator has at his disposal the possibility of highly selective choices in terms of interpretation of the original text and harnessing the expressive potential of the target linguistic system. Despite countless possible choice for the translator, words remain the smallest salient verbal signs for the readers of the translation. This salience is what Veale implicitly highlights (2012: 87 [emphasis mine]) in defining creative language as "a form of conceptual rewiring that allows us to influence, *with words alone*, how ideas are connected and emotions are channeled in the heads of an audience".

It can be hypothesized that the individual creativity of the translator regarding the utilization of linguistic resources in the receptor language is employed to match the conceptual (and/or) formal creativity of the source message and the creative efforts of the original author. As these complex correlations can hardly be operationalized for analysis, the choices of the translator of viable non-routinized expressive means in the target language in an attempt to achieve results truthful to the original's effects are recognized as tokens of creativity. Shakespeare is notorious for his neological creativity (Garner 1987; Lederer 1991) and to reflect this and to do justice to the pragmatic, aesthetic and linguistic characteristics of the original, the translator has adopted the strategy of deviating from firmly established routines in the target language, relying heavily on analogy, combined with paradigmatic relations (all instances of affixing non-extant bases as in *захая* [zahaja, 'start caring'], analysed below) and contextual framing (using an associated base or providing semantically suggestive cues in the surrounding co-text as in the case of *отстрекава* [otstrekava, 'dissuade').

---

3 The texts abound in neological compounds but the complexity of their analysis far exceeds the scope of the current chapter. The Bulgarian texts are replete with neological verb prefixation, creative name renditions, etc., but affix manipulation is the most frequently employed process. It is quite natural for affixation to be the most frequently employed process as Bulgarian is characterized as a predominantly affixing language (Avramova and Baltova 2016). In the four plays 43 examples of affixally encoded creativity have been identified against 11 instances of creative compounding (Bagasheva and Stamenov 2015).

4 The term is intended to capture both the reanalysis of a monomorphemic word as a derivative and the individuation of a presumed affix and the use of existing affixes on atypical bases or non-extant bases.

According to Roberts et al. (2021: 121) a "generative process with a special link to creativity that has undergone careful experimental examination is analogical reasoning", defined as "the reorganization of existing category knowledge to form ad-hoc or goal-derived categories to meet a particular need". The role of analogy (e.g., Blevins and Blevins 2009; Gaeta 2010; Fertig 2020; Fischer 2019; Itkonen 2005), more specifically four-part or proportional analogy, has recently been recognized also as an extremely powerful and almost ubiquitous process in language, including synchronic word formation (see Arndt-Lappe 2015 for an overview). Recognizing the pattern in a specific derivational model between the base and the derivative and applying it to establish the same pattern between a new base and a possible derivative or the reverse (a lexical item from which a possible base and a possible affix are produced as in the poem above), constitutes an act of creative interpretation. Such is the case in which the translator conscious of the systematic morphosemantic (directional) opposition between the inchoative/inceptive or resultative *за-* [za-, 'begin doing'] prefix and the privative, reversive prefix *от-* [ot-, 'reverse the result of', 'accomplish'] in Bulgarian in such pairs as *запиша* [zapisha, PREF$_{accomplishment}$-write, 'enroll'] and *отпиша* [otpisha, PREF$_{reversive}$-write, 'to delist, to drop out'], *заключа* [zaklyuča, PREF$_{accomplishment}$-lock, 'to lock'] and *отключа* [otklyuča, PREF$_{reversive}$-lock, 'to unlock'] relies that the perceived relational similarity will be strong enough to get activated in the pair *сторила*[5] [storila, Ø-do, 'to do'] – *отсторила* [otstorila, PREF$_{reversive}$-do, 'to undo'] (the original is "undo't"). The reversive meaning maintained in the opposition between *за-* [za-] and *от-* [ot-] is assumed to work even in the case of Ø and *от-* [ot-]. This is possible due to the lexical semantics of the base which encodes accomplishment whose reversal is a natural conceptual correlate. This intralingual analogy in prefix use is heavily exploited, for example in using the conventionalized monomorphemic verb *подстрекава* [podstrekava, 'to incite, to instigate'] as a source of reanalysis, removal of a presumed prefix *под-* [pod-, 'completion' / 'under'] and creating the novel reversive antonymic verb *отстрекава* [ot- + strekava$_{presumed\ base,\ resulting\ from\ deprefixation}$, 'to dissuade'] (in the original this is "unprovokes") by presumably substituting a prefix, which does not actually figure in the base. The translator freely uses this strategy, since analogy correlates with a textual function of word formation defined by Bauer (2000: 836) as "the tendency for new words to occur in close textual proximity either to their

---

5 The words analyzed are given in the form in which they appear in the translation but are glossed in the respective base form. All the examples are taken from the 2012 translation of the four tragedies by Alexander Shurbanov. In the paper the exact appearance of the examples is not indicated. A table with detailed information for all the excerpted examples, where both the play and the page are identified, can be provided upon request.

base words or to another derivative from the same base. This phenomenon most clearly helps the listener to interpret the new word, but it probably also helps the speaker to coin it". Thus the appearance of *подстрекава* [podstrekava, 'to incite, to instigate'] (synchronically unanalyzable) naturally facilitates the recognition and acceptance of the new coinage *отстрекава* [otstrekava, 'to dissuade'], which in the absence of the former might be difficult to interpret. The privative and reversive meaning of the prefix *от-* [ot-], substituting the presumed prefix *под-* [pod-] helps the perceiver to immediately recognize the meaning (which is also exploited in *откадят* [otkadyat, 'disperse smoke', PREF$_{causative, \, reversive}$ – ot-smoke] ("fire" in the original), *отродена* [otrodena, 'deprived of relatives'] ("unfriended" in the original), etc.).

A case similar to *подстрекава/отстрекава* [podstrekava/otstrekava] is *захая* [zahaja, 'start caring']. *Нехая* [nehaja, 'do not care'] is not analyzable synchronically and the negative particle *не-* [ne, 'not'] is spelled together with the verb violating the general rule for verb negation in the language. *Хая* is hardly more immediately understandable than would be *becile*, unless as an antonym of *imbecile.* Substituting the presumed negative prefix *не-* [ne-] with the inchoative prefix *за-* [za-] results in a neologism, which can only be understood by contemporary readers with reference to *нехая* [nehaja, 'do not care']. The reader/listener is again facilitated by the appearance of the new coinage *захая* [*zahaja*] (in the original "would make thee care for me") in the immediate verbal context of the well-established and familiar *нехая* [*nehaja*]. No matter whether such creations are embedded in a suggestive context or not, they will necessarily attract attention and make the cognitive prosody of the text choppy and have special effects on the readers.

The neologism *начелник* [načelnik, lit. PREF$_{over}$ – forehead-SUFF$_{entity}$, 'visor'] ("visor" in the original) enters a paradigm-driven analogy. The productive model (or a series of exemplars) on which it is based can be traced in *нагръдник* [nagradnik, lit. PREF$_{over}$ – chest-SUF$_{entity}$, 'breastplate'], *нашийник* [nashiynik, PREF$_{over}$-neck-SUFF$_{entity}$, 'neckpiece'], *намордник* [namordnik, PREF$_{over}$-muzzle-SUFF$_{entity}$, 'muzzle']. Commenting on the decisions of a previous translator, Shurbanov (2004: 52–53) points out as a determining factor in such choices "the recognizability" of the components in the lexical choice and praises one of the previous translators for his avoidance of *забрало* [zabralo, 'visor'], an old Slavic word for *visor*, and his resorting to a creative neologism *лицебран* [licebran, 'face-guard'] (a compound formation), whose second component is easily recognizable and renders the meaning of the coinage transparent and includes the new word in a series of literary and obsolete words with the same positional constituent. Shurbanov's choice, instead, relies for its effectiveness on the series of semantically related lexical items for different types of body part covers and may be analyzed as a suffixal quasi

phrasal compound based on the prepositional phrase *на чело* [na chelo, 'on the forehead'] + the suffix *-ник* [-nik] or alternatively as the result of circumfixation.

The list of noteworthy nonce-formations excerpted from the latest translation of the four great tragedies can be extended ad infinitum. The cumulation will be exclusively quantitative without any qualitative change, since the same word formation processes and mechanisms are employed. The comprehenders perceive them as creative since they involve deviations from or modifications of established routines.

3.3 The Creativity of Blends

The creativity of the products that are in general circulation and whose creator is unknown originate initially in the same manner as ones recognized for the creativity of the minds that created them, with the difference that in societal creativity the initial moment and immediate context of the creation are unknown. The inability to assign authorship immediately transfers creativity properties directly to the product.

Consequently, from the product's perspective in the complexity of the creative word formation act, linguists have ventured to assess the creativity of different word formation processes and blending has surfaced as the most ludic and comprehension-challenging (Hamans 2011; Kjellander 2018, 2019; Renner 2015; etc.) or the most creative one (Kjellander 2018; Gries 2012). The reasons for this high creativity ranking of blending can be read off some of the crucial features of blending as identified by Gries (2012: 145, [emphasis mine]): "not as rule-governed as other derivational processes; *more creative* than most derivational processes; usually involves *conscious effort* and wordplay, sometimes *violating* rigid morphological rules". The mere status of the process as "a phenomenon of word creativity, or [. . .] a regular and predictable mechanism of word formation, remains an open question" (Beliaeva 2019: 1). For the purposes of the current argument, blends are assumed to be lexical items arising from the non-morphemically motivated gluing of two (or more) base words into a single lexeme, with a unified meaning, resulting from conceptual integration, which is assumed to iconically mirror the structurally varied integration of the constituent elements participating in the blend (Beliaeva 2014; Kjellander 2018, 2019; Renner 2015; Ronneberger-Sibold 2010; etc).

One of the morphological rules that splinters (or blend constituents) violate is their non-discriminative behaviour in relation to the various kinds of elements they attach to: other splinters (e.g., *robogasm*), roots (e.g., *mindgasm*) or with inde-

terminate first element (e.g., *floorgasm < floor+-gasm / flo- + orgasm*). In other words, the secretion of a novel splinter, its utilization as a blending component and potential transformation to an affix-like element may be conceived as tracing the cline from word creation to full routinization of a word formation pattern. The more affix-like a splinter's behavior becomes, the less creative its products are considered (e.g., Bauer et al. 2019; Beliaeva 2019). This reduction of creativity, triggered by routinization, relates back to the wearing off of novelty, recognized as a property of creativity (see above).

3.3.1 On the Conceptual Creativity of *-gasms*

3.3.1.1 Theoretical Framework and Methodology

Evans (2016) contends that humans have at their disposal two semantic representational systems (the conceptual and the linguistic), which are characterized by qualitative differences, even though they constitute a co-evolved symbolic assembly for meaning construction in language. The two systems have different representational formats, and only the linguistic one can be interpersonally exchanged as it has a material, expressible side (Evans 2016: 1). Language is assumed to provide "an "executive" control function, operating over embodied concepts in the conceptual system. The essence of this function is to facilitate "access to knowledge representation – concepts – in the conceptual system, in order to construct meaning, during the course of communication" (Evans 2016: 2). Concepts in the conceptual system are directly grounded in the modalities of the perceptual, body-based experiences internally represented for an immersed experiencer (Barsalou 2008; Gallese and Lakoff 2005; Zwaan 2004; etc.). The dynamic view of concepts in the conceptual system is best summarized in the claim that "the concept is the skill or ability to produce a wide variety of situated conceptualizations that support goal achievement in specific contexts" (Barsalou 2005: 626). It is this flexibility of the conceptual system that underscores creativity in blending as a linguistic mechanism. For Evans (2016) the general-purpose conceptual system is analog in nature, while the linguistic conceptual system utilizes parameterized concepts. Parametric concepts provide "a level of schematic representation" which guides "how analog concepts are activated and, consequently, how simulations are constructed in the service of linguistically mediated meaning construction" (Evans 2016: 7). They provide access to large knowledge structures, streamlined by the specific linguistic encoding. Lexical concepts are associated with lexical items and function as alternative schematic evocative attentional cues for perceptual states encoded in memory. This coheres with the accounts of lexical semantics of both Fillmore (2006) and Langacker (2008), who insist that word

meaning is always relativized against larger knowledge structures, among which frames are central. Barsalou and Hale (1993: 131) contend, "[h]uman knowledge appears to be frames all the way down." Fillmore (2006: 378) defines the relation between frames, construal mechanisms and lexical items as a mutually implicating one: "[frame is] the structured way in which the scene is presented or remembered, we can say that the frame structures the word-meanings, and that the word 'evokes' the frame." Besides being central for the emancipation of lexical concepts, frames appear to be of relevance within word formation as well. This specific relevance is onomasiological in nature, i.e., it aids the human conceptualization capacity in forming parameterized linguistic concepts. As Koch (1999: 153, [emphasis mine]) recapitulates,

> [f]rames, which are relevant not only to metonymies but also to certain types of *word formation*, can – and in fact, should – be defined *onomasiologically*, so that even cross-over links within one and the same frame realized in different languages, concepts which have not yet been expressed, senses of a given word which do not yet exist, and new words which have not yet been fanned can all be provided for.

This provides an operationalized methodology for studying meaning-form correlations from a word-formational perspective. Adopting the definition of lexical meaning provided by Evans (2016) and the methodology of onomasiologically informed frame semantic analysis (Fillmore 2006), below I provide the third case study of word formation creativity.

3.3.1.2 The Nature and Functional Specialization of Blending

Blends are produced by gluing initially indeterminate constituents, employing the formal mechanism of composition. This allows users to produce complex concepts without the intervention of other parameterized linguistic elements. Blending provides the tools for maximal informativity with minimal linguistic parametrization. As a linguistic mechanism blending provides the formal side of "multistable meaning structures" (Kjellander 2019: 1) whose meaning computation requires cognitive creativity against formal similarity. Besides, blending specializes for the performance of two super-functions within word formation: the transconceptual and the compacting function. The latter two are defined within the "ecosystem view of English word-formation" (Renner 2020: 4) as two of the four basic lexical functions – transcategorial, transconceptual, evaluative, and compacting, which word formation processes in English specialize for. The transconceptual function is defined as "the function of changing the conceptual (or denotative) meaning of an input" (Renner 2020: 7) and is likened to Plag's (2018) "lexical-expanding" or "labeling" function (Renner 2020: 7). The transconceptual function accounts for modifications of concepts in terms of subcategorization, complex concept formation and conceptual in-

tegration. Renner (2020: 8) expounds that "compounding and blending encode the transconceptual function." Meanwhile the compacting function is defined as shortening "a preexisting complex lexical unit without any alteration of its meaning" (ibid.: 8). In short, as a mode of maintaining "multistable meaning structures", blending allows for parallel manipulation of the formal and meaning sides of an isomorphically constituted lexical items and provide space for heightened creativity. As Kemmer (2003: 71) claims, blends are "words that are cognitively linked to pre-existing words which are co-activated when the blend is used". This complexity of blends, associated with reduced linguistic parametrization and inherent multistable meaning construal influence the perception of blending as a highly creative word formation process requiring cognitive efforts for interpreting its products with even standardly glued elements.

3.3.2 The *-gasm* Family

The *-gasm* family has been rapidly expanding for the last twenty or so years (Barrena Jurado 2019 and Beliaeva 2019). Its cumulating members (attesting to type productivity (Beliaeva 2019), which within the F- vs. E-creativity space may be identified as growing routinization) raise interesting questions about the meaning contribution of supposedly the same splinter-constituent *-gasm*, which easily participates in such products as *scorgasm*, *eargasm*, *foodgasm*, and *chillgasm*. The creativity dimension in operation can safely be assumed to be exclusively conceptual since three formal parameters are identical: same process – blending, identical second constituent – *-gasm*. Most of the products are still considered neologisms (Bauer et al. 2019). The family can be conceived of as an intermediate constructional schema (on the types of schemas in construction morphology see Audring 2019; Booij 2007, 2010a, 2010b, 2016, 2017; Croft 2001; etc.) with a fixed second constituent:

[Xi gasm$_j$]Nk ↔ [SEMfeeling of pleasure or satisfaction$_j$ with relation R to SEMXi]Nk

Such schemas are intermediate not only in relation to levels of abstraction and item specification, but also as occupying an intermediate position along the gradient scale between creativity and routinization. The schema has the power to coerce an interpretation of new *-gasm* types roughly as 'pleasurable experience/ enthusiasm in relation to X', but this coercion involves specific conceptual combinations, whose role is not insignificant. It is exactly in this combinatorial calculation that the creativity of *-gasm* blends flourishes. Considering that "similarity and recognisability of the source words" are constitutive properties of blends (Gries 2012 and Beliaeva 2014, 2016, 2019), it takes some cognitive effort to recognize the similarity

of and combine concepts such as 'shoe' and 'orgasm' or 'rage' and 'orgasm'. It is not a trivial matter for both a creator and a comprehender to fathom out how *score,* for example, combines with *-gasm* in parallel to *ear* + *-gasm* (and *aurgasm*), *nerd* + *-gasm*, and *chill* + *-gasm*. A facilitating factor is the high anthropocentrically motivated recognizability of the source word *orgasm* and the remnant subversive inuendo associated with its history as a taboo word (on the essence and functioning of "the organs and acts of sex" as taboo words and their fate in English see Allan and Burridge 2006). Experimental data (Arndt-Lappe and Plag 2013 and Beliaeva 2016) have revealed that recognizability of the source words is a strong psycholinguistically relevant constraint for blending. In the majority of *-gasm* blends both source words are easily recognizable (especially cases in which *-gasm* attaches to roots, e.g., *foodgasm, eargasm, winegasm*, etc.). Unlike the recognizability condition, the similarity condition is far more difficult to satisfy (or analytically detect). Its violation is responsible for the perception of creativity in relation to novel *-gasm* products, which arises from the unexpectedness of the concepts combined with the recognizable splinter and depends on the plausibility of conceptually associating the components. No straightforward dimension of conceptual similarity can be established between *cry* and *orgasm, program* and *orgasm*. To be more precise, the conceptual computation of blends is based on the cognitive constraint identified by Kjellander (2018: 163–164) as ease of schema transfer. Schema transfer belongs to one of the two proto models describing conceptual combinations (Medin and Rips 2005; Murphy 2002; inter alia). Psycholinguists have spilled much ink in trying to establish the cognitive mechanisms behind conceptual combination and their efforts at present seem to gravitate towards two core models: schema-based models and relational models (Medin and Rips 2005; Estes et al. 2011). In schema-based models a conceptual schema with well-defined slots (properties or participants in case of event schemas) is established by one of the constituents (in compounding) and gets filled by relevant properties or participants from the schema of the second constituent (e.g., Murphy 1988; Wisnewski 1997). Relational theories on the other hand (e.g., Gagné and Shoben 1997; Gangé and Spalding 2006, 2015) emphasize the establishment of a plausible relation in which the constituents act as arguments. Or as Jones and Golonka (2012: 2) claim, the thematically related concepts play complementary roles in a given action or event and are amenable to a "script" interpretation. A thematic analysis of a compound will interpret *flu virus* as a virus causing flu (Medin and Rips 2005: 51), while a schema based one will interpret *truck soap* as soap for cleaning trucks (Medin and Rips 2005: 51).

Understanding the meaning of blends requires a plausible cognitive scenario of concept combination. Collapsing the two overarching models of concept combination in the psycholinguistic literature, Gagné (2000: 384) claims that both "relation interpretations and property interpretations" involve a process whereby "the

selection of a relation is followed by an elaboration process in which the properties/features of the newly formed combinations are derived". This is exactly the type of conceptual elaboration that according to Fillmore a frame coerces and aids in computing meaning. Frames (Fillmore 2006) as concept-language associated knowledge structures and as patterns of packaging cognition and the general cognitive principle of indexicality (cause-effect relations) (Ciecierski 2021) seem to be the most facilitatory ones in conceptual combinations in context. Kjellander (2019: 3) adds to these ambiguity as a major cognitive factor, where ambiguity is understood "as lexical structures allowing, or even driving, meaning construal involving unresolved semantic conflicts". Such semantic conflicts may be identified as mismatches whose resolution is achieved through coercion (Ziegeler 2007: 992). The necessity to identify and apply appropriate cognitive mechanisms to resolve the mismatch for successful communication to occur lies at the heart of creative linguistic processes (Bergs 2019). The creativity in the -*gasm* family is associated with the unspecified fusion of arbitrary parts (from whole words to chance splinters) of source words with a semantically more specific element, where the fusion coerces a mixture that has to be conceptually computed with creative effort, unlike in other blends, which "'mix' random parts of existing lexemes ('splinters') – structurally and semantically – and there is the additional semantic component blending/mixture" Fandrych (2008: 111), whose hybridity is iconically matched with the formal mixture (e.g., *labradoodle*). What renders most -*gasm* blends creative against their formal uniformity is the complexity of the *floating*[6] concept resulting from the cognitive efforts for "*multistable meaning construal*" (Kjellander 2019: 22 [emphasis in the original]).

The positionally fixed constituent (the splinter -*gasm*) is reported to derive from the source word *orgasm* (Beliaeva 2019: 10) which was borrowed into English in the "17th century from New Latin *orgasmus*, from Greek *orgasmos*, from *organ* – 'to mature, swell'" (Collins Dictionary). The Online Etymological Dictionary describes a semantic differentiation and specialization of the borrowed word, "1680s, 'sexual climax, the acme of venereal excitement', from French *orgasme* or Modern Latin orgasmus, from Greek orgasmos – 'excitement, swelling', from *organ* – 'be in heat, become ripe for', literally 'to swell, be excited', related

---

6 A floating concept is a rich and dynamic cognitive structure, which arises when "a new concept is forged out of the conceptual substance of two base concepts and from additional conceptual input. Especially with prototypical compounds whose attributes do not suggest an obvious conceptual link with the other concept, this is a difficult job" (Ungerer 2003: 564). Although the definition is provided for N + N compounds, the mechanism of composition described applies with double force to blends, where truncation of a source word further complicates conceptual computation.

to *orge* – 'impulse, excitement, anger', from PIE root *wrog- 'to burgeon, swell with strength'. Also used in 17th century of other violent excitements of emotion or other bodily functions; broader sense of 'immoderate excitement or action' is from 1763" (Etymonline).

This semantic differentiation and specialization in the word from which the splinter has been secreted is reflected in the members of the *-gasm* family in the 21st c., which can analytically be grouped into two overarching subschemas – pleasure in any sense-related realm (e.g., *eargasm*, *foodgasm*, *eyegasm*, and metonymic extensions thereof (of *foodgasm*) *beergasm*, *fruitgasm*, etc.) and "excitement, enthusiasm about something" (e.g., *bookgasm*, *postergasm*, *vetogasm*, *neologasm*, etc.). The subschemas can for analytical purposes be modeled via the method of frame semantics, which necessarily involves the study of the multidirectional backgrounding/foregrounding relations between concepts and the lexical items evoking and evoked by them. A frame for the purposes of semantic analysis is understood as a "system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text, or into a conversation, all of the others are automatically made available" (Fillmore 2006: 373). Frames constitute the gestalts against which the multistable semantic relations within a complex word are established.

The frame semantic analysis of 173[7] *gasms* confirms that there are two basic subschemas: a) 'physical or sensual pleasure' and b) 'excitement or enthusiasm about'. The subschemas differ in terms of the nature of the index *j* on the specified X constituent in the intermediate construction schema and the nature of *R*. The two subschemas within the *-gasm* family can be summarized in the subschemas:

[Xi gasm$_i$]Nk ↔ [SEMgasm$_i$ with *R physical or sensual pleasure* to SEMXi]Nk

[Xi gasm$_j$]Nk ↔ [SEMgasm$_j$ with *R excitement or enthusiasm R* to SEMXi]Nk

These two subschemas are associated with different frames, and these motivate the interpretations in the combinations of blend constituents of *-gasm* formations. The respective frame coerces the most plausible or cognitively least demanding interpretation that fits the Xj constituent into the relevant frame. The FrameNet (https://framenet.icsi.berkeley.edu/) provides the following descriptions of the respective frames:

PLEASURE

---

7 The data (173 *-gasm* blends) have been harvested from Barrena Jurado (2019), Beliaeva (2019), COCA, GloWbE, enTenTen, and Urban Dictionary.

The words in this frame describe an *Experiencer*'s emotions with respect to some *Content*. Although the *Content* may refer to an actual, current state of affairs, quite often it refers to a general situation which causes the emotion.

Content is what the *Experiencer*'s feelings or experiences are directed towards or based upon. The *Content* differs from a *Stimulus* because the *Content* is not construed as being directly responsible for causing the emotion.

The *Topic* is the area about which the Experiencer has the particular experience.

The Frame Element *Expressor* marks expressions that indicate a body part, gesture or other expression of the Experiencer that reflect his or her emotional state.

And of EXCITEMENT/ENTHUSIASM, respectively

an *Experiencer* who is feeling or experiencing a particular emotional response to a *Stimulus* or about a *Topic*. There can also be a *Circumstances* under which the response occurs or a *Reason* that the *Stimulus* evokes the particular response in the *Experiencer*.

The *Reason* is the explanation for why the *Stimulus* evokes a certain emotional response.

The *Stimulus* is the person, event, or state of affairs that evokes the emotional response in the *Experiencer*.

The differences between the two frames are captured in *Expressor* specialised for the Pleasure frame and *Reason* for the Enthusiasm frame. From this ensues a difference concerning conceptual combinations: the first subschema does not require extensive cognitive efforts, since it is easy to calibrate conceptually how a pleasurable feeling is associated with the *Stimulus*, *Expressor* or *Topic*. *Experiencer* is a bit more complicated but does not pose a real "intellectual challenge" in Kajellander's sense (2019: 22). The *Experiencer* in the Pleasure frame requires more extensive frame fusion as it is not immediately obvious which specific property of the *Experiencer* is elaborated. Consider *beergasm*[8], *chocogasm*, *eargasm* and *mouthgasm*. In the first two *Stimulus* is easily detectable, and people effortlessly recognize the meaning of the blends as 'pleasure of drinking beer' or 'eating/tasting chocolate'. In the second pair the cognitive effort is a bit more extensive since there is the need for an intraframe metonymic transfer to the *Expressor* 'ear' and 'mouth' as the receptors for the pleasurable experience – 'a sense of pleasure derived from listening to some-

---

8 Meaning definitions of all members of the *-gasm* family used in the chapter are provided in Appendix 1.

thing, particularly music. A sense of pleasure derived from physical stimulation of the earlobe or ear canal, such as when using a Q-Tip to clean the ear' and 'a sense of pleasure derived from eating food which tastes good'. In the EXCITEMENT frame-driven and supported meaning computation of *nerdgasm, fangasm, geekgasm*, etc., the *Experiencer* frame element is saturated but the fixation of the multistable conceptual construal is not as straightforward as in the case of *Expressor*. While in the *Expressor* script, the relation of the experiencer is directly evocated and is sensory in nature, in the *Experiencer* script, the nature of the experiencer is underspecified and remains semantically multistable and opened to interpretation, as it is basically similative, encoding a type of experience resembling the one a nerd experiences when enjoying their niche interest. First, the sensory pleasure meaning is overridden by the secondary, derived meaning of enthusiasm, hence the *-gasm* family follows the meaning differentiation and specializiation of the source word for the splinter. Second, the focus is on the nature of the experience: *nerdgasm* – 'a thrill of excitement felt in response to something relating to a subject in which a person has an obsessive interest', which leaves a wide margin for interpretation, which can be only contextually narrowed down to a relevant unitary meaning, since the nature of the experience is dependent on the subject of interest.

The multistability of the construal can also be inferred from the fact that all possible slots can be and have already been saturated as exemplified below:

Experiencer – *nerdgasm, horgasm, herogasm, Berbgasm*
Content – *laughgasm, chillgasm, coregasm*
Stimulus – *beergasm, beefgasm, cakegasm*
Topic – *yogasm, Berbgasm, deathgasm*
Expressor – *eyegasm, mouthgasm, footgasm*
Reason – *postergasm, bookgasm, ragegasm*

The most challenging conceptually and necessarily requiring multistable meaning construal are the *gasms* in which *Reason* determines the nature of *R* (especially when there is a clash or incongruity between the splinter *-gasm* and the left constituent such as in *ragegasm*). The challenge of conceptual combinations with *Reason* can be illustrated with revealing examples. The following three *-gasm* blends, incorporating *Reason*, illustrate the required creative effort (or the need for specific shared background knowledge) for understanding *gasms* as these are recorded in Urban Dictionary (on the use of crowdsourced dictionaries as a source for analyzing word formation phenomena see Cotter and Damaso 2007 and Sajous, Josselin-Leray and Hathout 2018):

Postergasm

1. n, A "Discordian" Project / activity involving putting posters and stickers up all over the place. Meme bombs, surreal phrases or images, and Operation: Mindfuck pages are preferable. 2. v, an expression of joy relating to postergasm.

> When I realized I wasn't the only one in my neighborhood putting up posters, I had a POSTERGASM. by Cramulus June 4, 2008 (https://www.urbandictionary.com/define.php?term=postergasm)

Bookgasm

Getting excited from sneaking Discordian, eccentric, alternative, controversial, pornographic or illegal books or tracts into places they would never be found. Also the act of introducing such books. Books may be brought in or donated, often hidden with other books to hide their true nature. Popular places to place books and tracts are public libraries, jails, prisons, houses of worship, church pews, motels, hotels, book stores, public restrooms, park benches, or almost anywhere in China, Iran or Iraq.

> I just snuck The Autobiography of Minnie Rae into the jail in Mad Dog, Texas, and OMG! I'm having a bookgasm! (https://www.urbandictionary.com/define.php?term=BookGASM)

Ragegasm

1) A fit of unrestrained [. . .] anger. One step up from apoplectic rage.; 2) The climax of make-up sex where the 'make-up' issues haven't been completely resolved and one or more of the involved parties are still angry.

> My girlfriend threw me out, I stubbed my toe and then dropped my wallet down the grid. I had a complete ragegasm in the middle of the street.
> by TheNexus May 19, 2011 (https://www.urbandictionary.com/define.php?term=Ragegasm)

All three examples belong to the EXCITEMENT subschema in the *-gasm* family. The first two look like the result of saturating the *Topic* role, while the third of the *Content* one. Their meaning, explained by users in Urban Dictionary, reveals that it is the *Reason* role that is saturated in all of them. All three pose an intellectual challenge (in the sense of Kjellander 2019) for figuring out the fitting interpretation of the multistable meaning construal. The most straightforward interpretation of the concept combination (the reading of *R* in the composition of the blends) in the first two may be 'seeing a nice poster or reading a nice book causes pleasure'. The lexical meaning of *postergasm*, different than this straightforward *R* interpretation, opens a wider script and encodes rich content which may contextually be easily interpretable, although it might require familiarity with the Discordianism movement (variously defined as a counterculture, a religion or so-

cial commentary). The same applies to *bookgasm*, where the excitement is associated with sneaking in books with marked content. The recognition of creativity derives from the complexity of the hidden scenario and the reference to a culturally specific practice. The creativity of *ragegasm* stems from the clash of the concepts that are combined in the blend. No specific (communally restricted shared) knowledge is necessary, but cognitive effort is required for resolving the incongruous combination of antonyms. Besides, the emotive-affective nature of the concepts combined further enhances emotional engagement and the perception of creativity. The discordancy of the concepts requires resolving the incongruity and computing the meaning of the blend. The semantics of 'excessive degree' in the first sense of the blend is not immediately derivable from the incongruous clash. The need to reconcile the two concepts into a meaningful unity invites an elaborate scenario in which the *-gasm* part via its association with the source word for the splinter implies an end or highest point and leads to the meaning of excess. The incongruity resolution in the second sense of the blend, which conforms to the PLEASURE frame subschema takes a different route: the pleasure is derived while in a rage. This interpretation does not correspond to the saturation of an immediately recognizable frame slot. The slot that ranks highest in saturation is *Content*. The interpretation requires less cognitive effort as it associates with the initial meaning of the splinter and alludes to the initial inuendo.

A further space for creativity is the possibility for recognizing different source words for the left constituent as in *exploragasm* (interpreted by Barrena Jurado [2019]) as *Experiencer*: *explorer* + *-gasm*, i.e., whatever an explorer feels when discovering something significant or alternatively it could have been derived from *exploration* + *-gasm* and may be interpreted as the enthusiasm for exploration, an attitude to life focused on exploring novel things). Even when the constituents are clear, multiple interpretations are possible due to the indeterminacy of the *R* link (and the chosen frame slot for saturation), as in *Berbagasm*, which may be assumed to express the pleasure Berbatov feels when scoring a goal (*Experiencer*) or it can have a *Content* interpretation of experiencing pleasure when Berbatov scores in a match or when just watching him play or even have a similative meaning, i.e. someone experiencing feelings imitative of the ones Berbatov has when scoring a goal (which explains the appearance of the *Berbgasm* lexeme with different saturated frame roles in the list above). It is exactly this pluripotentiality of conceptual combinations that still renders *gasm* blends creative (non-routinized production of new types or non-conventionalized interpretations), requiring cognitive effort for both creator and comprehender.

The *-gasm* family expands by left constituents derived from source words fulfilling any of the identified frame constituents and fragments into semantic niches (understood as defined by Hüning (2009: 183). Semantic niches are subgen-

eralized unities of "word-formation processes [which] often show semantic fragmentation: in the course of time, they develop 'semantic niches', i.e., groups of words (subsets of a morphological category) kept together by formal and semantic criteria and extendable via analogy". Such extensions are possible within each subschema via hyponymic relations among the left-hand source words. For example, in the first subschema, there is a whole series of *foodgasms*, which may be generalized into a *consumption* niche: *beergasm, soupgasm, beefgasm, cakegasm flavourgasm, chocogasm, deligasm, beefgasm, Swirlgasm*, etc. In the second subschema the generalized schema can be termed the *occupation* niche: *blogasm, bookgasm, astrogasm, yogasm, wordgasm*, etc., where the leftmost constituent names an activity with which the Experiencer of the excitement/enthusiasm is dealing (either as permanent occupation, a hobby or a chance activity) and which functions as the *Stimulus*. The family seems to have developed semi-hyponymic sets, i.e., ones based on metonymy-based frame switches (e.g., *Biebergasm, Ga-Gasm/Gagasm, guitargasm, bassgasm*, etc.), which testify to reduced creativity and enhancing productivity and routinization. None of the members of the family are recorded in OED (19 August 2023) and varying frequencies in different sources (e.g., COCA, GloWbE, enTenTen, for details see the frequency table in Barrena Jurado 2019 and Bauer et al. 2019, Belieaeva 2019) are reported. All *-gasm* products are qualified as neologisms (Bauer et al. 2019). Though advanced on the path of routinization, *-gasm* blends still retain properties that render many of them creative.

Their spontaneous creation and comprehension are aided by the common cognitive and communicative strategies. At the disposal of language users for diverse conceptual-linguistic human behaviors are various recurrent cognitive operations (for the use of cognitive operations in cognitive linguistics for analytical purposes see Anderson 2010; Ruiz de Mendoza 2011; Ruiz de Mendoza and Galera 2014: 92–96). Among these varied cognitive operations, the following have been identified as playing a significant role in establishing and interpreting meaning relations within lexical blends: parameterization (both metonymic and non-metonymic), generalization, contrast, resemblance, strengthening, expansion, and reduction (Peña-Cervel 2022). In the *gasm* family most frequently utilized seem to be parametrization, contrast, resemblance and expansion. Their wide application in language processing (Ruiz de Mendoza 2022) aids communicative intercomprehension despite unexpected source word choices for filling the first slot in the intermediate *-gasm* schema.

4 Contextualizing the Case Studies in the Creativity Debate

The differentiation between word creation and word formation (see Ronneberger-Sibold 2010, 2015) is temporal in nature and does not involve any qualitative difference between the formal mechanisms and cognitive operations involved. Each lexical item starts as a "product of nonce-formation" (Schmid 2008: 3) associated with a "floating concept" (Ungerer 2003: 563) and (usually) ends up as a conventionalized lexeme encoding a hypostatized concept (Schmid 2008: 3). This interpretation runs parallel to F- and E-creativity, or as the recognition of creativity and routine as the enclosing extremes of a scale. Drawing the parallel allows one to reinterpret word creation and word formation in the same manner as the extreme points of a gradient between nonce-formation and a conventionalized lexeme as fully corresponding to the endpoints encapsulating the complex ontological dynamicity between creativity and routine.

The formal mechanisms are constant and routine-bound, since a condition sine qua non for intercomprehension is a precarious balance between conventionalization (in terms of the product), routinization (in terms of shared behaviours) and creativity whereby effectiveness (mutual understanding) strives to constrain novelty. Most flexible for the hypostatization of new concepts or for naming familiar concepts in novel ways appear to be reanalysis + deaffixation, affixation and blending. The common feature across these processes is conceptual flexibility and formal experimentation.

It remains extremely difficult to discriminate between E and F creativity since "the cumulative production of ever more complex concepts and artifacts is unique to humans" (Roberts et al. 2021: 102). This cumulative effect and the essence of language as an emergent complex adaptive system significantly impede the teasing apart and analytical modeling of contributing factors, interrelated agencies and agents, processes and products, converging forces in creativity in distributed cognition and its inception in micro events (interpersonal communicative acts) with rippling consequences.

It is hard to deny that the "individual speaker is the central factor with regard to all linguistic phenomena" (Koefoed and van Marle 2000: 311), but it takes two to tango and a community of practice to recognize the tango, practice it and appreciate its beauty. Only when the comprehender is able to reverse-engineer the mechanisms employed by the creator of the unexpected product and establish the motivational links can creativity of a linguistic element be recognized and registered. This is a requirement of the effectiveness dimension in the constitution of creativity (at least in the scenario with a targeted balance between origi-

nality/novelty and effectiveness as in the standard definition of creativity, see Runco and Jaeger 2012). The aesthetic dimension (positive vs. negative valuation) of the product in an act of creativity imposes another dimension of distinguishing between creative and wrong, which lies entirely within the perception side and may encompass behaviors ranging from genuine creative acts to criminal deeds (see Uhrig 2020).

A 'creative' product or any neologism is not a qualitatively different type of word, rather it is the spatio-temporal uniqueness within the immediate context of co-semiosis in the initial fabrication of a word whose socio-pragmatically determined diffusion will result in varying degrees of entrenchment and conventionalization leading to routinization (for a detailed elaboration of the complex interaction of all agents and dimensions involved see Schmid 2020; Schmid this volume) that is recognized as creative. Such "'lexicogenetic' mechanisms [. . .] involve changes through which a concept, regardless of whether or not it has previously been lexicalized, comes to be expressed by a new or alternative lexical item" (Geeraerts 2010: 26). The sharedness in the execution and perception of creativity between creator and comprehender has been succinctly summarised by de Beaugrande (1978: 9 [emphasis mine]), as "the process whereby we become aware of the present and *possible conditions* for the *organization of cognition*, and whereby we enable others to *reenact* that awareness". Creativity is always embedded in the aesthetics of the specific context and despite the analytically recognized and modeled differences between psychological (individual) creativity vs. societal (historical/linguistic) creativity the mechanisms are the same.

The measurable dimensions of creativity are novelty and added value calculated within the relationship between creator, product and perceiver, which can be modeled analytically in attitudinal studies, psycholinguistic experimentation, computational modeling and corpus-based statistical analyses. The greatest difficulty in studying creativity in word formation arises from the ontological dynamics of creativity and routine as one of perpetual recycling.

The gradient concept of novelty can only be utterly comprehended in the context of the dimensions of conformity, since creativity understood as one of its features, novelty, actually defies some of these dimensions to differing degrees. In the Entrenchment-and-Conventionalization (EC) model of language (Schmid 2015, 2020), the basic forces involved in the maintenance of equilibrium of language, which is in constant flux between stability and change, are in perpetual motion as in a simple Tinguely machine (Schmid 2020: 4). Creativity starts within breaking up routine within entrenchment (individual's linguistic knowledge and habits), introducing novelty and moving away from conformity and via usage expands over into the realm of conventionalization via diffusion (communal usage habits and cognitive knowledge), where the endpoint is a fully established lex-

eme. Within this model novelty may be measured by degree of violation of a conformity type. The conformity types (Schmid 2020: 298) are: onomasiological conformity, semasiological conformity, syntagmatic conformity, contextual conformity, and community-related conformity. Various kinds of associations (which constitute different levels of associative networks) establish seamless continuity and correlations between the conformity types: symbolic associations, paradigmatic associations, syntagmatic associations and pragmatic associations. Most tightly knit are onomasiological and semasiological conformity, whose relations are supported by symbolic and paradigmatic associations (Schmid 2020: 298). These associations are driven and fed by

– cognitive forces: similarity, contiguity, salience, categorization, gestalt processing;
– pragmatic forces: settings, participants, event types, intentions, goals;
– emotive-affective forces: egocentrism, emotion, need for admiration, fun, empathy;
– social forces: social networks, identity, solidarity, peer-group pressure, prestige (Schmid 2015: 9).

The association types are subservient to entrenchment or the complexity of processes via which language "happens" in an individual's mind (Schmid 2015: 6): "association, routinization and schematization." The conformity types are subservient to the societal (communal) processes of the emergence of language "subsumed under the label *conventionalization:* innovation, co-adaptation, diffusion and normation" (Schmid 2015: 6). The two are interconnected "by the feedback-loop processes of usualization and diffusion" (Schmid 2020: 298). Extrapolating these to the analysis of creativity in word formation, we can notice that in the creation of a novel lexical item semasiological conformity is violated and a novel onomasiologically driven symbolic association is established, which may or may not get communally propagated via diffusion. The understanding of the novel lexical item necessitates the establishment of a novel syntagmatic association within the individual communicative act, and the initiation of a novel co-textual and contextual links. The novelty gradient (along which change occurs) is postulated to encompass "complete novelty" (under which "creative coining" is subsumed), "salient innovation" (or "highly unconventional variants of variable patterns") and "non-salient innovation" ("new words formed by recourse to a productive word-formation pattern") (Schmid 2020: 310–311). The three case studies that we presented above, showcase "salient innovation" in the novel types of *-gasms* and "salient innovation" in the tongue-in-cheek poem and the translation of Shakespearean tragedies into Bulgarian), where the novelty results in non-routine recombinance patterns of elements, employing fully productive processes. This leads to the conclusion that for novelty to

occur or register routinized linguistic habits (be they individual or societal) have to be creatively (originally and effectively) violated.

Routine is a precondition for creativity as without shared cognitive and communicative routines users cannot achieve mutual comprehension, while creativity has a natural temporal dimension where replications in instances of co-semiosis lead to conventionalization and routinization in the societal dimension, as well as entrenchment from the individual's cognitive perspective.

5 Conclusions

The three cases studies suggest that there is no correlation between the separate parameters of perceived or imputed heightened creativity and specific creativity marked strategies for expressing/encoding it linguistically. In relation to word formation phenomena individual creativity, societal creativity, product creativity and conceptual creativity cannot be easily teased apart as they constitute complexly intertwined cogs in the mechanism of language as an emergent, complex adaptive system, since all of them are involved in word formation creativity.

There are no sufficient grounds for ranking the separate word formation processes in terms of creativity as any of the existing ones may be employed creatively under the opportune pragmatic and contextual circumstances. As the analyses of blending indicates, even though it heightens creativity when present, individual conscious and effortful creativity is neither a necessary, nor a sufficient condition. The properties which surface as necessary and uncancellable for the recognition of creativity are the attention-grabbing properties of a lexical item (which involves violation of isomorphic relations or established routine mappings between meaning and form) and the presence of a clash (incongruity) or contrast in the concepts involved in the hypostatization of a novel one. For any of these to operate interactive relations between creator, comprehendor and product in a specified context are minimally necessary.

Interpreting the diverse theoretical notions employed in the analysis of creativity in language (F- and E-creativity, word creation vs. word formation, nonce-formation vs. conventionalized (listed) lexeme, etc.) leads to the conclusion that creativity and routine in word formation constitute a temporal scale enclosing the first use of a new type and its full conventionalization, potentially measurable by productivity and type and token frequency.

Appendix 1

Glossary of *-gasms* Arranged into Two Sub-Schemas and Alphabetically within Each Group

A) The Pleasure Frame Based Sub-Schema

[Xi gasm$_i$]Nk↔[SEMgasm$_i$ with R *physical or sensual pleasure* to SEMXi]Nk

bassgasm	the pleasure when one hears a bass riff/solo that is awesome beyond belief
beefgasm	A) the process of having an orgasm while eating an Arby's roast beef sandwich; B) an Arby's roast beef sandwich.
beergasm	the pleasure of consuming beer
cakegasm	a mind blowing double orgasm induced by cakes
chillgasm	an orgasm as a result of being really, really relaxed
chocogasm	the pleasure of eating chocolate
cor(e)gasm	an orgasm that happens while you're doing a core exercise or workout
deligasm	pleasure from eating great pastrami, corned beef sandwiches at a good Jewish, kosher or other kosher-style deli restaurant
eargasm	a sense of pleasure derived from listening to something, particularly music. A sense of pleasure derived from physical stimulation of the earlobe or ear canal, such as when using a Q-Tip to clean the ear
eyegasm	a feeling of pleasure derived from a sight
floorgasm	a feeling of excitement from dancing on a dance floor
fruitgasm	to devour any fruits that's so overfuckingwhelming unbelievable sweet which simulates orgasms
guitargasm	A) having an orgasm while playing an incredible guitar; playing such a sweet guitar that your fingers feel so good, its almost sexual; B) becoming sexually aroused while listening and or watching someone shred on a guitar
mouthgasm	a sense of pleasure derived from eating food which tastes good
robogasm	to have an orgasm with a machine or robot or other sex replicant
soupgasm	pleasure obtained from consuming soup
Swirlgasm	the pleasure of eating Swirl yogurt

*Any food or drink may become a constituent under this sub-schema and establish hyponymic relations with *foodgasm.*

B) The Enthusiams Frame Based Sub-Schema

[Xi gasm$_j$]Nk↔[SEMgasm$_j$ with R *excitement or enthusiasm R* to SEMXi]Nk

astrogasm	is about being and having an explosion of excitement about the stars, cosmos, and life
stargasm	when you see a shooting star you weren't expecting and get excited
Berbgasm	the feeling of euphoria or excitement derived from watching football player Dimitar Berbatov play a good match.
Biebergasm	feeling of excitement when listening to music by Canadian singer Justin Bieber
blogasm	the feeling you get when you check your blog stats and you see WAY more visitors than you expected.
bookgasm	A) feeling of enthusiasm experienced when reading a book.; B) getting excited from sneaking Discordian, eccentric, alternative, controversial, pornographic or illegal books or tracts into places they would never be found. Also the act of introducing such books. Books may be brought in or donated, often hidden with other books to hide their true nature. Popular places to place books and tracts are public libraries, jails, prisons, houses of worship, church pews, motels, hotels, book stores, public restrooms, park benches, or almost anywhere in China, Iran or Iraq
deathgasm	a strong feeling of excitement brought on by anything dark or of death, such as music and art felt by death enthusiasts
explorgasm	the feeling of excitement of an explorer
fangasm	A) the feeling or action resulting from experiencing a particularly exciting event within the canon of a certain fandom, especially an event that has been anticipated for a long period of time, has a particular amount of tension building up to it; B) any extreme outburst of energy related to one's obsession with a fandom or an aspect of the fandom, especially when seen in random unrelated situations; C) to experience a fangasm
flavourgasm	when two foods combined create a new never before tasted flavour.
Gagasm	the act of receiving an orgasm from seeing or hearing lady Gaga perform.
geekgasm	a feeling of intense excitement over something geeky
herogasm	an annual festival in which most of Vought-American's superheroes meet secretly at a remote tropical resort for a company-sponsored orgy; The euphoric feeling after getting 100% on a song on Guitar Hero. Identifiable by high pitched screaming, jumping, and the need to show others one's achievement.
horgasm	feeling of pleasure or excitement experienced by (Torstein) Horgmo
laughgasm	that "ooh" orgasm noise a person makes after a really good laugh
mindgasm	an orgasmic-like feeling of excitement from thoughts
neologasm	excitement form coining or finding a neologism
nerdgasm	a feeling of intense excitement felt by someone considered to be a nerd, esp. when caused by something relating to his or her leisure interests such as a new piece of technology, computer game, science fiction film, etc.
ragegasm	A) a fit of unrestrained orgiastic anger. One step up from apoplectic rage; B) the climax of make-up sex where the 'make-up' issues haven't been completely resolved and one or more of the involved parties are still angry
vetogasm	a feeling of excitement form imposing a veto on something or someone

wordgasm	A) when one is overwhelmed by the influx of literal ideas; B) a feeling of excitement or enthusiasm when using or reading a word or when reading a text that somebody has written
yogasm	the euphoric, orgasmic sensation that you feel as if you've just had an intense moment of intimacy. or similar to the deep love that rocks your body when you take a bite of belgian dark chocolate

[*]No clear tendency for predicting hyponymic elaborations can be detected as almost anything can evoke enthusiasm in a plethora of ways.

References

Allan, Keith & Kate Burridge. 2006. *Forbidden Words. Taboo and the Censoring of Language*. Cambridge: Cambridge University Press.

Anderson, John. 2010. *Cognitive psychology and its implications*. New York: Worth Publishers.

Arndt-Lappe, Sabine & Ingo Plag. 2013. The role of prosodic structure in the formation of English blends. *English Language and Linguistics* 17(3). 357–563. https://doi.org/10.1017/S1360674313000154

Arndt-Lappe, Sabine. 2015. Word-formation and analogy. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation – An International Handbook of the Languages of Europe*. Volume 2, 822–841. Berlin: de Gruyter Mouton.

Audring, Jenny. 2019. Mothers or sisters? The encoding of morphological knowledge. *Word Structure* 12(3). 274–296. doi:10.3366/word.2019.0150

Bagasheva, Alexandra & Christo Stamenov. 2014. Trans/formation – creativity in word-formation. A case study of nonce-formations. In Diana Yankova (ed.), *Cross-linguistic Interaction: Translation, Contrastive and Cognitive Studies. Liber Amicorum in Honour of Bistra Alexieva*, 100–124. Sofia: St. Kliment Ohridski University Press.

Barrena Jurado, Alejandro. 2019. A study on the 'wordgasm': The nature of blends' splinters. *Lexis. Journal in English Lexicology* 14. https://doi.org/10.4000/lexis.3916

Barsalou, Lawrence. 2005. Situated conceptualization. In Henri Cohen & Claire Lefebvre (eds.), *Handbook of categorization in cognitive science*, 619–650. St. Louis: Elsevier.

Barsalou, Lawrence. 2008. Cognitive and neural contributions to understanding the conceptual system. *Current Directions in Psychological Science* 17. 91–95.

Barsalou, Lawrence & Christopher Hale. 1993. Components of Conceptual Representation: From Feature Lists to Recursive Frames. In Iven van Mechelen, James Hampton, Ryszard Michalski & Peter Theuns (eds.), *Categories and Concepts: Theoretical Views and Inductive Data Analysis*, 97–144. San Diego: Academic Press.

Bauer, Laurie 2000. Compounding. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, Volume 1, 695–707. Berlin & New York: Walter de Gruyter.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie, Natalia Beliaeva & Elizaveta Tarasova. 2019. Recalibrating productivity: Factors involved. *Zeitschrift Für Wortbildung / Journal of Word Formation* 3(1). 44–81.

Beaugrande, Robert de. 1978. Linguistics and creativity. *Paper presented at the Interdisciplinary Conference on Linguistics* (Louisville, Kentucky, April 1978).

Beckner, Clay, Richard Blythe, Joan Bybee, Morten Christiansen, William Croft, Nick Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann (The "Five Graces Group"). 2009. Language is a complex adaptive system: Position paper. In Nick Ellis & Diane Larsen-Freeman (eds.), *Language as a Complex Adaptive System*, 1–26. New Jersey: Wiley Blackwell. doi: 10.1111/j.1467-9922.2009.00533.x.

Beliaeva, Natalia. 2014. A study of English blends: From structure to meaning and back again. *Word Structure* 7(1). 29–54. https://doi.org/10.3366/word.2014.0055.

Beliaeva, Natalia. 2016. Blends at the intersection of addition and subtraction: Evidence from processing. *SKASE Journal of Theoretical Linguistics* 13(2). 23–45.

Beliaeva, Natalia. 2019. Blending creativity and productivity: On the issue of delimiting the boundaries of blends as a type of word formation. *Lexis. Journal in English Lexicology* 14. https://doi.org/10.4000/lexis.4004

Bergs, Alexander. 2018. Learn the Rules like a Pro, so you can Break them like an Artist (Picasso): Linguistic Aberrancy from a Constructional Perspective. *ZAA* 66(3). 277–293. https://doi.org/10.1515/zaa-2018-0025.

Bergs, Alexander. 2019. What, if anything, is linguistic creativity? *Gestalt Theory* 41(2).173–184.

Boden, Margaret. 1990. *The creative mind: Myths and mechanisms*. New York: Basic Books.

Booij, Geert. 2007. *The grammar of words: An introduction to linguistic morphology*. Oxford: Oxford University Press.

Booij, Geert. 2010a. *Construction morphology*. Oxford: Oxford University Press. doi: 10.1111/j.1749-818X.2010.00213.x.

Booij, Geert. 2010b. Compound construction: Schemas or analogy? A construction morphology perspective. In Sergio Scalise & Irene Vogel (eds.), *Cross-Disciplinary Issues in Compounding*, 93–107. Amsterdam & Philadelphia: John Benjamins.

Booij, Geert. 2016. Construction morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 424–448. Cambridge: Cambridge University Press. doi: 10.1017/9781139814720.016.

Booij, Geert. 2017. The construction of words. In Barbara Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistic*s, 96–433. Cambridge: Cambridge University Press [Kindle edition].

Burridge, Kate & Alexander Bergs. 2017. *Understanding Language Change*. Oxon & New York: Routledge.

Ciecierski, Tadeusz. 2021. Indexicality, meaning, use. *Semiotica* 238. 73–89.

Cotter, Coleen & John Damaso. 2007. Online dictionaries as emerging archives of contemporary usage and collaborative lexicography. *Queen Mary Occasional Papers Advancing Linguistics* #9 (OPALS). Available at: https://www.qmul.ac.uk/sllf/media/sllf-new/department-of-linguistics/09-QMOPAL-Cotter-Damaso.pdf.

Croce, Benedetto (ed.). 1994 [1902]. *Aesthetic as science of expression and general linguistics*, 2nd edn. London & New York: Routledge.

Croft, William. 2001. *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.

Dressler, Wolfgang & Lavinia Barbaresi. 1994. *Morphopragmatics: Diminutives and intensifiers in Italian, German, and other languages*. Berlin: Walter de Gruyter.

Estes, Zachary, Sabrina Golonka & Lara Jones. 2011. Thematic thinking: the apprehension and consequences of thematic relations. In Brian Ross (ed.), *Psychology of Learning and Motivation*, vol. 54, 249–294. Burlington, MA: Academic Press.

Evans, Vyvyan. 2016. Design features for linguistically-mediated meaning construction: The relative roles of the linguistic and conceptual systems in subserving the ideational function of language. *Frontiers in Psychology 7* (art. 156). doi: 0.3389/fpsyg.2016.00156.

Fandrych, Ingrid. 2008. Sub-morphemic elements in the formation of acronyms, blends and clippings. *Lexis. Journal in English Lexicology* 2. 103–121.

Fernández-Domínguez, Jesus. 2010. Productivity vs. lexicalisation: Frequency-based hypotheses on word-formation. *Poznań Studies in Contemporary Linguistics* 46(2). 193–219.

Fillmore, Charles. 2006 [1982]. Frame semantics. In Dirk Geeraerts (ed.), *Cognitive Linguistics. Basic Readings*, 373–400. Berlin & New York: De Gruyter Mouton.

Gaeta, Livio. 2010. Analogical change. In Silvia Luraghi & Vit Bubenik (eds.), *Continuum Companion to Historical Linguistics*, 147–160. London & New York: Continuum.

Gagné, Christina. 2000. Relation-based combinations versus property-based combinations: A test of the CARIN theory and the dual-process theory of conceptual combination. *Journal of Memory and Language* 42(3). 365–389. doi: 10.1006/jmla.1999.2683.

Gagné, Christina & Edward Shoben. 1997. Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(1). 71–87. doi: 10.1037/0278-7393.23.1.71.

Gagné, Christina & Thomas Spalding. 2006. Conceptual Combination: Implications for the mental lexicon. In Gary Libben & Gonia Jarema (eds.), *The Representation and Processing of Compound Words*, 145–168. Oxford: Oxford University Press.

Gagné, Christina & Thomas Spalding. 2015. Semantics, concepts, and metacognition: Attributing properties and meanings to complex concepts. In Laurie Bauer, Livia Körtvélyessy & Pavol Štekauer (eds.), *Semantics of Complex Words*, 9–27. Dordrecht: Springer.

Gallese, Vittorio & George Lakoff. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology* 22(3). 455–479. doi: 10.1080/02643290442000310.

Garner, Bryan. 1987. Shakespeare's Latinate neologisms. In Vivian Salmon & Edwina Burness (eds.), *A reader in the language of Shakespearian drama*, 207–229. Amsterdam & Philadelphia: John Benjamins.

Geeraerts, Dirk. 2010. *Theories of Lexical Semantics*. Oxford: Oxford University Press.

Gries, Stefan. 2012. Quantitative corpus data on blend formation: Psycho- and cognitive linguistic perspectives. In Vincent Renner, François Maniez & Pierre Arnaud (eds.), *Crossdisciplinary perspectives on lexical blending*, 145–167. Berlin: De Gruyter Mouton.

Hamans, Camiel. 2011. About uniqueness and productivity of blends. *Folia Linguistica et Litteraria* 5. 69–79.

Haspelmath, Martin. 1995. The Growth of Affixes in Morphological Reanalysis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1994*, 1–29. Boston: Kluwer.

Haspelmath, Martin. 2002. *Understanding morphology*. London: Arnold.

Hoffmann, Thomas. 2018. Creativity and construction grammar. Cognitive and psychological issues. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 259–276.

Hunsaker, Michael & Raymond Kesner. 2013. The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neurosci Biobehav Rev* 37(1). 36–58. doi: 10.1016/j.neubiorev.2012.09.014.

Hutchins, Edwin. 2010. Cognitive Ecology. *Topics in Cognitive Science 2*. 705–715. https://doi.org/10.1111/j.1756-8765.2010.01089.x.

Hüning, Matthias. 2009. Semantic niches and analogy in word-formation: Evidence from contrastive linguistics. In *Languages in contrast* 9(2). 183–201.

Jacobson, Ronald. 2004 [2000]. On linguistic aspects of translation, in Lawrence Venuti (ed.), *The Translation Studies Reader*, 2nd edn. 138–143. London & New York: Routledge.

Jones, Rodney (ed.). 2016. *The Routledge Handbook of Language and Creati*vity. London & New York: Routledge.

Jones, Lara & Sabrina Golonka. 2012. Different influences on lexical priming for integrative, thematic, and taxonomic relations. *Frontiers in Human Neuroscience* 6(205). https://doi.org/10.3389/fnhum.2012.00205

Jordanous, Anna & Bill Keller. 2016. Modelling Creativity: Identifying Key Components through a Corpus-Based Approach. *PLoS ONE* 11(10), Article e0162959. https://doi.org/10.1371/journal.pone.0162959.

Kaufman, James & Robert Sternberg. 2010. *The Cambridge Handbook of Creativity*. Cambridge: Cambridge University Press.

Kemmer, Suzanne. 2003. Schemas and lexical blends. In Hubert Cuyckens, Thomas Berg, René Dirven & Klaus-Uwe Panther (eds.), *Motivation in Language: From Case Grammar to Cognitive Linguistics. Studies in Honour of Günter Radden*, 69–97. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.243.08kem.

Kjellander, Daniel. 2019. Gold Punning: studying multistable meaning structures using a systematically collected set of lexical blends. *Lexis Journal in English Lexicology* 14. 1–28. DOI: 10.4000/lexis.3962.

Kjellander, Daniel. 2018. Cognitive constraints in English Lexical Blending. A data collection methodology and an explanatory model. *Pragmatics and Cognition* 25 (1). 142–173. DOI: 10.1075/pc.18003.kje.

Koefoed, Geert and Jaap van Marle. 2000. Productivity. In Geert Booij, Christian Lehmann, Joachim Mugdan, Wolfgang Kesselheim & Stavros Skopeteas (eds.) *International Handbook on Inflection and Word-formation*. Vol. I, 303–311. Berlin: Walter de Gruyter.

Körtvélyessy, Lívia, Pavol Štekauer & Pavol Kačmár. 2021. On the role of creativity in the formation of new complex words. *Linguistics*. 1–39. https://doi.org/10.1515/ling-2020-0003.

Langacker, Ronald. 2008. *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195331967.001.0001.

Lederer, Richard. 1991. *The Miracle of Language*. New York: Pocket Books.

Lieber, Rochelle. 2010. *Introducing morphology*. Cambridge: Cambridge University Press.

Mattiello, Elisa. 2013. *Extra-grammatical morphology in English: abbreviations, blends, reduplicatives, and related phenomena*. Berlin: Walter de Gruyter.

Medin, Douglas & Lance Rips. 2005. Concepts and Categories: Memory, Meaning, and Metaphysics. In Holyoak, Keith & Robert Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning*, 37–72. Cambridge: Cambridge University Press.

Munat, Judith. 2010. Creation or recreation? Word formation processes and word creation strategies as components of lexical creativity. In Sandrine Sorlin (ed.), *InventiveLinguistics*, 147–157. Montpellier: Presses Universitaires de la Méditerranée.

Murphy, Gregory. 1988. Comprehending Complex Concepts. *Cognitive Science* 12 (4). 529–562. doi: 10.1207/s15516709cog12042.

New York Times (19 April 2010) https://archive.nytimes.com/schott.blogs.nytimes.com/2010/04/19/brev-is-the-soul-of-wit/ (last accessed 20 May 2024).

Oliveira, Roberta Piers de & Robson de Souza Bittencourt. 2008. An interview with Mark Johnson and Tim Rohrer: From neurons to sociocultural situatedness. In Roslyn Frank, René Dirven, Tom Ziemke & Enrique Bernárdez (eds.), *Body, Language and Mind, Vol. 2: Sociocultural Situatedness*, 21–51. Berlin & Boston: De Gruyter.

Pancheva, Evgeniya. 2014. Укротяване на публиката [Taming the audience]. Проблеми на изкуството [*Art Studies Quarterly*] 2. 18–23. ISSN: 0032-9371. https://artstudies.bg/wp-content/uploads/2015/12/BROI2-summaries-content.pdf

Pancheva, Evgeniya. 2013. Да пренесеш Шекспир през времето [To transport Shakespeare through time]. Култура [*Culture*] 4 (2975), 01 February 2013. https://newspaper.kultura.bg/bg/article/view/20562.

Peña-Cervel, Mª Sandra. 2022. Lexical blending in terms of cognitive modelling. In Alexandra Bagasheva, Bozhil Hristov & Nelly Tincheva (eds.), *Figurativity and Human Ecology*, 275–304. Amsterdam & Philadelphia: John Benjamins.

Plag, Ingo. 2018 [2003]. *Word-formation in English*, 2nd edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316771402.

Renner, Vincent. 2015. Lexical blending as wordplay. In Angelika Zirker & Esme Winter-Froemel (eds.), *Wordplay and metalinguistic/metadiscursive reflection: Authors, contexts, techniques, and meta-reflection*, 119–133. Berlin & Boston: Mouton de Gruyter.

Renner, Vincent. 2020. An ecosystem view of English word-formation. *The Mental Lexicon* 15 (1). 4–20. https://doi.org/10.1075/ml.00011.ren.

Roberts, Anne, Robert Sternberg, Mark Runco, Selcuk Acar, Thomas B. Ward, Yuliya Kolomyts & James C. Kaufman. 2021. Creativity and Cognition, Divergent Thinking, and Intelligence. In James C. Kaufman & Robert J. Sternberg (eds.), *Creativity. An Introduction*, 102–128. [Publikationsort]: Cambridge University Press.

Ronneberger-Sibold, Elke. 2006. Lexical Blends: Functionally Tuning the Transparency of Complex Words. *Folia Linguistica* 40 (1–2). 155–181.

Ronneberger-Sibold, Elke. 2010. Word Creation: Definition – Function – Typology. In Franz Rainer, Wolfgang U. Dressler, Dieter Kastovsky & Hans Christian Luschützky (eds.), *Variation and Change in Morphology*, 201–216. Amsterdam & Philadelphia: John Benjamins.

Ronneberger-Sibold, Elke. 2015. Word-creation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An International Handbook of the Languages of Europe*, vol. 1, 485–500. Berlin & Boston: De Gruyter.

Ruiz de Mendoza Ibáñez, Francisco José. (2011). Metonymy and cognitive operations. In Réka Benczes, Antonio Barcelona & Francisco José Ruiz de Mendoza Ibáñez (eds.), *Defining Metonymy in Cognitive Linguistics*, 103–124. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/hcp.28.06rui.

Ruiz de Mendoza, Francisco & Alicia Galera. 2014. *Cognitive modeling. A linguistic perspective*. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/hcp.45.

Ruiz de Mendoza Ibáñez, Francisco José. 2022. Linguistic and metalinguistic resemblance. In Bagasheva, Alexandra, Hristov, Bozhil and Nelly Tincheva (eds.), *Figurativity and Human Ecology*, 15–41. Amsterdam & Philadelphia: John Benjamins.

Runco, Mark & Garrett Jaeger. 2012. The standard definition of creativity. *Creativity Research Journal* 24 (1). 92–96. https://doi.org/10.1080/10400419.2012.650092.

Sajous, Franck, Amélie Josselin-Leray & Nabil Hathout. 2018. The Complementarity of Crowdsourced Dictionaries and Professional Dictionaries viewed through the Filter of Neology. *Lexis Journal in English Lexicology* 12. https://doi.org/10.4000/lexis.2322.

Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence. Experiment and argument in linguistics and philosophy of language*, 15–26. Bern: Peter Lang.

Schmid, Hans-Jörg. 2008. New words in the mind: concept-formation and entrenchment of neologisms. *Anglia-Zeitschrift für englische Philologie* 126 (1). 1–36. doi: 10.1515/angl.2008.002.

Schmid, Hans-Jörg. 2015. A blueprint of the Entrenchment-and-Conventionalization Model. *GCLA* 3. 3–25.

Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press.

Schopenhauer, Arthur. 1851. Parerga und Paralipomena: Kleine Philosophische Schriften. Volume 2, Section: 76, p. 93. Berlin: A. W. Hayn.

Sharifian, Farzad. 2015. Cultural Linguistics. In Farzad Sharifian (ed.), *The Routledge Handbook of Language and Culture*, 473–493. New York & London: Routledge/Taylor and Francis.

Sharifian, Farzad. 2017. *Cultural Linguistics*. Amsterdam & Philadelphia: John Benjamins.

Shurbanov, Alexander (translator). 2012a. *Шекспир. Великите трагедии*. [*Shakespeare. The Great Tragedies*]. Sofia: Iztok-Zapad.

Shurbanov, Alexander. 2012b. Шекспир трябва да се превежда на всеки 20–30 години [Shakespeare has to be translated every 20 or 30 years. Interview in *24 Hours* newspaper, 27.12.2012 12:11]. Available at: https://www.24chasa.bg/mneniya/article/1690614 (last accessed 27 August 2023).

Shurbanov, Alexander. 2004. The translatability of Shakespearean texts into an unrelated language/ culture. In Rui Carvalho Homem & Ton Hoenselaars (eds.), *Translating Shakespeare for the twenty-first century*, 51–64. Amsterdam & New York: Rodopi.

Simonton, Dean Keith. 2011. Creativity and discovery as blind variation: Campbell's (1960) BVSR model after the half-century mark. *Review of General Psychology* 15 (2). 158–174. https://doi.org/ 10.1037/a0022912.

Stein, Morris, J. 1953. Creativity and Culture. *The Journal of Psychology* 36 (2). 311–322.

Sternberg, Robert & Todd Lubart. 1999. The concept of creativity: Prospects and paradigms. In Robert Sternberg (ed.), *Handbook of creativity*, 3–15. Cambridge: Cambridge University Press.

Štekauer, Pavol. 2005. Onomasiological approach to word-formation. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 207–232. Dordrecht: Springer.

Talmy, Leonard. 2008 [2003]. Recombinance in the Evolution of Language. *CLS 39: The Panels* 2nd ed. 33–66.

Uhrig, Peter. 2020. Creative intentions – The fine line between 'creative' and 'wrong'. *Cognitive Semiotics* 13 (1). 20202027. https://doi.org/10.1515/cogsem-2020-2027.

Ungerer, Friedrich. 1999. Iconicity in word-formation. In Max Nänny & Olga Fischer (eds.), *Form Miming Meaning*, 307–324. Amsterdam/Philadelphia: John Benjamins.

Ungerer, Friedrich. 2003. The conceptual function of derivational word-formation in English. *Anglia* 120 (4). 534–567. doi: 10.1515/

Wisniewski, Edward. 1997. When concepts combine. *Psychonomic Bulletin and Review* 4 (2). 167–183. https://doi.org/10.3758/BF03209392.

Zawada, Britta. 2006. Linguistic Creativity from a cognitive perspective. *Southern African Linguistics and Applied Language Studies* 24 (2). 235–254.

Ziegeler, Debra. 2007. A word of caution on coercion. *Journal of Pragmatics* 39 .(5). 990–1028.

Zwaan, Rolf. 2004. The Immersed Experiencer: Toward an Embodied Theory of Language Comprehension. In Brian Ross (ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* Vol. 44, 35–62. Amsterdam: Elsevier Science.

Reference Materials

COCA: https://www.english-corpora.org/coca/ (last accessed 23 August 2023).

Collins English Dictionary – Complete & Unabridged 2012 Digital Edition © William Collins Sons & Co. Ltd. 1979, 1986 © HarperCollins Publishers.

enTenTen: https://www.sketchengine.eu/ententen-english-corpus/ (last accessed 17 August 2023).

Etymonline: https://www.etymonline.com/word/orgasm#etymonlinev7143 (last accessed 11 August 2023).

FrameNet: https://framenet.icsi.berkeley.edu/ (last accessed 21 August 2023).

GloWbE: https://www.english-corpora.org/glowbe/ (last accessed 19 August 2023).

Oxford English Dictionary: https://www.oed.com/ (last accessed 25 August 2023).

Urban Dictionary: https://www.urbandictionary.com/

Raymond W. Gibbs, Jr.

Are Creative and Routine Language Really that Different?

Abstract: There is a wide-spread belief that creative and routine language arise from different mental/linguistic processes and convey rather different types of linguistic meaning. This assumption may not accurately describe real, online language production and understanding processes. In most cases, creative language use is not that difficult or cognitively burdensome for speakers/writers. Creativity does not necessarily require people to be consciously aware of their production of innovative instances of language. At the same time, routine language is not interpreted automatically, but makes use of many of the same mental/linguistic routines employed in creative language use. A general theme of this chapter is that people rarely aim to be simply creative or routine when they speak or write as they typically have task-specific and context-specific goals that are interacting constraints in determining the language they use.

1 Introduction

Linguists and literary scholars take great pleasure in closely examining speech and writing for evidence of people's creativity. There is nothing better than coming across a new word or phrase, or some familiar expression that has been used in a novel manner. We ask "What does that mean?" and "How was that created?" We are interested in routine, familiar, sometime formulaic, language for the insights it provides on entrenched linguistic patterns and how such language communicates so effectively. But we also thoroughly enjoy contrasting creative and routine language, looking for systematic differences between these two types that may reveal something about the seemingly distinct ways people think and express themselves (e.g., creatively with deliberation or routinely with little forethought). One possibility is that creativity exists along a continuum, from widely creative to stubbornly routine. But the belief that creative and routine language are substantially different is a guiding assumption in much ongoing linguistic and psycholinguistic scholarship.

My goal in this chapter is to raise some complications regarding the possibility of drawing any simple division between creative and routine language. I am especially concerned with the tacit belief that creative language typically emerges from deliberate, conscious effort on the part of speakers and writers, while vari-

ous forms of routine, formulaic language may be easily, automatically, produced and interpreted. These complications are most evident when we go beyond just examining individual examples of words or phrases and dig deeper into the complex realities of people's specific language tasks (i.e., the implicit or explicit pragmatic goals that people aim to achieve). A look at these larger, contextual considerations actually suggests that creative and routine language may not be all that different. My claim here is that our identifications of some language as being creative or routine does not entail that these reflect different underlying mental processes through which they are created and interpreted.

2 Is Creative Language Deliberate?

Our rough intuitive judgment that some word, or sequence of words, is creative may be associated with the impression that it is both original and effective (Runco and Jaeger 2012). Creative language use is typically seen as the product of individual minds who possess special talents for engaging in different forms of convergent and divergent thinking (Amabile and Pratt 2016; Benedek and Fink 2019). We presume that creative language production was not an accident, but specifically emerges from mostly deliberate attempts to say something that is both novel and useful within some discourse context.

But there are important questions that need to be addressed given this standard view of creativity. First, although originality is undoubtedly relevant to creativity, how does one define what is really original? The possible demonstration that some word or phrase has never been produced before suggests that it may be new. However, random combinations of morphemes or words may be entirely new without necessarily being viewed as creative. Most creative acts recombine, or expand upon, existing knowledge in some form or manner.

Second, there is a long historical trend in psychology to study creativity as primarily a matter of "idea generation". Although considered a central phase of creative thinking, idea generation requires a goal, such as a problem to be solved or an opportunity to be addressed. In fact, ideation not only requires the discovery of a problem or opportunity, but the problem or opportunity must also be clearly identified and defined to prepare for the subsequent generation of ideas and solutions (Runco and Chand 1995). Generating novelty requires having some sense of the problem or opportunity that needs to be addressed.

Let's examine these ideas about context-specific, task-oriented creativity given one instance of creative language production. Consider the word *"rackjack"* that I recently heard for the first time. *"Rackjack"* technically refers to a method

of loading crates onto trucks using wooden ramps, something that few people would likely know of or have used in their own speech. In recent years, though, *"rackjack"* has appeared in talk of when one man is flirting with a woman and then another man comes in and manages to steal the woman away through his own flirtatious behaviors (e.g., "I was talking to a girl last night before some loser rackjacked me and left with her"[1]).

When I first heard *"rackjacked"* in a casual conversation, I had an immediate impression of what it meant, partly given the information about a second man coming in and leaving with the woman originally speaking with the first man, and doing so at "night". The word *"rack"* is a slang expression for women's breasts, particularly larger or more notable breasts, and is presumed to come from the Middle Dutch word "recken" (meaning "stretch out"), which is connected to the Proto-Germanic term *"rakjanan",* which has a similar meaning.[2] To "jack" something typically means, in American English, to "steal" something, which is a shortened use of the word "hijack" (e.g., *"The desperate man hijacked the plane on the way to Miami"*). *"Hijack"* is short for *"highway jacker",* describing a person who robs or seizes vehicles and airplanes when they are in the middle of travel along highways or in the air. *"Rackjack"* is still an unusual term for most English speakers and is believed to have emerged around 2006 on an American TV series.

Would you consider *"rackjack"* to be a case of creative word production? Is it an original and effective term in context? The word itself is not completely novel given its prior use in the context of loading truck beds. But the use of *"rackjack"* in a situation referring to one man stealing away a woman who was just talking with a different man makes the term "novel" because its original, loading trucks, meaning makes little sense in the new context.

The interesting point, though, is that the use of *"rackjack"* in this new context is not a straight-forward metaphorical extension of the earlier use. Stealing a woman away from another man is not related in any obvious way to a method for loading trucks. *"Rackjack"* appears to be original in the way that it borrowed the slang term *"rack"* referring to women's breasts as a metonymy for the entire person (e.g., a part standing for a whole) and combining that with a more familiar term "jack" meaning to steal something.

We may believe that the use of *"rackjack"* in this situation is sexist by referring to a woman by her *"rack"* and making it seem as if she were some man's possession and then, suddenly, hanging out with a different man. *"Rackjack"* may also be viewed, by some observers, as humorous given its alliterative form and

---

1 https://neologisms.rice.edu/index.php?a=index&d=1.

2 https://neologisms.rice.edu/index.php?a=index&d=1.

playful way of speaking about a case of a man stealing another man's woman. The sound of the word *"rackjack"* feels slightly iconic given that the meanings of *"rack"* and *"jack"* are somewhat related to their sounds (i.e., to jack something suggests taking possession of something quickly and forcefully, which is enhanced by the pronunciation of *"jack"*). This added iconicity makes the word especially effective in getting its meaning across in an efficient manner, which is certainly a big part of its utility or effectiveness in communicating a specific, novel idea. *"Rackjack"* in this context may be both original and effective because it is novel and provides a very compact way of describing a complicated social interaction.

Linguists often assume that being creative is cognitively burdensome, and adds time to the online process of formulating one's thoughts and expressing them in language. But it is not clear that this was necessarily the case with *"rackjacked?"* (and many other innovative instances of word play). Part of this depends on how transparent the context is in highlighting the possible meaning of the new word *"rackjacked"*. Putting together *"rack"* (when referring to women's breasts) and *"jack"* (when referring to stealing) is primarily a matter of combining two relatively familiar words that may not be very difficult to do, as experimental studies have shown (Swinney et al. 2007).

Consider now a different example that comes from American English and refers to political campaigns. When someone says *"to astroturf"* something, they intend to communicate the idea that "to engineer a campaign that looks as if it is originating naturally from the general population, though actually it is instigated and organized by large corporations or political parties" (Dancygier and Sweetser 2014: 106). This novel verb phrase is based on the artificial turf known as *"astroturf"*, often found in both indoor and outdoor sports stadiums, where fake, plastic grass is used on the field instead of real grass. The phrase *"to astroturf"* uses *"astroturf"* to metonymically stand for the idea of something that is artificial. By extension, the phrase *"to astroturf"* implies in context the metaphorical idea that the political campaign did not arise organically from the general public, but was artificially manufactured to make it seem as if their ideas originated more genuinely, from everyday citizens.

Once again, there is something novel and effective in the expression *"to astroturf"*, but that general judgment is very much guided by our understanding of what we believe the larger context was for its use, including the presumed communicative goal which motivated its adoption. We may also chose using *"astroturf"* because of the additional meaning effects it may elicit in people. Readers may also possibly infer what *"to astroturf"* means through recognition that the phrase was deliberately created as a metaphor.

Language production is not completely automatic in which speakers talk with little awareness of what they are doing. People are believed, in many situations, to be quite thoughtful when speaking and even produce very specific linguistic expressions with conscious deliberation. One concrete example of this idea is the proposal on "deliberate metaphor theory", which assumes that only a small select group of words or utterances really conveys metaphorical messages, namely those that are composed and delivered with a deliberate aim to alert others to particular cross-domain mappings (e.g., Shakespeare's Sonnet 17 line *"How shall I compare thee to a summer's day"*) (Steen 2008). Most verbal metaphors, under this view, are routine phrases that do not really convey metaphorical messages in which people actively compare one domain to another. Only expressions that are explicitly marked as "deliberately metaphorical" are understood as metaphors.

One possibility is that speakers and writers often explicitly signal their deliberate intent to use metaphor through various pragmatic signals (Goatly 1997; Steen 2008). For example, speakers may signal that they are using metaphor by including different discourse markers (e.g., *"well"*), comparatives (e.g., *"like"*), intensifiers (e.g., *"actually", "quite",* or *"utterly"*), words that indicate specific kinds of meaning (e.g., *"literally", "metaphorically"*), as well as phrases expressing meta-comments on the speaker's communicative intentions (e.g., *"so to speak", "one might say", "a figure of speech"*). These various discourse devices may generally be understood as "pragmatic signals" that act to alert listeners and readers to the special, creative, metaphorical nature of what people say.

The immediate difficulty with the proposal that people sometimes use specific pragmatic signals to alert others to their use of metaphor is that these devices are not at all specific to metaphor (Gibbs 2011). Words and phrases such as *"well", "like",* and *"one might say"* are found throughout spoken discourse and not just restricted to use with metaphor. One study examined a large corpus of language for the presence of so-called "signals" or "tuning devices" for metaphor and found that these are employed with non-metaphorical language 60% of the time (Shutova and Teufel 2010). Corpus analyses also reveal many cases of novel metaphor that are not marked by pragmatic signals and many instances of conventional metaphor that are accompanied by pragmatic signals (Nacey 2013). These observations cast doubt on any one-to-one link between specific pragmatic signals and metaphorically used words or expressions.

Psychological studies are really required to assess whether people signal their attempt to use creative metaphor in discourse. One experiment tested this possibility by examining what people consciously understood when encountering a conventional metaphor with, and without, various pragmatic signals presumed to be markers of deliberate metaphor (Gibbs 2015). Participants read the conven-

tional metaphorical statement in which Mark stated, *"We really have come a long way since the wedding"* in different contexts that explicitly mentioned, or not, different markers of metaphor. They were then asked to give ratings about what they understood the speaker's statement as communicating.

Deliberate metaphor theory assumes that conventional metaphors do not evoke their original cross-domain mappings (e.g., ROMANTIC RELATIONSHIPS ARE PHYSICAL JOURNEYS), But encountering a conventional metaphor with a pragmatic signal should alert listeners to a speaker's deliberate attempt to call attention to some cross-domain mapping. This should, therefore, increase listeners' understanding that the statement *"We really have come a long way since the wedding"* (a) is related to the ROMANTIC RELATIONSHIPS ARE PHYSICAL JOURNEYS, conceptual metaphor, (b) is particularly poetic or creative, (c) expresses the speaker's certainty in what he wanted to say, and (d) is intended to get the listener to think about the topic in a different way.

One-half of the participants viewed discourse contexts ending with a speaker's non-metaphorical reply *"We really are doing much better since the wedding".* This non-metaphorical condition was included to examine whether various pragmatic signals truly work to enhance deliberate metaphor understanding and not just any instance of language, metaphorical or otherwise. Participants were instructed to give their ratings of agreement on a 1–7 scale (1 indicating strong disagreement and 7 indicating strong agreement). The seven statements, presented below, were intended to elicit participants' interpretations of Mark's final utterance:

1. Mark was exactly sure what he wanted to say about his marriage.
2. Mark's statement implied that his marriage was now making more progress than earlier.
3. Mark's statement implied that his marriage was now built on a strong foundation.
4. Mark's statement was creative or poetic.
5. Mark's statement implied that he and his wife moved to a new home during their marriage.
6. Mark's statement was intended to compare his marriage to taking a physical journey.
7. Mark consciously wanted his listener, Larry, to think hard about the meaning of his final statement.

The data from this study did not confirm the main empirical predictions of the deliberate metaphor view, especially its claim that pragmatic signals enhance people's interpretation of cross-domain mappings underlying the meanings of conventional metaphors. Furthermore, there was no evidence to suggest that peo-

ple adopted an expectation to think differently or harder about a particular verbal metaphor given the presence of so-called pragmatic signals of deliberate metaphor. In many cases, people more readily drew cross-domain mappings as part of their understanding of a conventional metaphor alone than when it was accompanied by specific pragmatic signals. These results are consistent with the experimental literature demonstrating that people ordinarily recruit cross-domain mappings (i.e., conceptual metaphors) as part of their fast-acting understanding of metaphorical language, even when those metaphors are conventional (e.g., *"We have really come a long way since the wedding"*) (Gibbs 2017).

More generally, when people presume that they or others have performed some action with deliberative forethought or full awareness, they often mistakenly believe these behaviors are entirely the sole product of conscious mental processes. Experimental psychology has dozens of studies that drive home this important point (Gibbs 2011). We may view some instance of language as creative, and due to deliberate thought processes on the part of those who produced that language, but this assumption is clearly an inaccurate account of how much creative language is actually used and understood.

3 Creativity in Pragmatic Context

Many studies on linguistic creativity analyze words and phrases from various corpora without acknowledging their larger, pragmatic, communicative messages in context. This is readily seen in the use of creative metaphor. Consider what a speaker intends to communicate by the metaphor *"My marriage is an ice box"* in the following conversation:

> *Mary was talking to John about her husband.*
> *"We exchanged marriage vows ten years ago."*
> *"We have been married a long time."*
> *Mary continued,*
> *"We are still hanging in there."*
> *Mary then said,*
> *"My marriage is an ice box."*
>   (metaphorical assertion)

Mary's final statement is a declarative assertion that compares her marriage to an icebox from which a listener presumably draws a variety of inferences, such as that Mary's marriage is unemotional, confining, and perhaps lacking in sex. One reason why metaphors are believed to be challenging to understand is because they typically communicate more varied messages (e.g., the marriage is

confining, unemotional, possibly sexless) than is the case with most literal equivalent expressions (e.g., *"My marriage is confining"*).

Now compare the meaning of *"My marriage is an ice box"* when it is used in a slightly different context:

> *Mary was talking to John about her husband.*
> *"We exchanged marriage vows ten years ago."*
> *"We have been married a long time."*
> *John then asked,*
> *"Are you happy in your marriage?"*
> *Mary then said,*
> *"My marriage is an icebox."*
>   (metaphorical assertion + implicature)

In this situation, Mary's statement about her marriage not only conveys certain information about her marriage, but provides an answer to John's question about whether she is happy in her marriage (e.g., *"No, I am not happy in my marriage"*). Mary's final utterance conveys both a metaphorical assertion and a conversational implicature. Conversational implicatures are pragmatic inferences about what a statement contextually implies based on an initial assessment of what the statement literally said (Grice 1989). The meaning of *"My marriage is an icebox"* appears to convey *more* meanings in this second context (e.g., metaphorical assertion plus implicature) than in the first (e.g., metaphorical assertion) by virtue of the added conversational implicature in the second case, which was set up by John's question (e.g., *"Are you happy in your marriage?"*).

One may believe that people are likely to exert more effort, and more time, to comprehend the novel metaphorical utterance in the second context (metaphorical assertion plus implicature) than in the first (metaphorical assertion). In fact, the results of one study showed that this was not the case (Gibbs 2010). Readers took significantly less time to read *"My marriage is an icebox",* and many similar metaphors, when they also conveyed a conversational implicature than when it only expressed a metaphorical assertion (Gibbs 2010). The take-home message here is that people do not necessarily create a "full" or "complete" understanding of each creative metaphor, such as *"My marriage is an icebox",* where they infer all of its possible metaphorical meanings and only then derive pragmatic implications from this. Our understanding of creative meanings can be short-circuited because of expectations of optimal relevance in the specific pragmatic context at any one moment (Sperber and Wilson 2002). People may not automatically spend much effort to understand the potentially complex meanings of individual creative metaphors. They may only process, in this case, creative metaphorical utter-

ances in ways that are merely "good enough" for the present communicative purposes without deriving all of their possible meanings.

My point here is that a single linguistic expression may seem to be both creative and effective from a distant point of view, but the real cognitive effort needed to produce or understand the utterance differs in various pragmatic contexts. One cannot draw a direct association between what some language appears to be like (e.g., it is a creative metaphor) and the psychological effort needed to produce or interpret that language without consideration of the pragmatic context in which it appears.

A different demonstration of how pragmatic context affects cognitive processing of creative language is seen in a study of what metaphors imply in different discourse situations (Gibbs, Tendahl, and Okonski 2011). When people hear an expression such as *"Lawyers are also sharks"* in discourse, their aim is not to just understand the metaphoric meaning of this phrase, but to understand what pragmatic effect the speaker wishes to communicate by using this metaphor. For instance, in a conversation between two people, one may state a number of negative thoughts about lawyers with a second person supporting this argument by saying *"Lawyers are also sharks".* In this case, the metaphor simply strengthens the existing set of beliefs held by the conversational participants. But in a slightly different situation, one person may say various things about lawyers, to which the second speaker adds some completely new information by uttering *"Lawyers are also sharks".* Finally, in a third situation, one speaker may comment on many positive attributes of lawyers to which the second person responds *"Lawyers are also sharks"* in order to contradict the first speaker. Thus, the same metaphor can achieve, at least, three different pragmatic effects (i.e., strengthening an existing idea, adding new information, and contradicting an existing idea) depending on the context. Not surprisingly, people take more time to comprehend the metaphorical utterance *"Lawyers are also sharks"* in the contradictory situation than in the other two.

One implication of this work is that understanding what any creative word or phrase means is not simply a matter recovering a particular creative linguistic meaning per se, but it also involves understanding what a speaker pragmatically intends to achieve by the word or phrase in that specific discourse context. Most reading time (within psycholinguistics) or brain scanning studies (within cognitive neuroscience), fail to consider these pragmatic effects by focusing exclusively on crude, simple distinctions between "literal" and "figurative meanings" or between "routine" and "creative" language (Gibbs and Colston 2012). As a result, scholars sometimes incorrectly attribute variations in processing time or brain activity to constructions of those meanings as opposed to the different social and pragmatic effects that speakers' utterances often convey in real discourse. My

general argument is that understanding creativity demands more attention to people's pragmatic, context-dependent goals in the very moments when new language is produced (e.g., contradicting a previous speaker).

4 Task Defined Creativity

It would be quite nice to list all the personal/cognitive traits or characteristics that underlie creative thinking and language use. But it is difficult to clearly summarize the extant empirical findings on this relationship because they are inconsistent. This inconsistency in the results on creativity is likely due to the variety of tasks employed in these empirical studies and how creativity itself is assessed (Abraham 2018; Fink et al. 2014; Zhu et al. 2013). Different creative tasks may often lead to different conclusions about the nature of creativity.

For example, studies that examine the relationship between intelligence and creativity sometimes show positive effects, but sometimes do no (Abraham 2018). Studies on the relationship between creativity and metaphor production vary considerably depending on whether an individual is constrained in their linguistic choices when making a metaphor (e.g., being asked to craft a metaphor describing professors as smart), or whether they are relatively free to craft a metaphor of their choosing (Skalicky 2022). Participants with higher fluid intelligence scores, or who exhibited a greater desire to engage in cognitively difficult tasks, produced metaphors rated as more creative, but only when creating unconstrained, novel metaphors (Beaty and Silvia 2013). Moreover, people rate their own statements as being more creative than when others judge these same expressions (Skalicky 2022). Most creativity studies also ask participants to produce creative language in response to particular tasks (e.g., being asked to craft a metaphor describing professors as smart), but with no other social context, such as whom the addressees may be or what specific pragmatic messages about smart professors they wish to communicate.

These additional considerations are not simply factors that need to be included in future studies on creative language production. Instead, any theory of creative language use must always be tied to the specific tasks people are engaged in and who assesses the creativity in participants' verbal responses. More generally, it is difficult to create a specific list of personal characteristics that underlie all creative language use given that creativity is always tied to specific tasks. This last point is critically important. Judgments of creativity are always person and task-dependent. Speakers and listeners may judge the creativity of some new words in metaphor quite differently than external experts (e.g., linguists). We

must always ask ourselves "creativity for whom?" and "creativity in what task?" when determining whether some word or phrase is deemed as creative.

5 The Social Context for Creativity

Most creative language emerges in particular social contexts. For example, my recent research on metaphors for bodily experiences shows tremendous creativity in the ways people talk about their bodies (Gibbs, in press). In many instances, this creativity arises in competitive situations where individuals try to out-do other people. An extensive set of this type of examples comes from a website called "The Museum of Menstruation and Women's Health", which has collected many dozens of novel expressions for, and stories about, menstruation.[3] Among some of the creative expressions listed for menstruation include:

> *"Ax wound"*
> *"Antietam"* (bloodiest battle in the US Civil War)
> *"At high tide"*
> *"Attracting the lesbian vampires"*
> *"Having your pixies"*
> *"The dam has burst"*
> *"The devil's work"*
> *"I'm crying me a bloody river"*
> *"It's hunting season"*
> *"Leak week"*
> *"Monthly monster"*
> *"Plum pudding"*

Can you understand why these words and phrases refer to menstruation? Do they appear to be original and effective? Part of the focus on word play is solving the puzzle of why certain ideas or events (e.g., menstruation) are described in specific linguistic ways (e.g., *"The badger is angry"*). These different phrases for menstruation emerged from open, social understandings about what are otherwise private bodily experiences (e.g., the problem in dealing with menstruation and what that says about women's bodies). Once again, people contributing to this website are also engaged in a competition with others to come up with new and unusual expressions. In many cases, our impression that some phrase may be offensive also may affect people's judgments about the phrase's novelty and effectiveness. This implied competitive situation, and possible offensiveness of

---

3 http://www.mum.org/words.html.

the phrases, constitute specific tasks that surely affected the new phrases which were created and reported on the website.

The menstruation website also contains many examples of people describing the particular circumstances in which new phrases for menstruation were created. Consider several instances of these recollections.

> *"Antietam"*
> *"Hello, I don't know if you are still collecting these, but one I most often use is absent. It has been created by me, but is related to historical circumstances: Antietam: The Battle of Antietam during the Civil War is often considered the most costly single-day battle in terms of life loss on American soil. Whenever I refer to that time as 'Antietam' and somebody raises an eyebrow in confusion, I get that look on and say 'bloodiest battle of the war.'"*

> *"The Badger is Angry"*
> *"This euphemism arises from when I was studying Greek medical thought in regards to female anatomy. Plato is quoted as remarking: 'In the middle of the flanks of women lies the womb, a female viscus, closely resembling an animal . . .' and goes on to talk about how they would burn incense under a woman's parts or have the woman inhale smelling salts to get the internal creature to move this way or that. A diagram of these thoughts that I once saw looked like a badger, and I am also unfortunate to get very bad cramps, so sometimes I say that the badger is very angry."*

> *"Blowing a fuse"*
> *"Blowing a fuse: a term I made up, meaning leaking from a tampon. The string on the tampon reminds me of a fuse, and blowing reminds me of a blow-out on a car with all the air leaked out."*

> *"Expelling my hysteria"*
> *"I like to tell my husband that I am 'expelling my hysteria.' First, because of the etymology of the word hysteria (in relation to the [ancient Greek word for] womb), and also as a reassurance that in a week, after it's over, my moods will be normal again."*

> *"I can't churn the butter today"*
> *"I recently started using, 'I can't churn the butter today' because I found a line from one of my niece's books that had a list of things girls couldn't do when they had their 'time of the month' and one of them was, 'I am not allowed to churn the butter/cream.' My mother and I read this and found it funny, because it was one of a long list of things a girl wasn't allowed to do. I am in my teens. Oddly enough I started my period the day of my sister's wedding, and didn't have the nerve to ask my sister what was happening to me; it took me two days to find out."*

I present these examples to make several points about creative language production. First, many of the metaphors for menstruation, including the most creative ones, seem to have developed to enable people to talk indirectly about a topic that is normally taboo. Being euphemistic is an important social skill when talk-

ing about private body parts and bodily functions that serve as one of many context-specific problems for creative language use.

Second, these excerpts from the menstruation website are full of very idiosyncratic motivations for the novel metaphors submitted to the site (mostly by women). In many cases, the idiosyncratic background knowledge about Antietam, Plato, and Greek words, for example, are shared with others co-present at the time when these metaphors were created. The presence of a specific problem to be dealt with, namely finding alternative ways of speaking about one's menstruation, sets the stage for the recruitment of this somewhat obscure knowledge to be able to find novel ways of speaking about a taboo topic.

Some of the terms of menstruation may seem offensive. But as the above examples demonstrate, the actual contexts from which these novel phrases came into being do not appear to be ones in which speakers specifically aimed to be offensive.

Finally, our judgments as to the creativity of these novel expressions for menstruation are not applied to just single linguistic expressions (e.g., *"The badger is angry"* or *"I can't churn the butter today"*). We also consider the explanations offered for why a particular word or phrase may seem creative, novel, or apt for those individuals involved in creating and using these metaphors. Creativity does not arise from some special, modular part of our inner cognition, but emerges in the service of particular social situations where individuals are attempting to solve some task or problem (e.g., talking about very private bodily functions, yet doing so in a unique, understandable manner).

Consider a final example of the social context for creative word production in a study of word play that occurred in the house of Dr. Samuel Johnson, now a museum in London, dedicated to the man who compiled the first *Dictionary of the English Language.* There was a three-month exhibition entitled "The House of Words", which encouraged visitors to engage in activities that reflected the themes of the museum, namely the creation of a contemporary English dictionary (Patel et al. 2016). Visitors approached a substantial desk in the museum that had a book in which they could write a word, an accompanying definition, and, if they wished, a drawing that illustrated the word.

The visitors produced several hundred entries, many of which may be described as "creative" to one degree or another. Consider several examples. In some cases, visitors transformed a person's name to refer to personal characteristics associated with some person with that name. For instance, the name *"Edgar"* was defined as "a helpful fellow", while *"Ameerah"* was defined as "the correct definition of princess". In another case, the name of British tennis player *"Tim Henman"* was written down and defined as a verb meaning "to fail at the last stage of any endeavour", which was pertinent given that Henman repeatedly

failed to win the Wimbledon tennis championship. A different technique used by visitors was to redefine the meaning of extant words, to attribute new meaning to long-standing and familiar nouns. For instance, *"broccoli"* was defined as "a disgusting abuse of the taste buds".

Many entries played on the situation, or task, in which the visitors found themselves. For example, one word written down was *"Johnsonate",* which added the suffix *"ate"* to the word *"Johnson",* which turned the name of Dr. Johnson into a verb meaning "to invent a new word". It is a word that may only be familiar to those who know of Dr. Johnson, or those who have visited the exhibition. The word *"Jonesonian"* was put down to allude to the idiom *"keeping up with the Joneses".* There were also instances in which a suffix was added to the name of a person to produce a meaning that was associated with that individual's activities. Thus, one visitor added *"Abramofy",* a play on the name of the Russian entrepreneur, Roman Abramovich, who owns the Chelsea Football Club. *"Abramofy"* was defined as "to buy an English football team and make it universally hated by all true born Englishmen". The creation of these new words partly assumed that others may possess relevant knowledge of the people being referred to in the new words.

Aside from adding suffixes to existing words, visitors produced new words from combinations, in some cases highly idiosyncratic combinations, of distinct words. For instance, *"HappyFoodDance"* was defined as "how your upper body and/or butt wiggle when you eat something incredibly delicious". *"MuffinBrain"* was defined as "to be stupid – a muffin instead of a brain". Finally, the expression *"Lapoftheroom"* was defined as "an excuse for not completing this form", which the writer then used in context (e.g., *"I will just do a lap of the room while I think then never return"*).

Finally, visitors made small changes to individual words to create new and distinctive definitions. For example, a *"cycloptimist"* is "one who rides around on a bicycle thinking they will never have an accident" (i.e., a combination of combining *"cyclist"* and *"optimist"*). A *"horrorscope",* which combined *"horror"* and *"scope",* was defined as "reading your astrological forecast in the newspaper, and then having a sense of fear and foreboding at the predicted day ahead".

This in-vitro study of novel word play exposed a variety of different means by which people create new words. Some of these are surely employed by ordinary speakers in everyday conversation. But the participants' task here was unusual in that they were to "create new words for a dictionary", a challenge that many of them likely had not experienced before. Nonetheless, the participants offered their suggestions, as well as different definitions for their new words, given the larger context of engaging in an activity within the home/museum of the fa-

mous Dr. Samuel Johnson. The new words produced were creatively made in the specific context of this environment.

6 Are Routine/Formulaic Sequences Fixed or Frozen?

Linguists often compare creative language with what is routine or formulaic. Formulaic language, for example, is widely defined as, "a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray and Perkins 2000) (e.g., idioms, binomials, compounds). Using formulaic language contrasts with creative productivity, the ability to use the structural system of language (syntax, semantics, morphology, and phonology) in a combinatory way to create and understand novel utterances.

Common sequences of words are often more predictable and presumably easier to produce and interpret than creative phrases. Many psycholinguistic studies indeed show that familiar, routine language can often be understood quite quickly (Carrol and Conklin 2014; Gibbs and Colston 2012). Metaphorical language that is routine, such as idiomatic phrases, are typically processed faster than corresponding literal discourse (e.g., the idiom *"John blew his stack"* is processed faster than the literal paraphrase *"John got very angry"*) (Gibbs 1992). Faster processing of formulaic language may be explained probabilistically given speakers' knowledge about word co-occurrences.

But understanding routine language is more complicated than typically portrayed. To give one example, Siyanova-Chanturia, Conklin, and Schmitt (2011) conducted an eye-movement study with native and nonnative participants comparing reading of idioms (e.g., *"left a bad taste in my mouth"*) and matched control phrases (e.g. *"the bad taste left in his mouth"*) embedded in discourse passages. The eye-movement analyses of first-pass reading time, total reading time, and number of fixations showed that the native-speaking participants processed the idioms significantly faster than the nonformulaic controls. For the nonnative participants, there was no evidence that the idioms were processed any faster than the matched controls. Contrary to the general prediction, the figurative readings seemed to be read more slowly than literal readings. In this way, native and nonnative individuals approach routine language somewhat differently, and it is not the case that people interpret all idioms, for example, in the same manner.

Similar variations are seen in other studies on routine language use. Schmitt, Grandage, and Adolphs (2004) identified a number of different types of formulaic sequences from corpus evidence and embedded them in a spoken dictation task with native and nonnative English speakers. Each burst of dictation was longer than short-term memory could hold (i.e., 20–24 words), so the respondents were not able to repeat a burst from rote memory. This meant they were forced to re-construct what they just heard. An analysis of people's repetitions showed that many formulaic sequences were reproduced intact with no hesitation pauses or transformations. But this was not the case for all of the sequences, as some for-mulas were recalled incorrectly and there were important variations among par-ticipants in their abilities to recall speech formulas without errors. This set of findings suggested that even if some word sequence is deemed formulaic, this does not directly imply that it is represented holistically in our minds, and re-trieved as a frozen chunk, contrary to popular belief. I see this as one of the en-during myths about formulaic or routine language use (i.e., phrasal meanings are directly retrieved as a whole from some phrasal lexicon without the need to do any sort of compositional analysis).

Many idioms, in fact, are, to varying degrees, analyzable and understood through compositional processes, exactly as one would typically predict for crea-tive language use. Participants in one series of studies named targets that were syntactically appropriate or inappropriate completions of semantically unrelated sentence contexts (Peterson et al. 2001). Sentences ended with incomplete idioms (e.g., *"kick the"*) and were biased for either literal (e.g., *"ball"*) or idiomatic (e.g., *"bucket"*) completions. People gave faster responses to syntactically appropriate targets than to inappropriate ones. This finding suggests that people engaged in normal syntactic processing during online idiom processing even for many less analyzable idioms. Various other works demonstrate that idiom understanding depends on some compositional analysis that triggers a figurative interpretation when a specific key word is encountered (Tabossi, Fanari, and Wolf 2009). These data also do not support the claim that idioms are necessarily understood in a holistic manner as long words.

The analyzability of an idiom is really a matter of degree depending on the salience of its individual parts. For instance, many English speakers view the phrase *"fall off the wagon"* (meaning "to start drinking alcohol again") as being less decomposable than *"pop the question"* (meaning "to propose marriage"), be-cause the meaning that *"fall"* contributes to *"fall off the wagon"* is not as salient as the meaning that *"pop"* contributes to *"pop the question".* When speakers judge that the idiom *"let off steam"* (meaning "to release inner anger") is analyzable or decomposable, they are essentially finding some relationship between the compo-nents *"let off"* and *"steam"* with their figurative referents *"release"* and *"anger"*

(Gibbs and Nayak 1989; Moon 1998). It is not surprising that speakers find some relationship between the noun *"steam"* and the concept of anger, because anger is metaphorically understood in terms of heat and internal pressure.

But even nonanalyzable idioms still retain some degree of compositionality. For example, people judge the classic, presumably frozen, phrase *"kick the bucket"* to be more appropriate in a pragmatic context where the person died quickly, as opposed to dying in a longer, protracted manner (Hamblin and Gibbs 1999). This intuition is motivated by people's understanding of the separate semantics for *"kick"*, which alludes to a fast, sudden action. Furthermore, the parts of some idioms are more understandable than others and so their metaphoricity is not evenly spread across an entire phrase (Gibbs 1994; Moon 1998). *"Rock the boat"* is a transparent metaphor, but *"rock"* has an analogous metaphorical meaning, "upset", that is seen apart from idiomatic phrases. Verbs such as *"move, agitate, shake"*, and *"stir"* systematically have meanings to do with physical movement and metaphorical meanings to do with emotional disturbance. Similarly, the metaphor of *"spilling"* in *"spill the beans"* is simpler than that of *"beans"*. It is easier to draw an analogy between the action of spilling something physically and that of revealing a secret (compare *"let slip"* or *"drop"* as in *"drop something into a conversation"* and *"spill one's guts"*) than it is to draw an analogy between *"beans"* and *"secret"*. *"Beans"*, referring to individual pieces of knowledge, seems more metaphorical than *"spill"* and thus the idiom is asymmetrically metaphorical (Nayak and Gibbs 1990).

The different meanings of an idiom's parts may also be shaped by the overall figurative meaning of that phrase. For instance, the word *"spill"* now conventionally means "reveal" from its participation in the idiom phrase *"spill the beans"* (meaning "to reveal a secret"). Many dictionaries now see *"reveal"* as one of the primary senses of *"spill"*. Linguistic interpretation does not always operate in a strict bottom-up manner (going from literal to figurative meaning), but involves top-down processes as well (where figurative meanings shape literal ones) (Geeraerts 1995). This possibility shows how idiom analyzability is not strictly grounded in lexical meaning apart from how words are pragmatically used in idioms. Most importantly, the analyzability of an idiom does not depend on that word string being literally well-formed (Gibbs and Nayak 1989). For instance, *"pop the question"* is literally anomalous but semantically decomposable. All that matters for an idiom to be viewed as decomposable is for its parts to have meanings, either literal or figurative, that contribute independently to the phrase's overall figurative interpretation.

These various linguistic observations and behavioral data on the analyzability of idioms raise questions about whether formulaic expressions are readily stored in a fixed manner in a special phrasal lexicon. Many speech formulas actu-

ally exhibit great syntactic and lexical flexibility, something which is typically associated with productive or creative language use (Parizoska 2022). Consider the following exchange (Gibbs and Colston 2012):

Two friends, Maria and Sven, who have not spoken in a few weeks are having a conversation. The last time they had spoken, Maria learned that Sven's very old pet dog was in poor health. During the present conversation, Maria asks Sven about his dog and Sven replies, "The bucket was kicked".

This short scenario demonstrates that the seemingly fixed idiomatic phrase, *"kick the bucket",* used commonly to mean "die", is in fact a decomposable expression in that one needn't preserve a certain specific linguistic form to be effective in communication. Indeed, one can sometimes replace the entire original form and still effectively use that expression to convey idiomatic meaning. Consider, *"She's going to punt that pail one of these days".* We might consider the fixed expression *"kick the bucket"* is, in fact, merely one pointer to a conceptual structure that can be invoked by a most likely finite, but nevertheless highly variable, number of surface forms. Although *"kick the bucket"* is the most prototypical pointer, it is unclear as to what must be preserved in order to keep this "fixed" form intact.

People also possess tacit knowledge about the underlying embodied, conceptual motivations for many routine phrases. This knowledge is closely tied to the ways people produce and interpret many idioms. For example, people tacitly understand that conventional phrases with roughly similar figurative meanings, such as *"blow your top"* and *"jump down one's throat",* can be motivated by different conceptual metaphors. Nayak and Gibbs (1990) examined this question in a series of studies on people's intuitions about idioms and their relations to conceptual metaphors and their context-sensitive interpretations of idioms. Participants in a first study were quite good at linking idioms (e.g., *"blow your stack"*) with their underlying conceptual metaphors (e.g., ANGER IS HEATED FLUID IN THE BODILY CONTAINER), suggesting that they have tacit beliefs of conceptual metaphors that motivated their understanding of some idioms.

A later study asked people to read short scenarios that were constructed to prime different metaphorical mappings (e.g., ANGER IS HEAT IN A PRESSURIZED CONTAINER or ANGRY BEHAVIOR IS AGGRESSIVE ANIMAL BEHAVIOR). Participants were asked to rate the appropriateness of each idiom for the given scenario. If people access the metaphoric mappings underlying an idiom, they should choose one idiom as more appropriate given their metaphorical understanding of the story context. This is exactly what was found. These findings showed that idioms are not "dead metaphors" as traditionally assumed, because people recruit information about the conceptual metaphors underlying idiomatic phrases to make sense of why these routine expressions convey specific metaphorical mean-

ings. Many routine, formulaic expressions are used and understood via compositional analyses that access conceptual information, an idea that is much more closely associated with creative language (Nayak and Gibbs 1990).

Another reason for thinking about routine speech as being more similar to creative language is seen in the ways idioms are understood in different pragmatic contexts. In one set of studies, people read stories ending with either idiomatic phrases (e.g., *"John blew his stack"*) or literal paraphrases for these statements (e.g., *"John got very angry"*) (Gibbs 1992). But the story contexts differed across experimental conditions. Some explicitly referred to the entailments of the underlying conceptual metaphors for the idioms. For example, the idiom *"John blew his stack"* is motivated by the metaphor ANGER IS HEATED FLUID IN THE BODILY CONTAINER, which expresses several entailments regarding the causation of the anger (internal pressure), the intentionality of the action (unintentional), and the manner in which it is expressed (violently). In contexts which accurately expressed all three of these entailments, people read the idiom *"John blew his stack"* more quickly than when a context violated one of the entailments (e.g., the cause of the anger was not internal pressure). The same set of contexts had no effect on the time it took people to read the literal paraphrases (e.g., *"John got very angry"*), probably because the paraphrases were far more vague in their meaning than are the detailed idiomatic phrases. Idioms, metaphors, and other forms of figurative language often immediately elicit embodied simulation processes as they are being processed (Gibbs 2006). These simulation processes enable speakers to demonstrate, and not just describe, what they are thinking in some specific moment (e.g., I think of John's anger as it exploding under pressure out of his bodily container and so state *"John blew his stack"* rather than *"John got very angry"*) (Gibbs 2023).

These findings suggest that verbal metaphors, even highly familiar routine ones, may be understood with different effort depending on how the pragmatics of the discourse context matched, or mismatched, the complex conceptual meanings of these conventional expressions. Similarly to creative language, people's use and understanding of routine sequences is very much dependent in the specific ways they are pragmatically used in discourse.

Speakers can, at times, consciously create novel versions of familiar, formulas, usually with specific tasks in mind. For example, consider the following proverbs that have been specifically twisted to represent unusual (e.g., ironic, satirical, absurdist) perspectives on life's enduring themes, called "Proverbs from Purgatory" (Schwartz 1995; itself an allusion to Blake's "Proverbs from Hell"):

A bird in the hand makes waste.
It's like killing one bird with two stones.
Two heads are better than none.
A friend in need is worth two in the bush.
A stitch in time is only skin deep.
Too many cooks spoil the child.
I'll have him eating out of my lap.
Let's burn that bridge when we get to it.
A friend in need opens a can of worms.
Don't cross your chickens before they hatch.
He's just a chip of the old tooth.

These twisted proverbs blend together parts of familiar metaphorical expressions to creatively convey new insights on old "pearls of wisdom". Each phrase expresses a satirical view, or the dark side, of common metaphorical themes that play such an important role in shaping people's beliefs and actions. Thus, *"A stitch in time is only skin deep"* provides a rather profound alternative view of the worthy reminder that *"A stitch in time saves nine"* (i.e., even our most conservative actions taken to protect us from future harm may not guarantee that we always remain safe). Notice here, once more, that the new spin on old expressions was done with a very specific creative task in mind: turning proverbs from hell to proverbs in purgatory. However, we don't know exactly when, how, or necessarily if, people may realize that some new expression is a creative instantiation of something quite familiar. Some of these associations may arise after, but not before, people have already understood the creative expression.

7 Conclusions

Creative and routine language may be far more alike than commonly believed. Simply judging some words or multi-word units as being "creative" or "routine/ formulaic" does not imply that they are produced and interpreted according to different psychological processes. It is very tempting to assume that novel language emerges from conscious, deliberate attempts to be creative, just as it is appealing to believe that routine language is typically used and understood automatically, and perhaps holistically.

A closer look at the contexts in which people use both creative and routine speech suggests, nonetheless, that there are many complex factors that determine what words are produced and what they mean in context. All language use is tied to specific tasks that people are engaged in, and these pragmatic contexts must be at the forefront of research on creative and routine language use.

Linguists and psychologists face a difficult challenge when analyzing different linguistic forms. We need to not automatically assume simple connections between words and multi-word sequences with different mental processes. The difficulty here is that when people seem to easily produce or understand routine language, for example, without much deliberate thought, it appears as if these processes are not shaped by many interacting personal, interpersonal, and environmental constraints. To give an analogy, skilled drivers move around in their cars with little conscious thought, unless some problem is encountered. Yet this so-called automatic behavior is really organized by a complex set of cognitive, perceptual, and motor skills, all of which operate again without much conscious awareness (Gibbs 2018). For similar reasons, our intuitions that people mostly produce language in one of two modes, creative and routine, is far too simplistic and fails to acknowledge the dynamic reality of how people really work, including when they speak or write. It is important, if not essential, that we go beyond the study of single words and phrases and dig more into the background contexts that drive both creative and routine language use. This message may seem disappointing to scholars who maintain that creative and routine language reflect very different styles of thought, and are produced and understood by rather different mental processes. The methodological imperative is to look closer at the specific personal and social tasks people engage in when speaking creatively or routinely. When we do this, it becomes clearer that creative and routine language may be far more related than is often believed to be the case.

References

Abraham, Anna. 2018. *The Neuroscience of Creativity*. Cambridge: Cambridge University Press.

Amabile, Teresa & Michael Pratt. 2016. The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior* 36. 157–183.

Beaty, Roger & Paul Silvia. 2013. Metaphorically speaking: Cognitive abilities and the production of figurative language. *Memory & Cognition* 41. 255–267.

Benedek, Mathias & Anna Fink. 2019. Toward a neurocognitive framework of creative cognition: The role of memory, attention, and cognitive control. *Current Opinion in Behavioral Sciences* 27. 116–122.

Carrol, Gareth & Kathy Conklin. 2014. Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition* 17. 784–797.

Dancygier, Barbara & Eve Sweetser. 2014. *Figurative Language*. New York: Cambridge University Press.

Fink, Andreas, Bernhard Weber, Karl Koschutnig, Mathias Benedek, Gernot Reishofer, Franz Ebner, Ilona Papousek & Elisabeth M. Weiss. 2014. Creativity and schizotypy from the neuroscience perspective. *Cognitive, Affective, and Behavioral Neuroscience* 14. 378–387.

Geeraerts, Dirk. 1995. Specialization and reinterpretation in idioms. In Martin Everaert, E-J. van der Linden, Andre Schenk & Rob Schreuder (eds.), *Idioms: Structural and Psychological Perspectives*, 57–74. Hillsdale NJ: Erlbaum.

Gibbs, Raymond. 1992. What do idioms really mean? *Journal of Memory and Language* 31. 485–506.

Gibbs, Raymond. 1994. *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.

Gibbs, Raymond. 2006. *Embodiment and Cognitive Science*. New York: Cambridge University Press.

Gibbs, Raymond. 2010. The wonderful, chaotic, creative, heroic, challenging world of researching and applying metaphor: A celebration of the past and some peeks into the future. In Lynne Cameron (ed.), *Metaphor in the Real World*, 125–153. Amsterdam: Benjamins.

Gibbs, Raymond. 2011. Advancing the debate on deliberate metaphor. *Metaphor and the Social World* 1. 67–69.

Gibbs, Raymond. 2015. Do pragmatic signals affect conventional metaphor understanding? A failed test of deliberate metaphor theory. *Journal of Pragmatics* 90. 77–87.

Gibbs, Raymond. 2017. *Metaphor Wars: Conceptual Metaphor in Human Life*. New York: Cambridge University Press.

Gibbs, Raymond. 2018. Words making love together: Dynamics of metaphoric creativity. In Esme Winter-Froemel & Verena Thaler (eds.), *Cultures and Traditions of Wordplay and Wordplay Research*, 23–46. Berlin: De Gruyter.

Gibbs, Raymond. 2023. What's figurative about figurative language? *Lingua* 287, 103520.

Gibbs, Raymond. In press. *Our Metaphorical Bodies: Why Metaphor May Be Everywhere*. New York: Cambridge University Press.

Gibbs, Raymond & Herbert Colston. 2012. *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.

Gibbs, Raymond & Nandini Nayak. 1989. Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology* 21. 100–138.

Gibbs, Raymond, Markus Tendahl & Lacey Okonski. 2011. Inferring pragmatic messages from metaphor. *Lodz Papers in Pragmatics* 7. 3–28.

Goatly, Andrew. 1997. *The Language of Metaphors*. New York: Routledge.

Grice, Paul. 1989. *Studies in the Ways of Words*. Cambridge, MA: Harvard University Press.

Hamblin, Jennifer & Raymond Gibbs. 1999. Why you can't kick the bucket as you slowly die: Verbs in idiom understanding. *Journal of Psycholinguistic Research* 39. 25–39.

Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Oxford University Press.

Nacey, Susan. 2013. *Metaphors in Learner English.* Amsterdam: John Benjamins.

Nayak, Nandini & Raymond Gibbs. 1990. Conceptual knowledge in the interpretation of idioms. *Journal of Experimental Psychology: General* 119. 315–330.

Parizoska, Jelena. 2022. Idiom modifications: What grammar reveals about conceptual structure. *Lexis: Journal in English Lexicography* 19. https://journals.openedition.org/lexis/6293#entries.

Patel, Menisha, Christian Heath, Paul Luff, Dirk von Lehn & Jason Cleverly. 2016. Playing with words: Creativity and interaction in museums and galleries. *Museum Management and Curatorship* 31. 69–86.

Peterson, Robert, Curt Burgess, Gary Dell & Kathleen Eberhard. 2001. Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27. 1223–1237.

Runco, Mark & Garrett Jaeger. 2012. The standard definition of creativity. *Creativity Research Journal* 24. 92–96.

Runco, Mark & Ivonne Chand. 1995. Cognition and creativity. *Educational Psychology Review* 7. 243–267.

Schmitt, Norbert, Sarah Grandage & Svenja Adolphs. 2004. Are corpus derived recurrent clusters psychologically valid? In Norbert Schmitt (ed.), *Formulaic Sequences: Acquisition, Processing and Use*, 127–151. Amsterdam: Benjamins.

Schwartz, Lloyd. 1995. Proverbs from Purgatory. *Paris Review* 13. 25–26.

Shutova, Ekaterina & Simone Teufel. 2010. Metaphor corpus annotated for source – target domain mappings. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, 3255–3261. Valletta, Malta: European Language Resources Association (ELRA).

Siyanova-Chanturia, Anna, Kathy Conklin & Norbert Schmitt. 2011. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research* 27. 1–22.

Skalicky, Stephen. 2022. Liquid gold down the drain: Measuring perceptions of creativity associated with figurative language and play. *Cognitive Semantics* 8. 79–108.

Sperber, Dan & Deirdre Wilson. 2002. Truthfulness and relevance. *Mind* 111. 583–632.

Steen, Gerard. 2008. The paradox of metaphor: Why we need a three-dimensional model for metaphor. *Metaphor & Symbol* 23. 213–241.

Swinney, David, Tracy Love, Matthew Walenski & Edward Smith. 2007. Conceptual combination during sentence comprehension: Evidence for compositional processes. *Psychological Science* 18. 397–400.

Tabossi, Patrizia, Rachele Fanari & Kinou Wolf. 2009. Why are idioms recognized fast? *Memory & Cognition* 37. 529–540.

Wray, Allison & Michael Perkins. 2000. The functions of formulaic language: An integrated model. *Language and Communication* 20. 1–28.

Zhu, Feifei, Zhang, Qinglin & Qiu, Jiang. 2013. Relating inter-individual differences in verbal creative thinking to cerebral structures: An optimal voxel-based morphometry study. *PLoS One* 8:e79272. doi: 10.1371/journal.pone.0079272.

Vsevolod Kapatsinski

Creativity through Inhibition (of the First Production that Comes to Mind)

Abstract: This paper proposes that, from a speaker-internal mechanistic perspective, creativity results from the suppression of a prepotent production activated by the message and the context in which the speaker is trying to express it. It illustrates a recent proposal for how such suppression might be accomplished by the production system, the Negative Feedback Cycle (Kapatsinski 2022) by a few examples. The Negative Feedback Cycle suppresses the production of forms that are likely to have unintended consequences before they are blurted out. The Negative Feedback Cycle's main function is to avoid overextension of frequent forms. However, as a side effect, it generates avoidance behaviors, such as avoiding the production of forms that are likely to be misunderstood. And, as another, more pleasant side effect, it also generates many creative linguistic behaviors that seed linguistic change, including backformation and circumlocution.

1 Introduction

A contestant stands on the stage of a popular American TV show. He is sweating. Ten thousand dollars are on the line. The question, part of the third-grade curriculum: "What is the singular form of the word *lice*?" The man stammers. The host comments "I didn't even know there is a singular form. I thought they travel in packs." After several minutes of embarrassment, the contestant finally volunteers, uncertainly, the guess *lie* (perhaps, on analogy with *die-dice*). Also guessing are four professional cheerleaders. They are not competing for money. They converge immediately on a different answer, also prescriptively wrong but far more obvious – *lice*. (Video available at https://youtu.be/sGKuNhQ7uRA?t=968).[1]

The present paper asks the following question: given the immediate availability of the form *lice* provided by the host, how does the contestant decide not to

---

[1] As shown by Bybee and Slobin (1982) such overextensions are the most persistent type of error in morphological production (see also Harmon et al., 2023; Hoeffner and McClelland, 1993; Taatgen and Anderson, 2002). They are also common in language change where they are likewise predictable from frequency and semantic similarity (Brochhagen et al., 2023; Bybee and Brewer, 1980; Tiersma, 1982).

use that form to express the related singular meaning (like all four cheerleaders did)? In other words, how does the contestant avoid producing *lice* immediately – since it is clearly the most accessible form in the moment, having been provided to him by the host – and instead continue painstakingly searching for a better answer? That is, how does he avoid overextending the form *lice* to the singular meaning LOUSE? The painstaking search for a less obvious expression of the intended meaning is, I argue, a basic prerequisite for creativity.

The intuition I pursue in this paper is that *lice* is not produced because it is a good cue to a meaning that the speaker does *not* want to express – PLURAL. This intuition is spelled out in a connectionist interactive activation framework for language production (Kapatsinski 2022). I illustrate how this candidate mechanism for suppressing overextension may produce, as a side effect, a number of creative linguistic behaviors.

The proposed framework aims to assume a minimum of linguistic structure. It makes the constructionist assumption that the language system is a network of direct form-meaning mappings (Bybee 2001; Goldberg 1995), where *form* is a surface phonetic representation that mediates between acoustics and meaning in perception and between meaning and articulation in production (Kapatsinski 2021). The proposed framework assumes that there are forms of various sizes, as in interactive activation models (Dell 1985, 1986), though it follows the usage-based tradition in abandoning the assumption that these sizes correspond to traditional linguistic units (Bybee and McClelland 2005; Langacker 1987).

Most importantly, the present framework assumes a connectionist view of the mind, in which the forms and meanings are organized in a network, and processing consists entirely of the spread of activation and inhibition. Under such a view, forms are not either licensed by the language system or lie outside of it – all we have is activation, and activation is a matter of degree.

This has an interesting consequence for the notion of creativity. In an influential recent paper, Sampson (2016), has argued that the simple application of existing constructions (form-meaning mappings) to new input forms (i.e., productivity) is distinct from creativity, or F(ixed)-creativity vs. E(nlarging)-creativity in his terms. (E-)creativity requires *extension* beyond the system, breaking the rules. This distinction presupposes that linguistic generalizations rely on classical categories where an input either is or is not eligible to undergo a particular rule (see also Hoffman 2019).

From a connectionist perspective, forms do not have necessary and sufficient conditions on use. In such a system, forms are activated by distributed semantic / contextual representations. Because these representations are distributed, similar contexts and meanings share cues (activated nodes) – the same nodes participate in representing multiple similar meanings (Hinton et al. 1986). As we shall see,

extension of known forms to new contexts is an inevitable side effect of the distributed nature of these representations for the contexts and cannot be distinguished from following the rules (Bybee and McClelland 2005; see also Suttle and Goldberg 2011, for a related perspective).

Extensions can vary in how similar the original use of a form is to its new use, and in how they are perceived by listeners, but all rely on the same basic mechanism – activation of forms by distributed semantic patterns. From a mechanistic perspective, extensions – no matter how creative-looking – are therefore not true creativity because the producer simply says the first thing that comes to mind in accordance with the normal functioning of the system. Creativity requires following the path less traveled, which I hypothesize requires (conscious or unconscious) reflection on the likely consequences of what one is about to say. The NFC provides a possible implementation for such reflection.

As mentioned earlier, processing consists entirely of the spread of activation and inhibition. Activation and inhibition continue spreading through the network until a contextually-determined deadline, which is influenced by factors like whether one is in danger of losing the floor (Holler et al. 2021) and whether the choice is a consequential one (Nozari and Hepner 2019). Thus, in our original example, longer processing time is allowed by the contestant than by the cheerleaders, because for the contestant thousands of dollars are on the line. This longer processing is, I contend, what allows him to generate a less likely form because it takes time to suppress the most obvious answer and settle on a less obvious one.

In the present framework, the process of suppressing the most obvious answer and settling on a less obvious one is what makes creative production different from routine production – that is, creative productions do go beyond what the language system would normally generate (Sampson 2016), but not because they extend the acceptable range of productions. Rather, it is because they are not the productions that the speaker would most readily generate in the present context. They require time and effort for activation and inhibition to spread beyond the most likely outcome.[2]

Creativity involves producing something new and unexpected, yet appropriate for expressing the intended message (e.g., Heinen and Johnson 2018; Rastelli

---

2 A demonstration that such networks can produce different outputs depending on the length of time activation is allowed to spread is provided by McClelland (1981), reimplemented by Axel Cleeremans at https://axc.ulb.be/pages/interactive-activation-application. However, the architecture of that network is such that spread of activation will not produce more creative solutions: activation spread in that network suppresses retrieval of unusual facts about the gang members so that they resemble more typical members of the gang with more processing.

et al. 2022). Therefore, in this framework, creativity requires 1) activation to spread through the network from the intended message, so that the intended message affects form selection, and 2) some way to avoid producing the most obvious expression of the intended meaning – the first form that comes to mind – given time to come up with an alternative expression. That is, to do something new in a familiar context, one needs to suppress the familiar actions that the context activates most strongly (Harmon and Kapatsinski 2021). In other words, being creative usually requires overcoming the influence of habit (see also Wood and Neal 2007). I note that creativity in this sense is not necessarily conscious – it might be, but it need not be. That is, there is no causal connection between creativity and consciousness – a system could be creative without being conscious.

Given the above, I adopt the following as the operational definition of creativity: *a creative production is less accessible than some other production(s) given the intended message and the context in which it is expressed, in the moment of production*.[3] The lower accessibility of a creative production means that it is activated by the message and context less strongly and therefore takes longer to come to mind than the more accessible alternative(s) (Oldfield and Wingfield 1965). In this sense, the contestant's effortful production of *lie* as the singular form of *lice* is creative, whereas an immediate production of *lice* as the singular form of *lice* when prompted with the plural form *lice* is not.[4] Though both productions are prescriptively wrong, only one required suppressing a more accessible production.

This definition of creativity is intentionally speaker-internal, as our goal is modeling the functioning of the production system. The production may or may not succeed in transmitting the intended message to the audience, and may or may not look creative to the audience or an outside observer. Though many of the creative productions in the speaker-internal sense do look creative to others,

---

3 A creative production in this sense is different from an error. First, the erroneousness of a production is in the eye of the beholder. Therefore, errors can be generated in different ways, only some of them creative. Second, many errors are demonstrably not creative because they are quicker than or as quick as correct productions. This means that they involve producing the form that was most accessible in the moment of production. Others are slow, but only because the speaker is in a state of uncertainty about what to produce – multiple alternatives are activated about equally (Staub 2009). Indeed, it appears unlikely that many errors are creative in the present sense, because an error (by definition) matches the intended message less well than an alternative production. So it is unlikely to arise by the speaker detecting that this alternative production doesn't match the message well enough and suppressing it.

4 The contestant puts himself into a tip-of-the-tongue state, which – unlike the usual tip-of-the-tongue state caused by lack of any dominant response – could have been avoided by producing *lice*. The NFC puts him in this state and gets him out of it.

whether they do or don't makes no difference to whether the speaker suppressed their habitual way of expressing their intended message in producing what they produced. This definition is also not intended to subsume all apparently creative linguistic behaviors. For example, sometimes there is no established way to express a message. In this case, processing may not require the proposed mechanism to suppress a more likely expression, but the result can look or sound very creative to an observer.[5] I also do not wish to imply that creativity in this sense is intentionally creative. Indeed, it is often not. The contestant in our original example was not intentionally trying to be creative – he was intending to be as conventional as possible! However, he was (consciously or unconsciously) intending not to blurt out the most accessible form (*lice*). Behaviors that look creative are often an unintended consequence of this intention.

2 Extension: Apparent Creativity

Some behaviors that look creative to an observer arise out of the habitual functioning of the language production system, and therefore are not creative in our sense. Perhaps the best example of this kind is semantic (over)extension (Brochhagen, Boleda, Galdoni and Xu 2023; Gershkoff-Stowe and Smith 1997; Harmon and Kapatsinski 2017; Naigles and Gelman 1995), which (in the present framework) subsumes morphological paradigm leveling (Bybee and Brewer 1980; Harmon et al. 2023; Hoeffner and McClelland 1993; Tiersma 1982).

For example, the child who says *kitty* when presented with a cow is often not truly being creative. Instead, *kitty* is the form activated most strongly by the semantics of a cow: the child either has not yet learned the word *cow* and therefore has no better-matching word in their vocabulary for referring to a CUTE.BOVINE. ANIMAL, or the better-matching word *cow* is simply less frequent than *kitty* and therefore less accessible despite receiving more activation from the intended message – in the child's experience a CUTE.ANIMAL is usually a kitty (Gershkoff-Stowe and Smith 1997; Harmon and Kapatsinski 2017; Naigles and Gelman 1995). Similarly, extending *lice* to mean one louse would not be truly creative because the much greater frequency of *lice* compared to *louse* means that the message LOUSE.SINGULAR is likely to activate *lice* more than *louse* (Harmon and Kapatsinski 2017; Hoeffner and McClelland 1993).

---

5 This apparent creativity may, perhaps, be based on the listener's simulation of how difficult it would be for them to come up with the same expression for the same meaning. We leave this for future work

Harmon and Kapatsinski (2017) show that such accessibility-driven extensions are not restricted to children and can be elicited experimentally in adults – a form is more likely to be extended to new related meanings if it is frequent in the speaker's prior experience. They further show that there is no preference to extend frequent forms if accessibility differences between frequent and infrequent forms are leveled. This result suggests that extensions result from habit: they are produced because they are more accessible than alternatives, and are therefore accessed before these alternatives come to mind, or (more formally) reach a level of activation needed to be selected for production.

Accessibility-driven extension is illustrated in Figure 1. Here, arrow lengths represent connection strengths, which are proportional to the frequency with which the meaning was expressed by the form in the speaker's experience (see Kapatsinski and Harmon 2017, for a proof that more complex learning algorithms would yield the same result). The dashed line in Figure 1 demonstrates the state of the production system at a point at which the frequent form has been activated by the message (to a level sufficient for production) while the less frequent form has not. An accessibility-driven extension is inevitable at this point, *as long as the speaker starts speaking*.
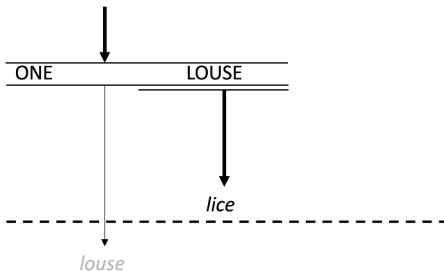
Figure 1: Overextension of *lice* to the meaning ONE LOUSE by a speaker who knows the forms *lice* and *louse* but uses the form *lice* more frequently because lice "usually travel in packs". In this speaker's experience, the form *lice* is much more probable than *louse* given a LOUSE-related message. As a result, *lice* becomes activated more quickly than *louse* (shown by the shorter and thicker arrow) even if the intended meaning activated by the message (rootless top-down arrow at the top of the figure) is ONE LOUSE, and *lice* does not fully match that meaning (matching LOUSE only). The dashed line shows a point in time at which *lice* has already been activated, and *louse* has not yet (which is why it is greyed out). At this point, only *lice* can be produced and overextension therefore appears inevitable.

3 Avoiding Overextension: The Negative Feedback Cycle

To avoid an accessibility-driven extension, production has to be delayed until the less frequent form is activated enough to have a fighting chance against its more frequent competitor. For the speaker to delay production despite having accessed a form, they must sense that there is likely to be something wrong with the form they are about to say, or else they must have an inkling that there is a better option. Otherwise, there is no reason to suppress the most accessible form and delay speaking. Fortunately, speakers can flexibly delay speaking when they have (and sense that they need) time to plan (Holler et al. 2021).

What would make a less frequent form worth waiting for is its superior ability to transmit the intended message to the listener. In fact, speakers who extend frequent forms to new meanings often consider them to be relatively poor expressions of these new meanings. For example, children calling a cow a *kitty* admit that it is not a kitty and would look at a kitty and not a cow when hearing the word *kitty* (Naigles and Gelman 1995). Similarly, Harmon and Kapatsinski's (2017) participants tended to extend frequent forms to new meanings in production but tended to map them onto the experienced meanings in comprehension. When the same form was rare, it was extended to a new meaning less in production, and was mapped onto it more in comprehension. Thus, frequent forms are extended to new meanings because they are more accessible than rarer forms, *even when* the less frequent form would be a better expression of that meaning (see also Koranda, Zettersten and MacDonald 2022, where match to the meaning is quantified objectively).

Even though a rare form may often be worth waiting for, the speaker who already accessed the frequent form and is deciding whether or not to plan more or start speaking (dashed line in Figure 1) has no way of knowing whether a better alternative to the form they have accessed will eventually come to mind. (The contestant does not have the form *lie* accessible when they decide not to say *lice* – it takes him another minute to come up with the form.) Therefore, for the speaker to delay speaking, they must think that there is something wrong with the form they have accessed. That is, the speaker cannot wait to suppress producing the frequent form by comparing how well it expresses the meaning compared to an infrequent form. The suppression must often not be driven by a comparison of alternative expressions but by an evaluation of the dominant expression.

Kapatsinski (2022) proposed a processing mechanism by which a form may be detected to be unsatisfactory before any other form is accessed, allowing the speaker to delay production. This mechanism is illustrated in Figures 2–3. Follow-

ing the connectionist framework, the mechanism assumes that processing works via spreading activation and inhibition.

In Figure 2, the accessed form sends feedback to semantics. Like in comprehension, the amount of feedback reaching a meaning from a form is proportional to how well the form cues the meaning; $p(\text{meaning}|\text{form})$ or $\Delta p = p(\text{meaning}|\text{form}) - p(\text{meaning}|\text{-form})$ depending on learning model (e.g., Gries and Ellis 2015; Kapatsinski 2018; Kapatsinski and Harmon 2017; Ramscar, Dye and Klein 2013). However, unlike in comprehension, this feedback – localized within the production system – is inhibitory: it reduces the activations of meanings that are strongly cued by the activated form.

For meanings that are part of the intended message (LOUSE in Figure 2), this negative feedback makes little difference because they are receiving strong excitatory input from the message. However, any meanings cued by the form that are not part of the intended message now have a negative activation level (i.e., inhibition).
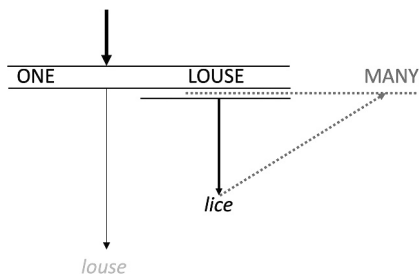
Figure 2: The form *lice* inhibits the meaning(s) it cues (shown by the red dashed arrow). The LOUSE part of the meaning remains activated because it is still receiving excitation from the message. But MANY is now inhibited and has some inhibition to pass on.

The inhibition then spreads from unintended semantics back down to the associated form(s), inhibiting them (Figure 3). Because feedback inhibition cycles back down to the form(s) that generated it, this mechanism is called the Negative Feedback Cycle (NFC). As a result forms that would strongly activate unintended meanings in comprehension are inhibited, and the speaker continues planning (Figure 3). Thus, the NFC allows the speaker to avoid producing frequent forms when they are likely to have unintended consequences.

4 Side Effects of the Negative Feedback Cycle

The primary function of the NFC is to prevent blurting out overextensions, which look creative but – on the production-internal view of creativity – aren't. However, despite its role in enforcing convention, the NFC can occasionally produce behaviors that both look and are creative.
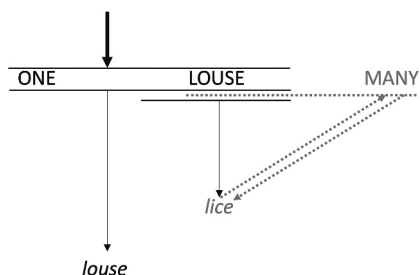
Figure 3: Inhibition then spreads from the unintended meaning MANY back down to the activated form(s) that cued it (*lice*), deactivating or inhibiting them. This both buys the speaker time to activate the initially less accessible form *louse* and ensures that it wins the competition against the initially more accessible form (*lice*).

The NFC suppresses the production activated first, or most strongly by the context when it is estimated by the speaker to be likely to have unintended consequences – to be misinterpreted by the listener as expressing a message that the speaker does not intend. Suppression of this production results in the selection and execution of a production that is less likely given the context. These productions are creative in the production-internal sense – they are not the most expected productions given the context, and require the speaker not to blurt out the first thing that comes to mind. They are also relatively effortful and take some time to produce – inhibition needs time to cycle. They can also look highly creative to an observer once produced. Nonetheless, they too are the product of the normal functioning of the production system.

4.1 Subtraction

The first creative consequence of NFC is deletion of units that express unintended meanings from larger forms, when no smaller form to express the intended message is available. The clearest example of such creative deletions is backformation, which refers to a process by which a speaker generates a new form by deleting what looks like a morph from a pre-existing form. For example, the speakers who first produced *edit* from *editor*, *burgle* from *burglar, destruct* from *destruction*, or *pea* from *peas* engaged in backformation. Currently these productions are of course no longer creative.

Let us look more closely at the case of *editor*. The speaker who created the verb *edit* must have wanted to express the message "perform the job of an editor" or "do what editors do", the act of editing. In the absence of the word *edit*, the closest expression of this message EDIT$_{ACT}$ would be the word *editor*. The speaker had the option of just verbing it: after all, we *author* papers and *engineer* language models rather than *authing* and *enging*. Simple conversion of nouns into

verbs, verbing, is by far the dominant way of forming verbs in English. However, the speaker decided to delete *-or*. Why? Presumably because *-or* has unintended semantics – it is a very good cue to agentivity, ONE.WHO.[. . .]$_{ACTS}$, and this is not part of the intended message.

More formally, we can describe the process as in Figure 4. The speaker's message EDIT$_{ACT}$ first activates the closest matching form, *editor*, as there is no form *edit* yet. At this point, *editor* could have been converted into the verb *editor* and produced to mean EDIT. However, the *-or* is a good cue to ONE.WHO.[. . .]$_{ACTS}$ = AGENT. It therefore inhibits these unintended semantics, which inhibit it in return through the Negative Feedback Cycle. A similar process can account for other instances of backformation in which a morph is deleted, like *peas > pea*, where *-s* would be inhibited by the unintended PLURAL meaning.
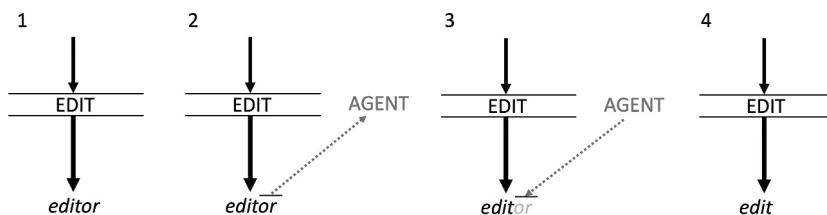
Figure 4: Left: First stage: *editor* activated by EDIT. Middle: Second stage: *-or* inhibits AGENT and is inhibited by it. Right: *-or* is inhibited by inhibition cycling back from AGENT and *edit* is produced.

Notice again that there are verbs like *author* and *engineer*. Backformation is a sporadic process. This sporadicity is actually expected under the NFC account: the NFC needs time to act, and backformations should therefore take time for reflection. After all, four out of five adults asked for the singular form of *lice* in our initial example, quickly wrote *lice* rather than backforming *lie*. Most of the time the speaker does not have enough motivation to wait (production errors do not usually cost thousands of dollars), and may have motivation to start speaking as quickly as possible (Holler et al. 2021). An interesting prediction of the NFC account is therefore that backformations can be distinguished from speech errors by being less, rather than more, likely to occur under time pressure.

4.2 Avoidance and Circumlocution

The strongest evidence for NFC is provided by avoidance behaviors, where the speaker avoids a form that is demonstrably the most likely one given the intended message. Avoidance results in selection of a less likely form, a (possibly

novel) circumlocution, or sometimes nothing at all (i.e., a paradigm gap). In either case, the suppression of the most accessible form is a necessary prerequisite for the emergence of a novel creative solution.

Avoidance is strongest and most successful when the avoided form has taboo connotations. Specifically, Motley et al. (1982) and Dhooge and Hartsuiker (2011) show that speech errors that would result in taboo utterances (like the exchange of initial consonants in *hit shed*) are avoided more successfully than errors that would not result in a taboo utterance (e.g., a similar exchange in *hip shed* would be more likely to be produced). Dhooge and Hartsuiker (2011) further show that speakers take longer to initiate word production when a taboo utterance is likely to result from an error. These results suggest suppression of taboo words like *shit* before they are executed – something the NFC is designed to account for.

Trask (1996) provides several textbook examples of lexical replacement due to the original word becoming tabooed. One well-known example is the replacement of the Proto-Indo-European word bear by $m^j edv^j ed^j$ < *med-o-jed* 'honey eater' in Russian (Fasmer 1986).[6] Another is the avoidance of *lie* in the sense of lying flat (rather than being untruthful) in favor of *lay* as in *I would lay on the couch* (COCA).

The NFC provides an account of these types of replacements, as shown in Figure 5. In the first diachronic stage of the language, *bear* is the normal way to say BEAR. However, when *bear* becomes tabooed (for whatever social reason) it acquires additional connotations. That is, the listener would think "I can't believe he just said *bear*!" (or *shit* or *God* or the name of a dead relative depending on the particular nature of the taboo). The negative nature of the connotation means that it has negative activation by default, and this negative activation can always spread to corresponding forms. In fact, the forms themselves will be likely to have a negative resting activation level as a result.

The negative resting activation can be overcome by excitation coming from the message – i.e., when the connotation is intended. Harry Potter can say *Voldemort*, and Voldemort can say *avada kedavra*, even though both utterances are heavily tabooed for wizards in the universe. Harry and Voldemort say these things because they intend the tabooed consequences (Voldemort intends killing someone by saying *avada kedavra*). However, when a tabooed meaning is unintended, its default negative activation level spreads to the corresponding forms

---

6 Following Fasmer (1986), I assume that the form was originally $m^j ed$-*o-jed* 'honey eater', with the common compound interfix *-o-*, [o] reduced into a glide in this common form, and the form was then reinterpreted as 'honey knower'. However, it could also be assumed, following other etymologies, that *ved* 'know' is the original form.

and makes them particularly easy to inhibit and therefore avoid producing. (If the taboo is strong, one really needs to make an effort to say the tabooed form even when it is intended, keeping the message in mind for longer until the negative resting activation is overcome.)

Returning to our bear example, as *bear* is suppressed, less likely forms can win the competition. *Honey eater* is one of the many possible such forms that can become conventionalized. Its initial production by some past speaker seems undeniably creative, as *bear* would be a more accessible form at the time, and would probably have seemed creative at the time (imagine someone referring to bears as *honey-eaters* in English!). As illustrated in Figure 5, this production likely comes from the semantics of BEAR containing properties like eating honey, which activate associated forms.
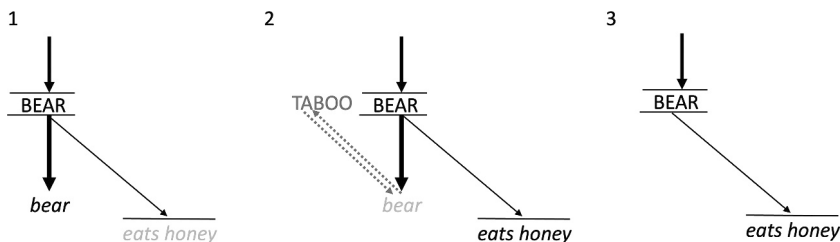
Figure 5: Left: Initially, BEAR strongly activates *bear* and weakly activates semantically similar utterances one could use as circumlocutions like [that creature that] *eats honey*. Middle: *bear* cues the taboo connotation and is therefore inhibited. Right: This leaves *eats* and *honey*, which are then slotted into the common [. . .]$_N$-o-[. . .]$_N$ construction for nouns referring to agents of transitive actions (the unification with the construction not shown; see Dell 1986; Kapatsinski 2017, 2021 for possible mechanisms).

Another instructive example discussed by Trask (1996: 41) is avoidance of forms that resemble names of dead relatives or community members. For example, the death of a community member called *djajila* in 1975 led speakers of the same community to avoid the verb *djäl*, which until then was the most common way to say WANT. The verb was replaced by the hitherto less frequent alternative *duktuk*. This example is interesting because it is not only the form *djajila* that is avoided: forms that are similar enough to *djajila* to activate its meaning are avoided as well. Thus, even though *djäl* means WANT, its production is suppressed because it activates DJAJILA, and the associated memories, enough. The phenomenon is illustrated in Figure 6.

This example supports the NFC thesis that taboo avoidance comes about because the speaker notices, implicitly, that the form they are about to produce would have unintended consequences. In other words, the form selected for pro-
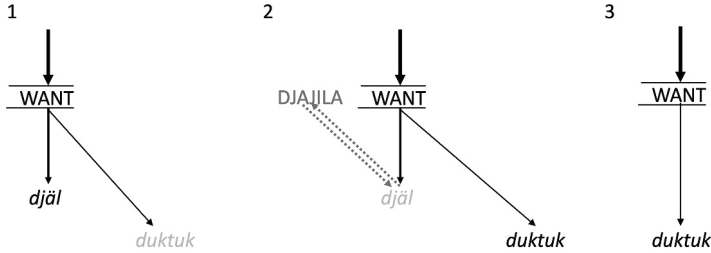
Figure 6: Left: Initially, WANT strongly activates *djäl* and weakly activates *duktuk*. Middle: *djäl* cues the semantics of a dead community member DJAJILA and is therefore inhibited. Right: This leaves *duktuk* to win the competition for WANT.

duction is the ultimate source of inhibition. A speaker who is about to say *djäl* would not have DJAJILA activated as part of their message: WANT and DJAJILA are semantically unrelated, so the message is unlikely to contain, or even activate DJAJILA. The only reason DJAJILA would come to mind is the accidental resemblance between the form *djäl* and the form *djajila*. Access to *djäl* is therefore necessary to activate the taboo semantics, launching the NFC. Furthermore, this example supports the associative nature of the NFC: the NFC suppresses not only forms that *refer* to a taboo meaning, but forms that strongly *cue* a taboo meaning. The form *Djajila* is suppressed most strongly only because it is the best cue to DJAJILA: forms that merely evoke DJAJILA can also be suppressed. In this case, suppression of the regular, run-of-the-mill way of expressing the meaning WANT leads the speaker to produce the less likely *duktuk*, initially a borrowing from another language. This may be akin to an English speaker replacing *want* with *desire*, which would likely be perceived as creative. In *Lord of the Rings*, Celeborn saying *Where is Gandalf, for I much desire to speak to him* is one of the most memorable lines in the story, and works well in conveying the ancient and ethereal character of the elves, in part because of the unusual verb *desire*. However, whether or not it is *perceived* as creative, the use of *duktuk* in place of *djäl* is creative for the speaker – it requires suppressing the usual way of expressing a meaning after it has come to mind enough to activate the taboo semantics.

Another consequence of the NFC is Gresham's Law of Semantic Change – "bad meanings drive out good" (on analogy with the original Gresham's Law, "bad money drives out good", in economics; Burridge 2012; Trask 2003: 45). By providing a mechanistic account of Gresham's Law, the NFC accounts for the common semantic change of pejoration. Specifically, pejoration occurs as a sequence of two changes: extension to a new but related meaning, which just happens to be negative or tabooed, followed by avoidance of the term when the new meaning is not intended. For example, consider the word *intercourse*. The Corpus of Histori-

cal American English (COHA, Davies 2012) shows numerous 19th century examples of its use to mean EXCHANGE or INTERACTION. For example, Jane Austen in *Pride and Prejudice* writes that *Mr. Darcy and Elisabeth had no intercourse but what the commonest civility required*, by which she means that they barely exchanged a word. There are also numerous 19th century bureaucratic documents with titles like *Rules and regulations concerning commercial intercourse with and in states and parts of states declared in insurrection* (from 1864, the American Civil War), where the word means an exchange (of goods). However, around 1890, *sexual intercourse* begins to appear. This of course is a simple extension to a new context (SEXUAL INTERACTION). However, *intercourse* is now a cue to SEX. Since SEX is a taboo meaning, the NFC will now suppress the production of *intercourse* when SEX is not intended (as part of the message; Figure 7). This means that *intercourse* stops being used in non-sexual contexts and therefore strengthens its co-occurrence and association with SEX. As a result, *intercourse* can now mean SEX without the word *sexual*. As *intercourse* can no longer be used to mean INTERACTION, other words for the same concept must take its place – the speaker needs to come up with some solution to this problem. Thus, the word *interaction*, previously rare, is increasingly selected for production when SEX is not intended, and rises in frequency (about 10-fold from 1890 to 1980 in COHA). This would be quite akin to a current speaker avoiding the term *interaction* by referring to a verbal interaction as a *thought-dance,* which I have obtained from ChatGPT by telling it to be creative instead of giving me established synonyms for *interaction* (prompts: *synonyms for interaction* followed by *Can you make a creative one?*). By avoiding the common terms, it is possible to have the same concept generate a less common, or even novel form.
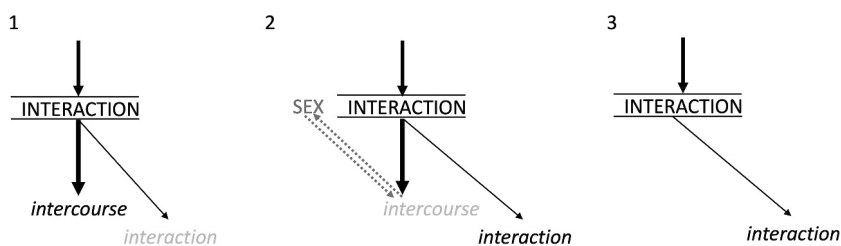
Figure 7: Left: Initially, INTERACTION strongly activates *intercourse* and weakly activates *interaction*. Middle: after 1890, *Intercourse* cues the semantics of SEX is therefore inhibited. Right: This leaves *interaction* to win the competition for non-sexual INTERACTION. Conversely, because *intercourse* now occurs only when SEX is intended, the association between *intercourse* and SEX continues to strengthen.

Not all avoidance behaviors are driven by taboo. Sometimes, the high likelihood of misinterpretation is sufficient. For example, Ceuppens and DeSmet (2024) show that when a form is extended to a new meaning by inference-driven metonymy, the original meaning of the form is often lost, but when the extension is meta-phorical, the original meaning persists. They argue that metonymic extensions tend to result in an ambiguous form whose ambiguity is not resolved by context. Indeed, the existence of bridging contexts in which the form is ambiguous be-tween old and new meanings is what allows inference-driven metonymy to occur (Traugott 1988). For example, in contexts like *It's been disaster after disaster since Clark was elected president*, *since* can be interpreted as both AFTER and BECAUSE, allowing for the metonymic extension of *since* from AFTER to BECAUSE. As a re-sult, one might expect avoidance of *since* in contexts where BECAUSE is not in-tended (and this is observed by Ceuppens and DeSmet). In contrast, metaphoric extensions do not require ambiguous contexts, and tend to result in ambiguity that is resolved by context. For example, one knows whether one means a literal chair or the chair of some organization when the word *chair* is used. As a result, the use of the form in the original meaning is not avoided, allowing that meaning to survive.

Misinterpretation-driven avoidance is controversial in morphology, but is now well documented in phonetics. For example, several researchers have shown that speakers hyperarticulate the phonetic cues that have just been misperceived by the interlocutor (e.g., Buz, Tanenhaus and Jaeger 2016; Schertz 2013; Wedel, Nelson and Sharp 2018). For example, speakers increase average voice onset time (VOT) in *cod* if it the listener misperceived as *God* last time it was said, but not if it was misperceived as *pod*. Importantly, the hyperarticulation in question ap-pears to be based on avoidance of the most ambiguous articulations (here, short, [g]-like VOTs), rather than targeting of the least ambiguous articulations (super-long VOTs). Buz and colleagues find that misperception-driven hyperarticulation results in tucking in of the left tail of the distribution of voiceless VOTs, rather than extension of the right tail or a shift in the central tendency. While there is no space to address this phenomenon fully here, this is what one would expect from the NFC, because what the NFC suppresses most strongly are potential pro-ductions that are most likely to be misperceived (here, productions with short VOT). Indeed, Stern and Shaw's (2023) successful model of misperception-based hyperarticulation implements the same principle, in suggesting that hyperarticu-lation involves inhibiting potential productions in proportion to the frequency with which they realize the unintended phone or message.

The NFC provides a novel perspective on the role of homophony avoidance in the emergence of paradigm gaps. A gap is an environment where creativity is re-quired of the speaker to express the message because all conventional expres-

sions of the message have been suppressed. For example, in Spanish, a famous paradigm gap in the first person singular present is *abuelo* (Figure 8), whose production as the first person singular of the verb *abolir* 'abolish' is avoided. The NFC suggests, contra most other accounts of gaps (Albright 2003; Gorman & Yang 2019; Sims 2015), that it is not an accident that *abuelo* is the word for GRANDFATHER in Spanish. Crucially, the meaning GRANDFATHER is far more frequent than ABOLISH. Therefore, when the message I.ABOLISH activates the form *abuelo*, the form will cue GRANDFATHER much more strongly than it cues I.ABOLISH. As a result, its production is likely to be suppressed. Because there is no other active alternative, a paradigm gap results, as shown in Figure 8.
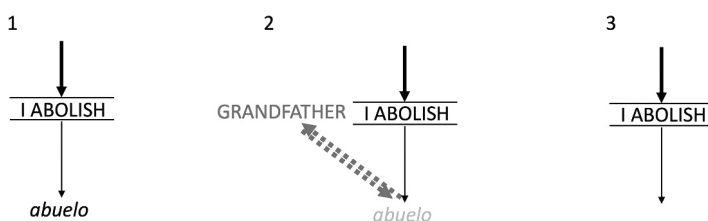
Figure 8: Left: I ABOLISH activates *abuelo*. Middle: *abuelo* strongly cues GRANDFATHER and is therefore suppressed. Right: nothing is left to say.

Many researchers are skeptical of homophony avoidance as an explanation for this gap. One objection is that avoidance occurs in other paradigm cells where there is no complete homophony. Albright (2003: 8) writes "not all parts of the paradigm would be affected by homophony, so even if *abuelo* happens to mean 'grandfather', there would be no reason to avoid the 3pl *abuelen*, which is not a possible noun form". However, complete form overlap is not necessary for avoidance, as exemplified above by the avoidance of *djäl* because it is close enough to activate DJAJILA. It is sufficient for the form to cue an unintended meaning. The sequence *abuel . . .* is surely more than enough to cue GRANDFATHER, which is much more likely than ABOLISH. Far shorter and more ambiguous word-initial chunks have been shown to elicit activation of meanings of the most likely words, and even their semantic associates. For example, in visual world eyetracking studies, listeners look at referents of words that begin with what they have heard so far (Allopenna, Magnuson and Tanenhaus 1998; Teruya and Kapatsinski 2019), and even at their semantic associates (Yee and Sedivy 2006) more than they look at pictures of unrelated meanings. Similarly, Pirog Revill et al. (2008) show, using fMRI, that words with no motion semantics activate the motion area of the brain (MT) if they overlap phonologically with motion words by one syllable. Thus,

abuel is likely enough to activate the semantics of *abuelo* and result in NFC suppressing the activated form.

Another objection is that there are other forms that have homophones and are nonetheless produced (Albright 2003; Gorman and Yang 2019; Halle 1973; Sims 2015). For example, Albright (2003: 8) writes "most importantly, there are many cases in which homophony is tolerated: *creo* 'I create'/'I believe', *avengo* 'I avenge'/'I reconcile', *suelo* 'I am used to'/'I pave', etc". However, none of these cases have the massive imbalance in token frequency between the intended meaning and the unintended meaning that is true of *abuelo*. For example, in the Corpus del Español (corpusdelespanol.org, Davies 2002) there are 79 *abolir* 'abolish' and 1266 *abuelo* 'grandfather', compared to 2894 *creer* vs. 1941 *crear*.[7] The size of the token frequency asymmetry is predicted to be crucial for the avoidance to happen by the NFC: *abuelo* is a much stronger cue to the unintended meaning (GRANDFATHER) than *creo* is.

Consider also the following example from Russian (Figure 9; raised by Halle 1973, and echoed in both Gorman & Yang 2019, and Sims 2015, despite their theoretical disagreements). Russian has gaps in the 1st person singular non-past in verbs of the *-i-* conjugation, in which stem-final coronals become alveopalatals, e.g., [d] becomes [ž]. For example, *deržu* is the expected 1st person singular nonpast of the verb *derzitʲ* 'to speak impudently'. The NFC suggests that *deržu* is avoided when the speaker tries to produce I.SPEAK.IMPUDENTLY, a rare expression, because *deržu* is also the 1st person singular non-past form of a far more frequent verb, *deržatʲ* 'to hold'. According to the Russian National Corpus (available at ruscorpora.ru; see also Grishina 2006), *deržatʲ* is 300 times more frequent than *derzitʲ* (77562 vs. 252; lemma search in the main corpus on 8/27/23). Therefore, if produced, *deržatʲ* would cue the unintended message I.HOLD more strongly than the intended message I.SPEAK.IMPUDENTLY. It would therefore be suppressed by the NFC.

An objection to this reasoning is that ambiguity is tolerated in the form *vožu*, which could mean either I.DRIVE.VEHICLE/I.LEAD.AROUND (*voditʲ*) or I.CARRY.BY.VEHICLE (*vozitʲ*). However, searching the Russian National Corpus reveals that the two verbs are about equally frequent: both are about 10 times the frequency of I.SPEAK.IMPUDENTLY (11328 vs. 9985 respectively). Therefore, inhibition from the unintended meaning would be counterbalanced by activation from the equally frequent intended meaning. The NFC is therefore less likely to succeed in

---

7 The other two examples where ambiguity is supposedly tolerated are not findable: *avenir* has a single token, while no examples of *solir* or *soler* are found.
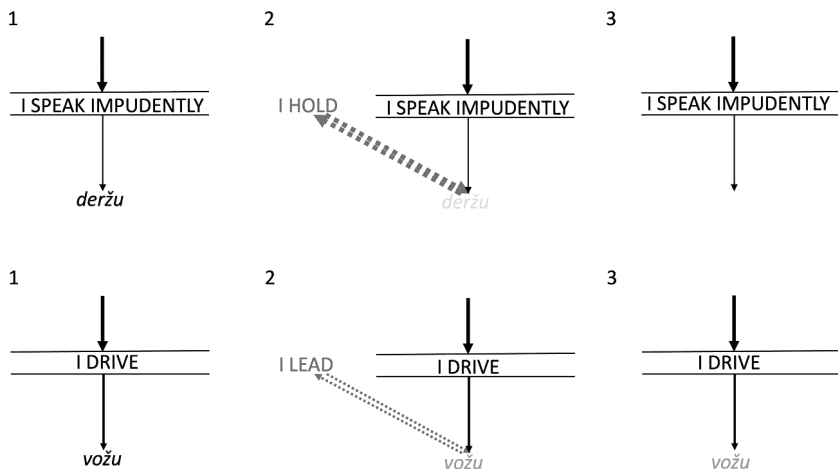
Figure 9: Top: *deržu* strongly cues the unintended message I.HOLD and is suppressed by it. Bottom: *vožu* does not strongly cue the unintended message I.LEAD and survives.

suppressing *vožu* in either sense than *deržu* in the sense of impudent speaking, explaining why only the latter is avoided. This difference is illustrated in Figure 9.

Finally, another objection to ambiguity avoidance as a source of gaps is that there are gaps not explained by this mechanism (e.g., Albright 2003: 8, mentions two verbs in Spanish that have gaps despite absence of homophony). However, we do not expect all gaps to arise for the same reason: there are many reasons that a form might be unacceptable. For example, Russian feminine nouns that have no vowel in the stem have gaps in the plural genitive because deletion of the suffix would leave them with no vowels, e.g., *xna* would become *xn*. Although gaps like this can also be explained by avoidance, it would be avoidance caused by pronunciation difficulty (see also Berg 1998; Martin 2007; Schwartz & Leonard 1982; for other cases of such avoidance). This kind of avoidance could be caused by negative feedback to form selection, but from articulation rather than from semantics (Berg 1998; Martin 2007) or through experience (trying to say [xn] and not liking the consequences; Kapatsinski 2018). Other forms might be gapped because unacceptability becomes associated with certain sublexical chunks through generalization from forms that do have infelicities that cause them to be avoided or stigmatized (Daland, Sims and Pierrehumbert 2007).

Our last example illustrates that avoidance is caused by ambiguity only if one of the meanings is unintended. In some cases, the form is intended to bring to mind another referent. Poetry of course comes to mind – a good poem should have more than one interpretation – but a more prosaic example is presented by

names in societies where children are often named in honor of their parents, other relatives or people the namers admire (e.g., *Albus Severus Potter*). For an accessible example, my name *Vsevolod* was given to be in honor of my great-grandfather and was intended to bring him to mind. Although this is an extension of a form to a new referent, it is a deliberate one, and not entirely accessibility-driven as the name is otherwise rare, and the naming occurred many years after my greatgrandfather has died.

Hypocoristics bring out the interplay of ambiguity avoidance and ambiguity-seeking well. For example, all Russian names have conventional hypocoristics. There are two conventional shortenings for *Vsevolod*, *Seva* and *Vol'a*. The former is more common, and yet, *Vol'a* was selected for me because it matched my great-grandfather, who was intended to come to mind when the name is uttered. In contrast, unintended ambiguity is avoided. For example, one would think that *Volod'a* is a possible shortening of *Vsevolod*. However, it is a conventional short-ening of the much more frequent name *Vladimir*. As a result, its use to refer to *Vsevolod*s is blocked (in fact, I am frequently called *Volod'a* by non-native speak-ers, who overextend the name). (While *Seva* might be argued to be favored over *Volod'a* because it preserves the stressed syllable, this is not true of *Vol'a*.)

As in the case of gaps, ambiguity in hypocoristics appears to be tolerated when frequency asymmetries are smaller: *Slava* could be the shortening to a wide range of names ending in *slav*: *Vladislav, Rostislav, Izjaslav, Svjatoslav*. It is probably not an accident that these names are relatively uncommon: *Vsevolod* vs. *Vladimir* shows a huge frequency asymmetry (2700 vs 32000 in the Russian Na-tional Corpus)[8] while the *slav*s are more closely matched in frequency (1600, 1500, 800, 600, respectively). It is therefore likely that *Slava* will not activate the wrong Slava in context, while *Volod'a* naming a Vsevolod is likely to. The overall lower frequency of the *Slava* names is also relevant: one is likely to know a *Volod'a* who is a Vladimir when naming a Vsevolod but is less likely to know a *Slava* who is a Vladislav when naming a Rostislav. Finally, individuals avoid naming children after people they dislike, and one is more likely to dislike a Vladimir than a Rosti-slav, simply because there are so many more Vladimirs.[9] Therefore, the unin-

---

8 Here and subsequently, Russian frequencies were obtained by using wordform search for the base Nominative Singular form in the main subcorpus of the Russian National Corpus (Savčuk, Arxangelskij, Bonč-Osmolovskaja, Donina, Kuznecova, Ljashevskaja, Orexov and Podrjadčikova 2024) and rounding to the nearest 100.

9 These frequency asymmetries might be somewhat skewed by Putin, who is a *Vladimir*, but should hold nonetheless: Vladimirs are numerous, so there will often be at least one politician named *Vladimir* (e.g., Lenin and Zhirinovskij were Vladimirs as well) causing its frequency to skyrocket in Zipfian fashion.

tended connotations of the former should be much more effective in driving the NFC. The contrast is illustrated in Figure 10.
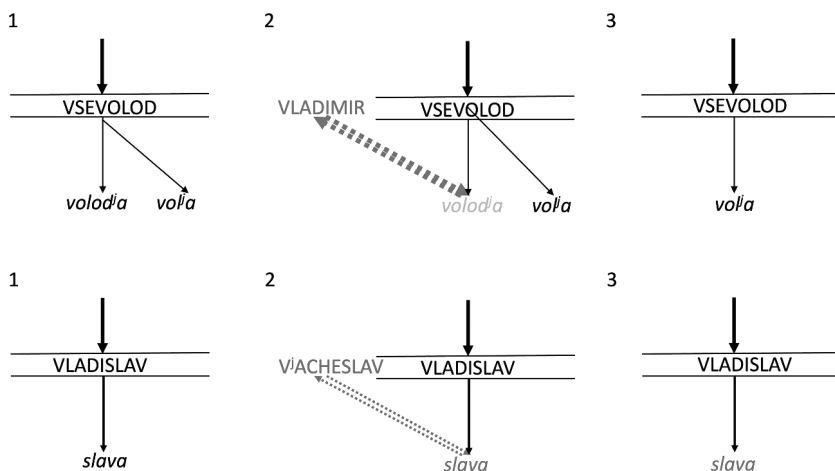
Figure 10: Top: *volodʲa* strongly cues the unintended referent VLADIMIR and is suppressed by it in naming VSEVOLOD. Bottom: *slava* does not strongly cue the unintended referent VʲACHESLAV and survives in naming VLADISLAV.

The need for creativity often arises from ambiguity avoidance in naming when one becomes friends with two people who have the same full names and hypocoristics (a common occurrence). In such cases, people will often deliberately change the hypocoristic of one of the people because it evokes the other, or devise ad hoc nicknames to differentiate them.[10]

---

10 Note that Stankiewicz (1957: 199) proposed that ambiguity is common within gender but avoided between genders. I would tentatively suggest the opposite, taking into account frequency asymmetries: in the case of different-gender names that differ only in the suffix, like *Valentin/ Valentina, Aleksandr/Aleksandra*, ambiguity is largely tolerated; perhaps because the frequency differences are not dramatic but also because the gender of the referent and the agreement markers on verbs and adjectives tend to disambiguate the referent in context. So *Stepanida* can be *Stëpa* despite the greater frequency of *Stepan* (11000/1600). For within-gender names, the pairs in Stankiewicz (1957) are of names of similar frequency mentioned above, and very rare names that are obsolete, making the preferred hypocoristics difficult to verify (*Mitrofan, Valerian*), e.g., were Valerians mostly Valʲas like Valentins or Valeras like Valerijs?

The most problematic example is *Mitja* as the shortening of both *Dmitrij* and *Mitrofan*. *Dmitrij* is much more frequent than *Mitrofan* (15.4K/0.9K) and *Mitʲa* is slightly more common than *Dima* (9.6K *Mitja* vs. 7.2K *Dima*), an alternative shortening for *Dmitrij* stated to be rare in Stankiewicz (1957). Both hypocoristics are also much more frequent than *Mitrofan*. It is possible that *Mitrofan*

5 General Discussion

The Negative Feedback Cycle suppresses productions that are likely to have unintended consequences; more specifically, productions that are likely to transmit unintended meanings. Importantly, suppression is only needed when the risky production is also the most likely production given the context – the first expression that comes to mind. Because of this, the NFC results in productions that are not only safer, but also often perceived as creative – circumlocutions to fill a paradigm gap, backformations, novel or unusual hypochoristics, and generally forms that the speaker would not have chosen to express the same meaning without a moment of reflection.

It appears impossible to avoid postulating something like the NFC if one takes seriously the finding that forms can be produced even when they do not fully match the speaker's intended message – simply because they are more accessible than forms that would express the message better (V. Ferreira & Griffin 2002; Harmon & Kapatsinski 2017; Koranda, Zettersten and MacDonald 2022), and yet speakers have the choice to continue planning and avoid blurting out the first thing that comes to mind. In the extreme context of a talk show, the example we started with, when ten thousand dollars are on the line, the speaker can spend several minutes trying to come up with the right word. To do this, the speaker must suppress production of forms that have already been activated as *not good enough to produce*. The idea of good-enough production was recently emphasized by Goldberg and F. Ferreira (2022) and Koranda et al. (2022). However, good-enough production begs the question of how the speaker would know whether what they are about to say is good enough. The NFC provides the first explicit mechanism by which the speaker could accomplish this goal. The fact that the NFC also accounts for a number of creative behaviors in language production, and makes novel predictions about these behaviors (such as the role of frequency asymmetries in gaps) is a pleasant side effect. Nonetheless, all of the behaviors discussed here demand proper studies that cannot be accomplished here in the available time and space, but represent promising directions for future work.

---

was only ever *Mit^{j}a* (since usually the first syllable is retained in a hypocoristic) and *Dmitrij* also strongly tended to be *Mit^{j}a* at the same time. This would suggest parents of Mitrofans tolerated the ambiguity with *Dmitrij*. However, this is hard to test in the corpus because *Mitrofan* de facto has no hypocoristics: of the first 100 examples of *Mitrofan* in the Russian National Corpus, all are names adopted by adults when they became monks in the Orthodox Church, abandoning their prior secular name, and monks are not referred to with hypocoristics.

The demise of these names whose likely hypocoristics are likely to be misinterpreted may not be an accident. In other words, instead of developing a new creative hypocoristic for a rare name to avoid misinterpretation, one could also avoid the full name, making it even rarer.

In particular, the NFC makes specific predictions about when a form's production is likely to be suppressed, given sufficient time: when the form has a taboo meaning (and that meaning is not intended), when the form has many specific but unintended meanings, and when the unintended meaning(s) of an ambiguous form are frequent relative to the intended meaning. All of these characteristics of a form, importantly, are expected to matter specifically when the NFC has had time to operate: early in processing, the probability of a form's production should depend only on the degree to which it is cued by the message – increasing with the number of semantic features of the form activated by the message times the probability of the form given each feature, and decreasing with the number of the form's semantic features inhibited by the message (Kapatsinski 2022).

The specific characteristics of the NFC are also, of course, up for further investigation and debate (see, e.g., Chuang et al. 2021; Dhooge and Hartsuiker 2011; Hartsuiker and Kolk 2001; Nozari, Dell and Schwartz 2011, for related ideas about how monitoring and suppression might work). The present paper has only scratched the surface of the field by digging up a few illustrative examples. For example, I have assumed that there must be something wrong with the form that a speaker is about to produce for them to reject it – the accessed form inhibits itself because it is a cue to unintended semantics. This is clear in the example of taboo avoidance driven by phonological similarity to the taboo form, in the absence of any semantic similarity.

Alternatively, a careful speaker may delay execution regardless of the appropriateness of what they have planned, and continue pumping activation into the system from the message until all possible alternatives are activated. At this point, the speaker may be able to compare them on how well each alternative production would express their intended message, with the eventually selected form matching the message better because it does not activate any unintended semantics. The NFC is only preferred over this alternative on a priori grounds at present: it has the functional advantage of delaying production only when a delay is needed, and is computationally simpler because it does not require comparison operations. Comparing the meaning activated by each form and the intended meaning would require computing a predicted semantic vector for each form in working memory to compare with the intended semantic vector. However, this advantage in prior probability could be overturned by empirical findings showing that the appropriateness of the original form accessed has no effect on how long a careful speaker takes to plan an utterance. This could be investigated, for example, by priming contextually appropriate vs. inappropriate forms along the lines of V. Ferreira and Griffin (2002).

The NFC proposes that the speaker decides to avoid starting to speak before having accessed an appropriate replacement for the initially accessed form. Alternatively, one could propose that the form eventually produced is what blocks the production of a more frequent, primed, or otherwise accessible form (a mechanism referred to as blocking in Aronoff, 1976, or statistical pre-emption in Boyd and Goldberg, 2011). However, blocking and statistical preemption do not account for the existence of defectivity / paradigm gaps, where the speaker struggles to come up with *any* acceptable production for a while. To return to our initial example of a talk show contestant struggling to produce *lie* as the singular form of *lice*, the contestant does not know what to say for several minutes, but still avoids producing the only form of the word that he does know.

The NFC proposes that the activated form sends *inhibition* up to the semantics it cues. This would, of course, be a non-starter if the feedback took place in the comprehension system. In comprehension, the form *activates* the unintended semantics, rather than inhibiting them. However, there is good evidence that feedback in production is internal to the production system and separate from the comprehension system. In particular, error monitoring (which is function of the NFC) appears to be production-internal because it can be damaged in aphasia independently of comprehension (Hartsuiker and Kolk 2001; Nozari, Dell and Schwartz 2011). The main motivation for the bottom-up inhibition is implementational simplicity – by assuming that the form sends up inhibition, the NFC can be implemented using exclusively spreading inhibition. If the form were sending up excitation, we'd have to somehow turn it into inhibition before it comes back down. However, this is again an empirical hypothesis. The present proposal suggests that unintended meanings associated with the suppressed form should be inhibited, and therefore harder to activate in the immediate future.

The NFC takes time to operate. As a result, when the speaker needs to start speaking quickly (e.g., in a multi-party conversation where other speakers would jump in at any sign of hesitation, Holler et al. 2021), the NFC may not have time to suppress accessibility-driven production choices. Conversely, a writer of a research article like this one – who has nearly unlimited time to plan, and Reviewer 2 to contend with – will often produce and discard multiple possible formulations of the same message because all end up having unintended interpretations, and the consequences of misinterpretation are relatively severe.

One interesting question at the form level is what is suppressed when a writer decides that an abbreviation is not sufficiently unambiguous. For example, in taking notes in the margins of a book, I recently initially wrote down *habit* as an abbreviation for *habituation* but, realizing that I would be likely to misinterpret *habit* as HABIT, continued into *uat*, after a moment's hesitation. Although this is a case in which the producer continues producing, rather than continuing

planning, upon reflection, it would be desirable to account for this phenomenon with the same mechanism as the cases of ambiguity avoidance we have discussed. However, what is being suppressed here? An interesting possibility is that what is suppressed is the action of stopping production (what Diesburg and Wessel 2022 call the "cancel process"), rather than the production *habit*. However, NFC would not be able to suppress it because the action of stopping is not associated with the meaning HABIT. From the NFC perspective, we are forced to assume that what is suppressed is *habit*, allowing the otherwise more costly *habituat* to win. A possible advantage of this account is that it explains why typing was not stopped after *u* or *a*, where the string is as unambiguous as after the second *t*: *habituat* is a chunk (stem) while *habitu* and *habitua* are not.

At the semantic level, one question is whether discrete semantic nodes are needed, or if semantics can be represented as a continuous space (e.g., Chuang et al. 2021). The phrasing of the present paper suggests that semantic representations are composed of discrete unary features like PLURAL or BOVINE. With this representational format, hypernyms are special: *thing*, *stuff*, and *this* may not be effectively suppressed by NFC in producing more specific words because they do not have any unintended semantic features. This may not be desirable because speakers are sometimes dissatisfied with a hypernym and produce a hyponym to it upon reflection (e.g., replacing *dog* with the name of a particular breed). If so, then the absence of a feature could be an unintended consequence, and unary features would be insufficient. Ultimately, points, regions, or distributions over semantic space may also present a better alternative to features (e.g., continuous patterns of activation over a set of hidden nodes as in Rogers and McClelland 2004; or dynamic neural fields, as in Stern and Shaw 2023).

Returning to the nature of creativity, the connectionist framework within which the NFC account is situated provides a different perspective on creativity than traditional symbolic grammar. Specifically, Sampson (2016) has argued that the simple application of existing constructions (form-meaning mappings) to new input forms (i.e., productivity) is distinct from creativity, or F(ixed)-creativity vs. E(nlarging)-creativity in his terms. (E-)creativity requires extension beyond the system, breaking the rules. While this distinction is intuitive, it presupposes that linguistic generalizations rely on classical categories where an input either is or is not eligible to undergo a particular rule (see also Hoffman 2019). From a connectionist perspective, forms do not have necessary and sufficient conditions on use; the selection of a form depends on simultaneous combination of a multitude of contextual and semantic influences (e.g., Kapatsinski 2009). In such a system, extension is an inevitable side effect of the distributed nature of mental representations and cannot be distinguished from following the rules (Bybee & McClelland 2005; see also Suttle and Goldberg 2011, for a related perspective). Extensions can

vary in how similar the original use of a form is to its new use, and in how they are perceived by listeners, but all rely on the same basic mechanism – activation of forms by distributed semantic patterns. From this perspective then, extensions – no matter how creative-looking – are not true (E-)creativity because the producer simply says the first thing that comes to mind in accordance with the normal functioning of the system. Creativity requires following the path less traveled, which we hypothesize requires reflection on the likely consequences of what one is about to say. The NFC provides a possible implementation for such reflection.

Importantly, the NFC's main function is not to produce creative behaviors that would surprise and delight a listener, but rather to avoid otherwise inevitable overextensions, and guard against productions that are likely to have unintended consequences. In other words, the NFC improves the precision of message transmission. That creative productions can then arise is a pleasant side effect.

6 Conclusion

This paper has proposed a production-internal definition of creativity – creativity involves the speaker suppressing the most accessible expression of a message in the moment of production (the first expression that comes to mind) and producing a less accessible expression that still expresses the intended message. This paper detailed how this process can take place within a connectionist framework for language production, and how it can give rise to innovative expressions like backformations and circumlocutions.

Importantly, the first expression that comes to mind tends to be the usual way of expressing the speaker's message in the speech community. Therefore, avoiding it usually leads to a production that is relatively novel and therefore likely to be perceived as creative by the listener. In this framework, all intentionally creative behavior such as writing a poem involves suppressing prepotent, habitual responses to a combination of message and context. However, much linguistic behavior that is not produced with the intention to be creative also involves the same mechanism of suppressing the first production that came to mind, and is therefore creative from a mechanistic, speaker-internal perspective. These behaviors (like backformations) are sometimes perceived to be creative, but the mechanism that produces them is the same whether they are perceived to be creative or not. In contrast, some behaviors that are often perceived to be creative are not creative from the speaker-internal perspective, because they do not involve suppression of a prepotent production.

For example, a child overextending *kitty* to mean COW is not being creative if they don't know the word *cow* and therefore have no alternative expression for the intended meaning (*cow*) to suppress. In contrast, a child who knows the word *cow* and accesses it first when naming the picture of a cow but opts to say *kitty* instead *is* being creative. Mechanistically-speaking, the first child simply allows activation from the message to spread to the associated forms and produces the first form that reaches a sufficient level of activation. The second child does something additional, suppressing the most active form (*cow*) and allowing the activation from the message to select another competitor. Creativity is about reaching the same intended destination by taking a path less traveled.

References

Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. *West Coast Conference on Formal Linguistics 22 Proceedings*. 1–14. Somerville, MA: Cascadilla Press.

Allopenna, Paul D., James S. Magnuson & Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38. 419–439.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Berg, Thomas. 1998. *Linguistic structure and change: An explanation from language processing*. Oxford: Oxford University Press.

Boyd, Jeremy K., & Adele E. Goldberg. 2011. Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language* 87. 55–83.

Brochhagen, Thomas, Gemma Boleda, Eleanora Gualdoni & Yang Xu. 2023. From language development to language evolution: A unified view of human lexical creativity. *Science* 381. 431–436.

Burridge, Kate. 2012. Euphemism and language change: The sixth and seventh ages. *Lexis. Journal in English Lexicology*. 7.

Buz, Esteban, Michael K. Tanenhaus, & T. Florian Jaeger. 2016. Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language* 89. 68–86.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan L. & Mary A. Brewer. 1980. Explanation in morphophonemics: changes in Provençal and Spanish preterite forms. *Lingua* 52. 201–242.

Bybee, Joan, & James L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22. 381–410.

Bybee, Joan, & Dan Slobin. 1982. Why small children cannot change language on their own: Evidence from the English past tense. In A. Alqvist (ed.), *Papers from the 5th International Conference on Historical Linguistics*, 29–37. Amsterdam: John Benjamins.

Ceuppens, Hilke & Hendrik DeSmet. 2024. When does semantic change lead to semantic loss? Metaphor vs. inference-driven metonymy. Paper presented at the 57th Annual Meeting of the Societas Linguistica Europaea, Helsinki, Finland, August 21–24.

Chuang, Yu-Ying, Marie Lenka Vollmer, Elnaz Shafaei-Bajestan, Susanne Gahl, Peter Hendrix & R. Harlald Baayen. 2021. The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods*, 53. 945–976.

Daland, Robert, Andrea D. Sims & Janet Pierrehumbert. 2007. Much ado about nothing: A social network model of Russian paradigmatic gaps. In *Proceedings of the 45th annual meeting of the Association of Computational Linguistics*. 936–943.

Davies, Mark. 2002. Un corpus anotado de 100.000.000 de palabras del Español histórico y moderno. *Procesamiento del Lenguaje Natural* 29.

Davies, Mark. 2012. Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora* 7. 121–157.

Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93. 283–321.

Dell, Gary S. 1985. Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science* 9. 3–23.

Dhooge, Elisah & Robert J. Hartsuiker. 2011. How do speakers resist distraction? Evidence from a taboo picture-word interference task. *Psychological Science* 22. 855–859.

Diesburg, Darcy A., & Jan R. Wessel. 2021. The Pause-then-Cancel model of human action-stopping: Theoretical considerations and empirical evidence. *Neuroscience & Biobehavioral Reviews* 129. 17–34.

Fasmer, M. 1986. *Etymological dictionary of the Russian language*. Moscow: Progress.

Ferreira, Victor S. & Zenzi M. Griffin. 2003. Phonological influences on lexical (mis)selection. *Psychological Science* 14. 86–90.

Gershkoff-Stowe, Lisa & Linda B. Smith. 1997. A curvilinear trend in naming errors as a function of early vocabulary growth. *Cognitive Psychology* 34. 37–71.

Goldberg, Adele E. 1995. *Constructions*. Chicago: University of Chicago Press.

Goldberg, Adele E. & Fernanda Ferreira. 2022. Good-enough language production. *Trends in Cognitive Sciences* 26. 300–311.

Gorman, Kyle & Charles Yang. 2019. When nobody wins. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler, & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 169–193. Cham: Springer.

Gries, Stefan T. & Nick C. Ellis. 2015. Statistical measures for usage-based linguistics. *Language Learning* 65. 228–255.

Grishina, Elena. 2006. Spoken Russian in the Russian National Corpus (RNC). In *Proceedings of LREC*, 121–124. Genoa: International Conference on Language Resources and Evaluation.

Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4. 3–16.

Harmon, Zara & Vsevolod Kapatsinski. 2017. Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology* 98. 22–44.

Harmon, Zara & Vsevolod Kapatsinski. 2021. A theory of repetition and retrieval in language production. *Psychological Review* 128. 1112–1144.

Hartsuiker, Robert J. & Herman H. J. Kolk. 2001. Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology* 42. 113–157.

Heinen, David JP, & Dan R. Johnson. 2018. Semantic distance: An automated measure of creativity that is novel and appropriate. *Psychology of Aesthetics, Creativity, and the* Arts 12. 144.

Hinton, Geoffrey, James L. McClelland, & David E. Rumelhart. 1986. Distributed representations. In David E. Rumelhart, James L. McClelland, & the PDP Research Group (eds.), *Parallel distributed*

processing. Explorations in the microstructure of cognition. Vol. 1: Foundations, 77–109. Cambridge, MA: MIT Press.

Hoeffner, James H. & James L. McClelland. 1993. Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation. In *Proceedings of the 25th Annual Child Language Research Forum*. Stanford, CA: CSLI.

Hoffmann, Thomas. 2019. Language and creativity: A Construction Grammar approach to linguistic creativity. *Linguistics Vanguard* 5. 20190019.

Holler, Judith, Phillip M. Alday, Caitlin Decuyper, Mareike Geiger, Kobin H. Kendrick & Antje S. Meyer. 2021. Competition reduces response times in multiparty conversation. *Frontiers in Psychology* 12. 693124.

Kapatsinski, Vsevolod. 2009. Adversative conjunction choice in Russian (no, da, odnako): Semantic and syntactic influences on lexical selection. *Language Variation and Change* 21. 157–173.

Kapatsinski, Vsevolod. 2017. Copying, the source of creativity. In Anna Makarova, Stephen Dickey & Dagmar Divjak (eds.), *Each venture a new beginning: Studies in honor of Laura A. Janda*, 57–70. Bloomington, IN: Slavica.

Kapatsinski, Vsevolod. 2018. *Changing minds changing tools: From learning theory to language acquisition to language change*. Cambridge, MA: MIT Press.

Kapatsinski, Vsevolod. 2021. What are constructions, and what else is out there? An associationist perspective. *Frontiers in Communication* 5. 575242.

Kapatsinski, Vsevolod. 2022. Morphology in a parallel, distributed, interactive architecture of language production. *Frontiers in Artificial Intelligence* 5. 803259.

Kapatsinski, Vsevolod & Zara Harmon. 2017. A Hebbian account of entrenchment and (over)-extension in language learning. In Glenn Gunzelmann, Andrew Howes, Thora Tenbrink & Eddy Davelaar (eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 2366–2371. Austin: Cognitive Science Society.

Koranda, Mark J., Martin Zettersten & Maryellen C. MacDonald. 2022. Good-enough production: Selecting easier words instead of more accurate ones. *Psychological Science* 33. 1440–1451.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Volume I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Martin, Andrew T. 2007. *The evolving lexicon*. Ph.D. Dissertation, UCLA.

McClelland, James L. 1981. Retrieving general and specific information from stored knowledge of specifics. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *3*, 170–172.

Motley, Michael T., Carl T. Camden & Bernard J. Baars. 1982. Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 21. 578–594.

Naigles, Letitia G. & Susan A. Gelman. 1995. Overextensions in comprehension and production revisited: Preferential-looking in a study of *dog*, *cat*, and *cow*. *Journal of Child Language* 22. 19–46.

Nozari, Nazbanou, Gary S. Dell & Myrna F. Schwartz. 2011. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology* 63. 1–33.

Nozari, Nazbanou & Christopher R. Hepner. 2019. To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology* 36. 193–207.

Oldfield, Richard & Arthur Wingfield. 1965. Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17. 273–281.

Pirog Revill, Kathleen, Richard N. Aslin, Michael K. Tanenhaus & Daphne Bavelier. 2008. Neural correlates of partial lexical activation. *Proceedings of the National Academy of Sciences* 105. 13111–13115.

Ramscar, Michael, Melody Dye & Joseph Klein. 2013. Children value informativity over logic in word learning. *Psychological Science* 24. 1017–1023.

Rastelli, Clara, Antonino Greco, Nicola De Pisapia & Chiara Finocchiaro. 2022. Balancing novelty and appropriateness leads to creative associations in children. *PNAS Nexus* 1. pgac273.

Rogers, Timothy T. & James L. McClelland. 2004. *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, experiment and argument in linguistics and philosophy of language*, 15–26. Bern: Peter Lang.

Savčuk, S. O., Arxangelskij, T. A., Bonč-Osmolovskaja, A. A., Donina, O. V., Kuznecova, Yu. N., Ljashevskaja, O. N., Orexov, B. V. & Podrjadčikova, M. V. 2024. Nacional'nyj korpus russkogo jazyka 2.0: Novye vozmožnosti i perspektivy razvitija. [National corpus of Russian 2.0: New capacities and prospects for development.] *Voprosy Jazykoznanija* 2. 7–34.

Schertz, Jessamyn. 2013. Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics* 41. 249–263.

Schwartz, Richard G. & Lawrence B. Leonard. 1982. Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9. 319–336

Sims, Andrea D. 2015. *Inflectional defectiveness*. Cambridge: Cambridge University Press.

Stankiewicz, Edward. 1957. The expression of affection in Russian proper names. *The Slavic and East European Journal* 1. 196–210.

Staub, Adrian. 2009. On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, 60 (2). 308–327.

Stern, Michael C. & Jason A. Shaw. 2023. Neural inhibition during speech planning contributes to contrastive hyperarticulation. *Journal of Memory and Language* 132. 104443.

Suttle, Laura & Adele E. Goldberg. 2011. The partial productivity of constructions as induction. *Linguistics* 49. 1237–1269.

Teruya, Hideko & Vsevolod Kapatsinski. 2019. Deciding to look: Revisiting the linking hypothesis for spoken word recognition in the visual world. *Language, Cognition and Neuroscience* 34. 861–880.

Tiersma, Peter Meijes. 1982. Local and general markedness. *Language* 58. 832–849.

Trask, Larry. 1996. *Historical linguistics*. London: Arnold.

Traugott, Elizabeth Closs. 1988. Pragmatic strengthening and grammaticalization. In *Annual Meeting of the Berkeley Linguistics Society*. 406–416.

Wedel, Andrew, Noah Nelson & Rebecca Sharp. 2018. The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language* 100. 61–88.

Wood, Wendy, and David T. Neal. 2007. A new look at habits and the habit-goal interface." *Psychological Review* 114. 843–861.

Yee, Eiling & Julie C. Sedivy. 2006. Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32. 1–14.

Hans-Jörg Schmid

What is Creative to Whom and Why? Creativity in Word-Formation and Phraseology Against the Backdrop of Shared Conventions and Individual Routines

Abstract: Creativity can be defined in terms of a balance of originality and effectiveness. I interpret originality in terms of degrees of deviation from convention and effectiveness in terms of degrees of mutual understanding. The challenge for exploring creativity in the domains of word-formation and phraseology is that the schemas and templates that are available in these domains offer considerable scope for regular creativity (i.e. productivity), but also for daring deviations from existing conventions. I resolve this tension by relying on Sampson's (2006) distinction between F-creativity and E-creativity and my own Entrenchment-and-Conventionalization Model (Schmid 2020). I discuss systematic ways in which speakers exploit word-formation and phraseology for acts of linguistic creativity which are salient to different degrees and can be creative vis-à-vis different dimensions of regularity. Essentially, they can pertain to the symbolic dimension of regularity and affect form-meaning pairings, the syntagmatic dimension, affecting combinatorial regularities, and/or the socio-pragmatic dimension, affecting functional, situational and social regularities. The role of factors widening or limiting the scope for creativity, among them individual cognitive routines, is also discussed and included in the proposed definition of linguistic creativity.

1 Introduction

The notion of *creativity* is a slippery customer. According to frequently cited sources, it can be defined in terms of 'originality and effectiveness' (Runco and Jaeger 2012: 92) or 'novelty and appropriateness' (Sternberg and Lubart 1999: 3). Definitions of creativity in linguistics depend heavily on definitions of language and on the linguistic level in focus (see Bergs 2019; Hoffmann 2022; and Tin 2022 for surveys from different perspectives). Transferring the general definitions of creativity to language in a preliminary manner, utterances can be considered 'original' and 'novel' to the extent that they are unconventional; they are 'effective' to the

extent that they facilitate effortless mutual understanding and successful communication; and they are 'appropriate' to the extent that they are suitable for reaching relevant communicative goals and fit the current context.

When looking at creativity in word-formation and phraseology, the domains of linguistics at hand, it pays off to take into account Sampson's (2016) distinction between two types of creativity: *F-creativity* (short for *fixed creativity*) and *E-creativity* (i.e. *extending creativity*). In the present context, F-creativity refers to novel word-formation products or multi-word expressions that are based on existing word-formation patterns or phraseological templates, respectively. This notion largely corresponds to the more traditional notion of *productivity* (Bauer 2001; Plag 1999). E-creativity denotes original formations that go beyond regular applications of productive patterns in one way or other and tend to be salient enough to be recognized as being motivated by the intention to be creative. Anticipating examples to be discussed in more detail below, we can say that the attested but rare form *unrouteable* is F-creative, since it is based on the form of and meanings associated with the productive pattern *un-* [V *-able*]. In contrast, a fabricated form such as *unhappyable* may be considered E-creative, because the use of the adjective *happy* as a basis constitutes a more marked deviation from this pattern. It should be noted right from the start that the boundary between F-creativity and E-creativity is fuzzy. This does not mean, however, that it is pointless to make the distinction. Instead, it will remain useful for theoretical and argumentative purposes to distinguish between these two types of creativity, especially when it is important to highlight differences between them.

The notions of *effectiveness* and *appropriateness* mentioned in the definitions quoted above can be considered to stand for counter-forces that constrain novelty. Only novel utterances that are communicatively successful and considered appropriate count as creative, depending crucially, of course, on the context. Original ways of speaking which cross the limits set by effectiveness and appropriateness are no longer regarded as creative ways of using language, but as unsuccessful or downright wrong (see, e.g., Bergs 2019; Uhrig 2018). How a given utterance is assessed or experienced will depend on numerous other factors, in addition to context, e.g. the genre and register. The linguistic routines as well as the observed linguistic and stylistic competence of the speaker and the hearer will also affect whether an utterance is intended and perceived to be creative. An unconventional expression found in a verse written by a renowned poet is less likely to be viewed as 'unsuccessful' or 'aberrant' than a creative form used by a language learner who is obviously grappling with the grammatical and lexical resources needed to get across what they want to say.

Useful as the definitions just discussed certainly are, they come with a list of follow-up questions related to the notions which serve as criteria for defining creativity. These questions include the following:

1. If creativity is defined by unconventionality, novelty and originality, what is the benchmark or base line for deciding what is or is not conventional, novel or original?
2. What are the systematic ways in which creative utterances can deviate from this base line?
3. How do effectiveness and appropriateness constrain creativity, and what other forces working against creativity can be identified?
4. What are the factors that can loosen the constraints on creativity imposed by effectiveness, appropriateness and other potentially constraining forces?
5. What is the role of the cognitive routines of individual speakers in the production and perception of creative utterances?

In this contribution, I would like to demonstrate how the Entrenchment-and-Conventionalization Model (or EC-Model, see Schmid 2015, 2020) can contribute to answering these questions. This model provides a unified account of how linguistic conventions emerge and are sustained in a community of speakers and how linguistic knowledge emerges and is sustained in the minds of individual speakers. As such, it provides a good starting point for answering question 1), and it will also turn out that it is helpful for answering the other four questions.

The two linguistic domains that are in the focus of this volume and the present contribution can be demarcated as follows. The domain of word-formation encompasses recurrent patterns that can be used to create new lexemes from existing morphological resources: derivation, compounding and conversion, as well as blending, clipping, back-formation and the formation of acronyms. Regarding creativity, it is important to keep in mind that the core morphemic processes are either partly or fully variable patterns, in the sense of construction grammar (see, e.g., Goldberg 2019). This means that in contrast to lexically or morphologically specific items and chunks, they include at least one variable slot (in the case of partly variable patterns) or consist only of variable slots (in the case of fully variable schemas):

- derivations are partly variable patterns, because the prefixes and suffixes used are specific morphological elements (the fixed part) that are combined with different bases (the variable slot; e.g. *un-* + Adj or V + *-able*);
- conversion is a fully variable pattern, because it is not marked by any type of constant morphological substance and therefore does not include any fixed elements (e.g. *hammer*N and *hammer*V);

– compounding varies from fully variable patterns, such as N + N (*flower pot, paper tissue*) or Adj + N (*hotbed, blackbird*), to partly specific ones, e.g. V + *-ing* + N (*swimming pool, walking stick*) or N + V + *-er* (*laser printer, dish washer*).

The domain of phraseology encompasses a wide range of linguistic phenomena sharing the property that several words are combined to build more or less fixed chunks bearing more or less idiomatic meanings. On one pole of these continua, proverbs, commonplaces, routine formulae, idioms and lexical bundles are relatively fixed and lexically 'unique', so to speak. For example, the proverbs *a stitch in time saves nine* and *the early bird catches the worm* are only comparable regarding abstract properties such as their functions and usage contexts, but they do not share any lexically specific material. On the other end, phraseological templates, multi-word verbs, prepositions and conjunctions as well as collocations tend to occur in numerous variant forms and therefore lend themselves to being regarded as productive patterns, due to the variable slot or slots they include. For example, the commonality shared by utterances of the type *as white as a sheet, as free as a bird* or *as light as a feather* can be captured as a partly variable phraseological template of the form *as* Adj *as a* N, which can serve as a productive source for new formations.

It is precisely the tension between complex lexemes and fixed expressions which are established as lexically specific items, partly variable patterns and fully variable patterns that makes the domains of word-formation and phraseology so interesting for the study of linguistic creativity. This tension provides speakers with the opportunity to be creative to different degrees, in a variety of ways and in the pursuit of different goals. One of the main challenges to be tackled here is to do justice to this flexibility and yet manage to come up with a unified concept of creativity.

To enrich the discussion with examples from word-formation and phraseology, I focus on two patterns (but will add other examples where necessary or helpful): the word-formation pattern *un-* + X + *-able*$_{Adj}$ (e.g. *uncomfortable, unpredictable*) and the phraseological template *what the N* (e.g. *what the devil, what the hell*).

2 Two Exemplary Patterns: *un-* + X + *-able*$_{Adj}$ and *what the N*

Adjectives exhibiting the form *un-* + X + *-able*$_{Adj}$ can have as many as five different constituent structures:

1. The frequent binary right-branching verb-based structure *un*-[V-*able*]$_{Adj}$ found in *unavoidable* or *unpredictable*, i.e. *un*-prefixations of adjectives that are derived from verbs by means of the suffix *-able*.
2. The rarer binary left-branching verb-based structure [*un*-V]-*able*$_{Adj}$ as in *unscrewable*, i.e. verbs that are prefixed by reversative *un*- and then transformed into adjectives by means of the suffix *-able*.
3. The binary right-branching noun-based structure *un*-[N-*able*]$_{Adj}$ as in *unknowledgeable* or *unpleasurable*, i.e. negated adjectives that are derived from nouns with the suffix *-able*.
4. The ternary, 'synthetic' structure [*un*- X -able]$_{Adj}$ for forms which are difficult or impossible to align with one of the two possible binary analyses, as in *unskillable* or *unseatable*.
5. A structure derived from a bound root or base of unclear status, as in *untenable*, *unconscionable* or *unviable*.

Regarding the semantic structure of items sharing the form *un*- + X + *-able*, we can distinguish three dominant types (see Bauer, Lieber and Plag et al. 2013: 307–309):
– subject reference, as in *uncomfortable* or *unknowledgeable*,
– object reference, as in *unpredictable* or *unmarriageable*,
– and, more rarely, adverbial reference, as in *unpigeonholeable.*

As far as meanings associated with this pattern are concerned, the literature on prefixation by *un*- and suffixation by *-able* (e.g. Bauer, Lieber and Plag 2013: 291–303, 354–389; Hansen et al. 1985: 85, 112–113; Plag 2004) provides at least the following options (to which the adjectives can often not be allocated unambiguously):
– 'that cannot be V-ed', 'that is unable/unfit to be V-ed', e.g. *unavoidable, undeniable*
– 'that is not worthy of being V-ed', 'that should not be V-ed', e.g. *unacceptable, unlikeable*
– 'that does not V', 'that is unlikely to V', e.g. *unsuitable, unalterable*
– 'that is unable/unfit/unlikely to be subjected to N', e.g. *unmarriageable, unobjectionable*
– 'not having/being characterized by N', e.g. *unknowledgeable, unreasonable*
– 'not being in accordance with N', e.g. *unfashionable, unreasonable*
– 'not being capable of giving/bringing N', e.g. *uncomfortable, unpleasurable.*

To provide data illustrating this pattern from the British National Corpus (BNC), I ran the simple search un+able_AJ0 to extract examples of adjectives beginning with *un*- and ending in *-able* (excluding *unable* by means of the wildcard "+",

Table 1: Extract of the frequency list of items harvested by the search un+able_AJ0 from the BNC.

Rank	Item	Freq.	Prop.	Rank	Item	Freq.	Prop.	Rank	Item	Freq.	Prop.
1	uncomfortable	1327	8.33%	123	unanalysable	7	0.04%	382	unscrewable	1	0.01%
2	unacceptable	1219	7.66%	124	unassimilable	7	0.04%	383	unseatable	1	0.01%
3	unreasonable	978	6.14%	125	unattributable	7	0.04%	384	unseverable	1	0.01%
4	understandable	829	5.21%	126	unforgiveable	7	0.04%	385	unsexable	1	0.01%
5	unstable	697	4.38%	127	unmatchable	7	0.04%	386	unshaveable	1	0.01%
6	unsuitable	690	4.33%	128	unrealisable	7	0.04%	387	unshirkable	1	0.01%
7	unpredictable	673	4.23%	129	unresectable	7	0.04%	388	unsignable	1	0.01%
8	undesirable	617	3.88%	130	unteachable	7	0.04%	389	unsingable	1	0.01%
9	unbelievable	532	3.34%	131	unverifiable	7	0.04%	390	unskillable	1	0.01%
10	unreliable	476	2.99%	132	unclimbable	6	0.04%	391	unslakeable	1	0.01%
11	unmistakable	402	2.52%	133	undrinkable	6	0.04%	392	unsliceable	1	0.01%
12	unavoidable	394	2.47%	134	unmarriageable	6	0.04%	393	unsoftenable	1	0.01%
13	unavailable	377	2.37%	135	unnavigable	6	0.04%	394	unspellable	1	0.01%
14	unthinkable	351	2.2%	136	unpleasurable	6	0.04%	395	unstemmable	1	0.01%
15	unbearable	337	2.12%	137	unputdownable	6	0.04%	396	unstereotypeable	1	0.01%

which requires the variable slot to be filled by at least one letter). Table 1 reports three segments of the frequency list of the 424 types found: the 15 top-ranking items, followed by two sets of 15 items selected opportunistically from the middle range and from the hapax legomena of this pattern in this corpus. Frequencies range from 1,327 attestations for the form *uncomfortable* to a tail of as many as 189 hapaxes.

According to Stangl (2021), the phraseological template *what the N* started out in the late 15th century in the specific form *what the devil*, probably calqued from French *que diable*. Over time, the fixed form became variable in the nominal slot (see Figure 1), probably mainly due to euphemistic strategies for avoiding the taboo word *devil* (and later for dysphemistic strategies of promoting the use of taboo words).

As the figure shows, more and more new variants of the pattern were introduced, but all innovations have fairly strong formal and/or semantic associations to existing variants. For example, *deuce* and *dickens* are euphemistic substitutes for *devil*; their choice is motivated by formal similarity. *Hell* shows a strong semantic connection to *devil* and in turn gives rise to the formally associated substitute *heck*. The latter is a good candidate for serving as the motivation for *fuck* and formally and/or semantically related words derived from it. While Figure 1 lists frequent variants of the pattern, we will see that there are also rare and highly creative modifications. According to the Corpus of Contemporary American (COCA, Davies 2008-), the most frequent manifestations of the pattern are *what the hell, what the fuck* and *what the heck*, boasting normalized frequencies of 32, 14 and 3 occurrences per million words, respectively.

In the next section, I will give a highly condensed survey of the basic ideas of the Entrenchment-and-Conventionalization Model. In the remaining sections of the paper, I will discuss the questions raised above from the perspective of this model and probe its potential for contributing to a better understanding of creativity in word-formation and phraseology.

3 A Survey of the Entrenchment-and-Conventionalization Model

The EC-Model has the following basic design: What it tries to explain is the very nature of language (or grammar) itself, including the questions why languages have structure, why they are subject to variation, and why they keep changing. What it offers as an explanation is a systematic description of the way in which three components interact under the influence of various forces:
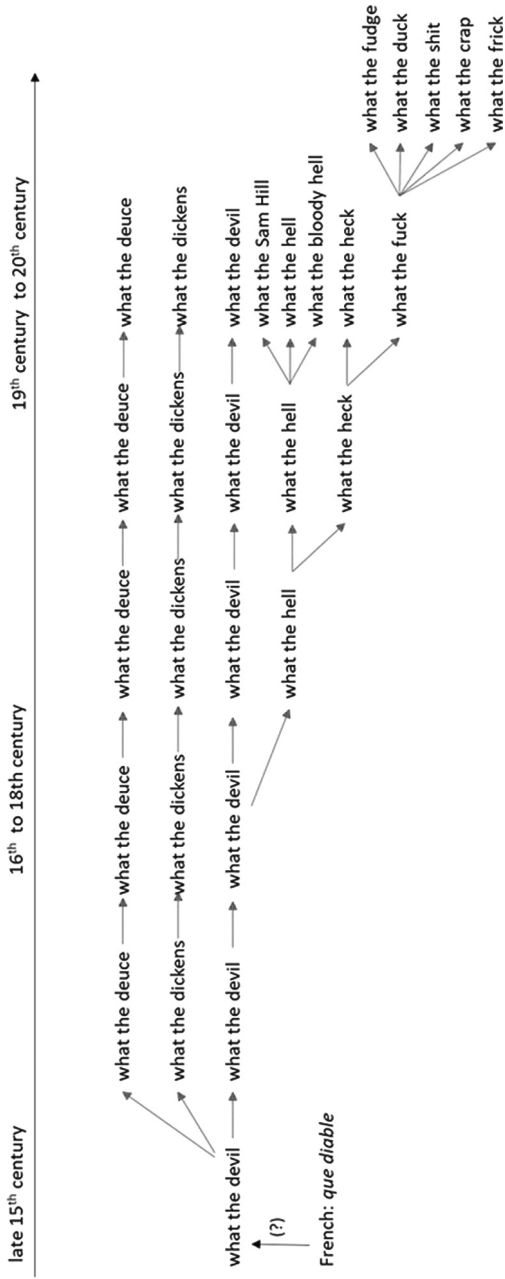
Figure 1: Rough sketch of the historical development of the pattern *what the N* (based on Stangl 2021).

- the use of language in concrete acts of communication,
- social conventionalization processes taking place in communities,
- and cognitive entrenchment processes taking place in the minds of speakers.

Put at its briefest, the model makes predictions regarding the question how usage, conventionalization and entrenchment interact to create and sustain linguistic structure, variation and change and how this interaction is influenced by internal and external forces.

In Schmid (2020), I explain the EC-Model with the help of the graph shown in Figure 2. The 'machine' shown in the figure is inspired by the work of the Swiss artist Jean Tinguely. Overall, it consists of an inner wheel, which is brought into motion by the communicative activities of speakers, and two outer flywheels driven by the inner wheel. The unorderly bundle of black circles in the centre of the figure represents the repetition of similar usage events, i.e. choices of linguistic elements and patterns instantiating conventional utterance types. Such repeated usage events can be token-repetitions of specific words such as *uncomfortable*, but also type-repetitions of a pattern underlying different words such as *unbelievable, unmarriageable* or *unskillable*. Even different tokens representing fully variable patterns such as N+N compounding count as repetitions of similar usage events from this perspective.

Whenever communication takes place, the energy produced in the inner wheel is transmitted to the two outer flywheels, which also begin to spin. This is meant to indicate that each act of communication has effects on the conventions shared by a community (represented on the left) and on patterns of associations in minds of the speakers present (represented on the right). Repetitions of words (*uncomfortable*) and patterns (*un* + V + *able*) will reinforce their conventionality and strengthen corresponding associations. In turn, the conventions shared by a community and the associations entrenched in the minds of speakers feed back into what happens when language is used, since conventions and routinized associations make communication possible in the first place. The whole system has the design of a double feedback loop centered in usage.

For the loop labelled "conventionalization feedback cycle", this means that the repetition of usage events feeds into and is in turn fed by the two conventionalization processes of usualization and diffusion. Conventions are not defined as linguistic units or rules, but instead as more flexible and probabilistic regularities of behaviour which are labelled *utterance types* (see Section 4 for more details).

The processes of usualization and diffusion establish and sustain these regularities in a feedback loop manner. Utterance types such as lexemes, patterns and schemas that are used in concrete acts of communication rely on the conventionality they already bring along due to their prior usage. With every repetition in
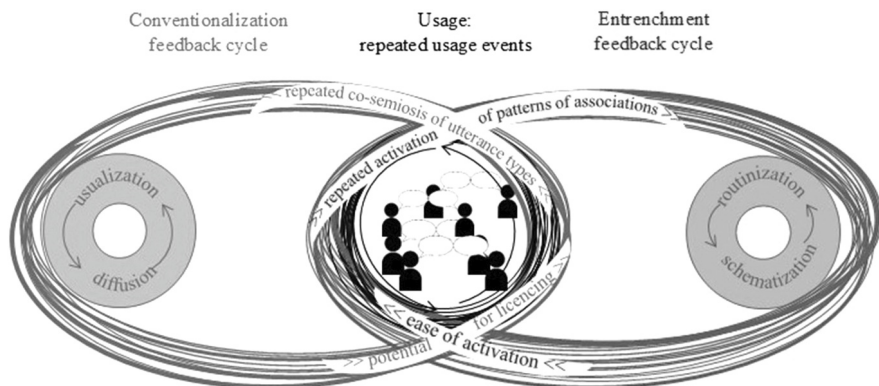
Figure 2: The 'Tinguely machine' model of language (adapted from Schmid 2020: 4).

an act of mutual belief of mutual understanding ("co-semiosis"), their conventionality is confirmed and increased, as their use provides yet another round of refreshment for their conventionality. The more frequently a given utterance type is repeated, the more firmly the underlying regularity becomes conventionalized by the processes of usualization (affecting the degree of regularity among speakers) and diffusion (affecting the spread of the regularity across speakers, groups and contexts). The frequent specific items listed in Table 1, e.g. *uncomfortable, unacceptable* and *unreasonable*, have become more conventional by repeated usage than the rare formations *unsingable, unshirkable* or *unskillable*. The frequent repetition of these and many more adjectives sharing formal and semantic properties sustains the conventionality of the productive semi-variable pattern *un-* + V + *-able*. The more conventionalized a given utterance type has become by past usage, the higher its potential for licensing future usage events. This is because conventional utterance types are more likely to be communicatively effective than unconventional ones. In this way, the conventionalization feedback cycle comes full circle.

The entrenchment feedback cycle works in an analogous way. It operates over patterns of linguistic associations which become active in the minds of those who are present in a usage event. For example, the form *uncomfortable* is symbolically associated with the meanings 'not having the quality of being comforting' and 'not at ease'. It is paradigmatically associated, formally and semantically, with its opposite *comfortable*, among others. Roughly speaking (see Section 8 below for more details), the pattern *un* + V + *able* is represented by a syntagmatic association between the three constituents *un-*, V and *-able* (with a variable slot in the middle) and a symbolic association to the meanings 'that can be V-ed, that has the quality of V-ing' (among other things, see Section 2). Patterns of associations

that are active in usage become routinized and strengthened by repetition, leaving increasingly strong traces and eventually serving as attractors in the associative network, i.e. as routines. The more frequently a given pattern of associations has been activated, the more strongly (or 'deeply') it becomes entrenched (i.e. represented) as a routine. The more strongly entrenched a routine has become, the more easily, effortlessly, automatically and thus also frequently it is activated in future usage events (Langacker 1987: 100; Schmid 2020: 43–72, 205–216).

Conventionalization in a community and entrenchment in the minds of individuals can only 'meet' and be tallied in usage. Conventionalization unfolds under the influence of social processes and forces (e.g. social order, prestige, power), but partly depends on degrees of entrenchment in the minds of the members of the community. Conversely, entrenchment is a cognitive process controlled by psychological factors such as salience, similarity or embodiment, but it is subject to the input supplied by the communities that a given speaker is a member of.[1]

4 What is the Base Line for Deciding what is or is not New or Original?

As mentioned above, the general notion of creativity crucially hinges upon the notions of novelty and originality. To decide whether or not, and to what degree, an utterance is original, we need a base line defining the nature of those utterances that are not original. The EC-Model suggests two candidates to serve as a base line: shared social conventions and individual cognitive routines. Creative utterances can either be considered original because they deviate from the conventions shared by the members of a speech community or because they are not part of the routinized repertoire of a given speaker. For instance, we can regard the word *unsexable* in the right column in Table 1 as creative, because it is not entered in the *OED*, suggesting that it is not a conventionalized solution for a communication task (if we take the dictionary as reflecting what is conventional). Alternatively, I could regard it as being creative because I believe I have never heard it before. Picking shared conventions as a base line has the advantage that we identify originality against the backdrop of language viewed as a socially distributed system of regularities of linguistic behaviour. The downside of taking

---

1 The interaction of the three main components of the model – usage, conventionalization and entrenchment – is subject to a range of forces which are neglected here.

large-scale, community-wide conventionality as a base line is that it is more diffi-
cult to explain that utterance types can be considered as perfectly conventional
by speakers who use them frequently and at the same time as original by speak-
ers who have not come across them.[2] Inter-individual variation of this type can
be explained quite well if one takes routines as a base line for originality, but
then we lose sight of the social dimension of language and have to deal with ram-
pant subjectivity. Therefore, I would suggest to accept conventions as a base line,
but keep routines in view to account for inter-individual and perhaps also social
variation (see Section 8).

In Schmid (2020: 88), I define conventions as mutually known regularities of
behaviour that the members of a community conform to because they expect each
other to conform to them. Note that the concept of *regularity* is understood in a
purely descriptive and empirical way here to identify observable similarities of ut-
terances produced by different speakers. But what exactly is regular about speak-
ers' linguistic behaviour in this sense? This question is usually answered by point-
ing to lexical entries, linguistic rules or formulaic expressions as prime examples of
linguistic conventions (see, e.g., Clark 1996: 76–77). In order to provide a more fine-
grained concept of conventions, I propose a multi-dimensional and probabilistic ac-
count and describe them in terms of five dimensions of regularity:

- *Onomasiological regularities*, i.e. repeated correspondences between commu-
  nicative intentions and linguistic forms available for encoding them; e.g. to
  encode the communicative intention 'express the opposite of a quality', use
  the patterns *un-* + Adj (*unable, unlikely, unusual*) or *in-* + Adj (*independent, in-
  formal, inadequate*) or *non-* + Adj (*non-alcoholic, non-adjustable, non-existent*)
  (examples and semantic classification taken from Hansen et al. 1985).
- *Semasiological regularities*, i.e. repeated correspondences between linguistic
  forms and meanings; e.g. use the pattern *re-*V to encode the meanings 'V-
  again' (*rebuild, reopen, reconstruct*), 'V in a different way' (*reshuffle, reorga-
  nize, retell*), 'V . . . back' (*reimport, retranslate, repurchase*) (examples and se-
  mantic classification taken from Hansen et al. 1985).
- *Syntagmatic regularities*, i.e. regularities regarding the linear arrangement of
  linguistic elements; e.g. the patterns *un-* + V + *-able* and *what the* N, comple-
  mentation patterns (*believe -> that/in, regard -> as, prone -> to*) or collocations
  (*make -> a difference/decision/point, take -> a picture/break/risk*).
- *Contextual regularities*, i.e. repeated correspondences between linguistic ele-
  ments and situational factors, genres, registers, styles, etc.; e.g. the correspon-

---

2 This aspect is integrated in the EC-Model in the way that conventions are defined as being rela-
tive to communities of speakers (see Schmid 2020: 88–90).

dence of the routine formula *good morning* to the situational parameter 'time', the use of *POV* in social media posts or of the formula *I hereby* V in official statements with a declarative illocution.

– *Community-related regularities*, i.e. repeated correspondences between linguistic elements and social groups or variables captured by concepts such as *dialect*, *sociolect* or *genderlect*.

In principle, all five dimensions of regularity can be operationalized in terms of conditional probabilities derived from frequency counts in corpora. The most familiar one of these is the conditional probability in the syntagmatic dimension known as *transitional probability*, e.g. the forward transitional probability of the suffix *-able* given the prefix *un-*. Using frequency counts from the BNC to calculate this probability, we arrive at the value p(*-able*|*un-*) = c(un+able_[AJ0])/c(un*_[AJ0]) = 22,056/147,014 = 0.15. This means that 15% of the adjectives that begin with *un* end in *able* in the BNC, corresponding to a transitional probability of 0.15. The reciprocal backward transitional probability of the occurrence of the prefix *un-* given the suffix *-able* is slightly lower: p(*un-*|*-able*) = c(un+able_[AJ0]/c(+*able*_[AJ0]) = 22,056/170,115 = 0.13. Nevertheless, both types of transitional probabilities indicate strong degrees of syntagmatic regularity within the pattern *un-* + X + *-able*.

Transitional probabilities can also be calculated for each of the specific items found as bases in the variable slot of the pattern. One such measure is in fact listed in Table 1, where we can see that *comfort* accounts for 8.33% of all occurrences of the pattern *un-* + X + $able_{Adj}$, followed by *accept* (7.66%) and *reason* (6.14%). This measure has been called *attraction* (Schmid 2000: 54), since it reflects the degree to which the basis is – metaphorically speaking – attracted by the pattern. A reciprocal measure to be taken into consideration is *reliance*, i.e. the degree to which a word relies on the pattern for its usage. This measure can be calculated by taking into account the overall number of tokens of a given basis in the corpus, both in the pattern and elsewhere, and dividing the frequency in the pattern by this number. The three bases $comfort_V$, $accept_V$ and $reason_V$ behave very differently with regard to this measure: as verbs, *comfort* and *reason* occur 570 and 320 times respectively in the BNC outside the pattern *un-* + V + *-able*, the verb *accept* as many as 9,604 times. Relating this to their frequencies in the pattern as listed in Table 1, this means that *comfort* and *accept* rely on the pattern to 70% and 75% respectively, whereas *accept* does so only for 11% of its usage. Therefore, the degree of syntagmatic regularity for occurrence in the pattern is much higher for *comfort* and *reason* than for *accept*.

Similar measures can be applied in the domain of phraseology. According to the simple search what the N in the COCA, the attraction scores of the top-ranking

forms *hell* and *fuck* are 23% and 10% respectively. More extreme scores can be found for highly routinized patterns. For example, based on the BNC, the transitional probability of *mean* given the sequence *you know what I* is 0.76, reflecting the strong syntagmatic regularity and concomitant predictability of the phrase *you know what I mean*.[3]

In principle, conditional probabilities in the other dimensions of regularity work in an analogous way, but their application is encumbered by major practical obstacles: the frequency counts required for calculating conditional probabilities are much more difficult to come by. Schmid (2024) describes a method referred to as *Multi-Dimensional Regularity Analysis* and shows how it can be applied using simple descriptive statistics or more advanced vector space models. In theory, which is the focus of the present programmatic paper, onomasiological regularity is operationalized as the conditional probabilities of a range of linguistic forms suitable for encoding a given communicative intention. This requires counting communicative intentions – which remains a laborious manual task until we find an automatic way of annotating corpora for intentions. Semasiological regularity is operationalized as the conditional probability of one or more meanings given a linguistic form (see Schmid 2020: 90–92). In the idealized case that a given form has only one meaning, this probability would amount to 1. Perhaps the verb *procrastinate* with the sole meaning 'put off' is a case in point. For highly polysemous lexemes such as *get* or *run*, different meanings (e.g. 'obtain', 'become', 'turn' and many more for *get*) account for different proportions of their usage. According to Glynn's (2014) corpus study of the verb *run*, the meaning 'fast pedestrian motion' has a conditional semasiological probability of 0.32, followed by 'escape' (0.11), 'motion' (0.05) and 'fast motion', 'free motion' as well as 'execute' (0.04 each).

In this framework, the conventionality of the pattern *un-* + V + *-able* is described as a multi-dimensional probabilistic regularity configuration. The syntagmatic, onomasiological and semasiological dimensions of regularity dominate this configuration. Regularity in the syntagmatic dimension relates to the co-occurrence tendencies of the forms *un-*, V and *-able*. The onomasiological and semasiological regularity associated with this syntagmatic regularity was described in Section 2 above, where I listed typical structures and meanings of the pattern. The other two dimensions of regularity are less relevant in this case, because there are no obvious contextual or social regularities to be observed. In contrast, 'learned' suffixation patterns of Romance or Greek origin such as *-(a)tion*$_N$, *-ify*$_V$, *-ate*$_V$, *-ize*$_V$ or *-ism*$_N$ do

---

3 For the sake of transparency, I neglect the more sophisticated and realistic inferential statistical measures that can be used for assessing conditional probabilities, see Evert (2005), Stefanowitsch and Flach (2017: 115–116) and Schmid (2020: 52) for more information.

show contextual and social regularities, on top of syntagmatic, onomasiological and semasiological ones, as they are mainly used in formal genres and contexts (Schmid 2016: 181) and perhaps also more frequently by more highly educated speakers.

The inclusion of contextual and social regularities in the definition of conventions is also helpful for explaining words like *unsexable*, mentioned above and listed with a frequency of 1 in Table 1. Counter to my intuition and despite its not being entered in the *OED*, this word seems to be conventional in communities of practice revolving around fish and is used there to state that the sex of a given specimen cannot be determined.[4] This nicely illustrates that using the cognitive routines of a single speaker as a base line for creativity may be useful, since it helps to highlight differences in people's judgements of what is or is not creative (see Section 8 below).

The five dimensions of regularity chart a five-dimensional space. In theory, this space is populated by utterance types represented as five-dimensional vectors that reflect any type of regularity found in the complete historical record of the language or variety at hand.[5] In order to visualize this space for illustrative purposes (see Figure 3), we can conflate the onomasiological and the semasiological dimension under the label of *symbolic regularity* and the contextual and the social dimensions as *socio-pragmatic regularity*. The syntagmatic dimension can stay the way it is.

Highly regular utterances, i.e. utterances that have a usage history marked by high usage conformity, find their place near the origin (marked as "100% regular"), highly unconventional ones would be located towards the outer edges of the space. If we recall the prediction of the EC-Model that conventional utterance types are more effective than unconventional ones and assume that speakers value effectiveness highly, we can also assume that the historical record of utterance types is not distributed evenly in this space. Instead, we would expect a distribution of the type shown in Figure 3 (using toy, i.e. fictive data), which is intended to show a positive correlation between the density of vectors in the space and the degree of regularity (or conventionality) they indicate. The bubbles cluster in the bottom front left corner, indicating high conventionality, whereas the

---

4 The example in the BNC reads: "The stock consists of: one pair of Ps. lombardi, two Ps. socolofi (which I can't sex), one pair of Labidochromis ewarti, a pair of Labidochromis chizimulu and two Labidochromis caeruleus (unsexable)."

5 It should be noted that all five dimensions are internally multi-dimensional, in the sense that there is not really a single score or value for them, but instead a set of scores reflecting different aspects of regularity (see Schmid 2024 for details). The discussion of the syntagmatic dimension, for example, has provided an idea of the various types of transitional probabilities that are factored in. Likewise, for the social dimension, usage regularities with regard to numerous variables such as age, gender, education, social class and others can be taken into consideration.
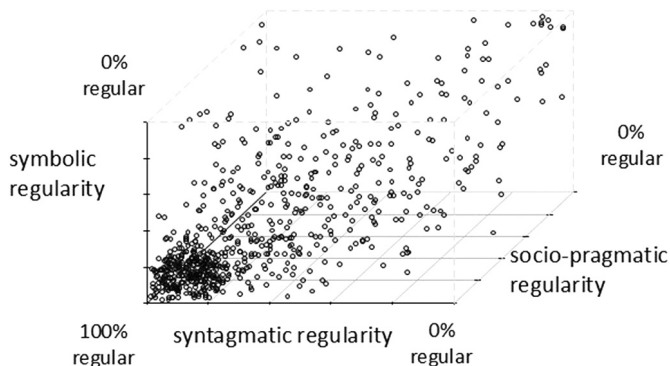
Figure 3: Fictive visualization of points in vector space representing utterance types in three-dimensional regularity space (bubbles stand for differently regular utterance types; the graph is not based on real data, but only serves illustrative purposes).

space towards the edges of the cube and the top right rear corner, containing utterance types marked by a low degree of regularity, is less densely populated. Obviously, it is in these remoter parts of the space that we have to start looking for creative uses of language.

5 What are the Ways in which Creative Utterances can Deviate from this Base Line?

5.1 Creativity in the Symbolic Dimension

Given the definition of conventions in terms of five probabilistic dimensions of regularity, we would expect that creative utterances can be non-conventional and thus original in one or more of these dimensions and to various degrees. Figure 4 visualizes what this means for symbolic (i.e. onomasiological and semasiological dimensions of) regularity. The bubbles located towards the ceiling of the cube representing multi-dimensional conventionality space indicate utterance types showing a low degree of symbolic regularity which remains, however, within the scope of existing usage variants. These utterance types correspond to cases of F-creativity in Sampson's (2016) terms. Cases of E-creativity are defined – in an idealized manner, i.e. taking into account the fuzzy boundary to F-creativity – by literally going 'out of the box', indicated by the four arrows, i.e. by going beyond the limits of what is conventional according to the historical record (see Bergs and Kompa 2020).

unhappyable (invented)
uncomfortable = 'impossible to comfort'
unmarriageable (BNC)

what the Nancy Pelosi (invented)
what the fuckety fuck ... (COCA)
Ø the hell/fuck ... (COCA)

symbolic
E-creativity

unsleepable (BNC)
unhaveable (BNC)

what the fudge ... (COCA)
what the frick ... (COCA)

symbolic
F-creativity

unrhymable (BNC)

what the bloody hell ... (COCA)

100%
regular

syntagmatic regularity

0%
regular
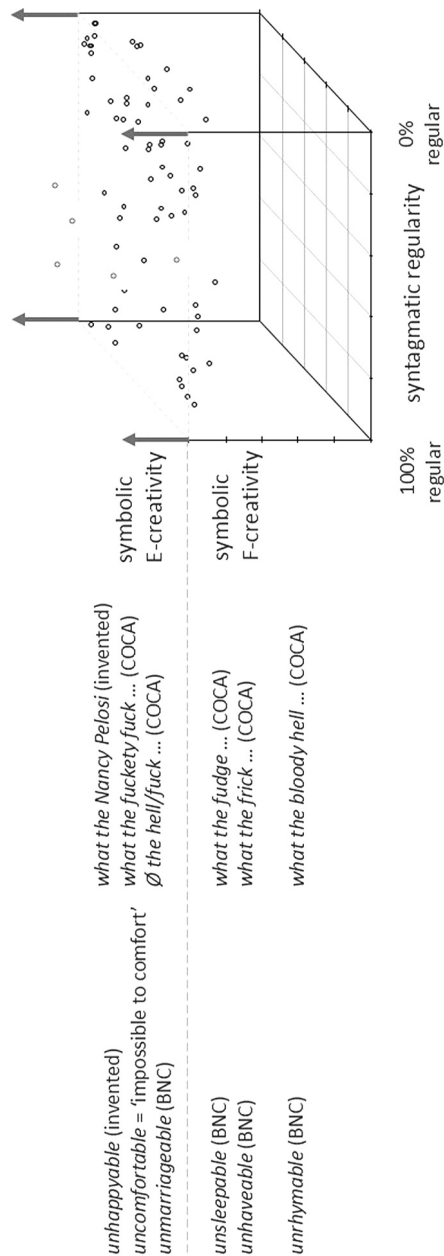
Figure 4: Symbolic F-creativity and E-creativity relative to three-dimensional regularity space (bubbles stand for differently regular utterance types; the graph is not based on real data, but only serves illustrative purposes).

5.1.1 Creativity in the Onomasiological Dimension

Speakers can be creative in the *onomasiological dimension* by selecting a novel linguistic form for encoding a communicative intention or concept (see Šteka-uer 1998, 2005). What will perhaps first come to mind here are those word-formation patterns that are sometimes labelled as *creative,* first and foremost blending. Coining novel blends such as *glampsite* or *hamdog* can be regarded as onomasiologically creative, since the coiner comes up with a new form for ex-pressing an existing or new concept. Another relevant pattern is secretion, i.e. the use of non-morphemic parts of words such as *-gate, -athon, -aholic* or *-erati* as productive suffix-like formatives (Warren 1990; Schmid 2016: 130, 167; see also Szymanek 2005; Callies 2016). These, too, are onomasiologically creative, since they offer new ways of encoding ideas. The same goes for innovative compounds, i.e. compounds that combine existing lexemes in new ways. While established compounds show many regularities that have been described in terms of formal, structural and semantic patterns (see, e.g. Bauer, Lieber and Plag 2013; Schmid 2016; Warren 1978), root-compounding in particular following the patterns N + N and Adj + N is extremely flexible. As a consequence, on the one hand, there is enormous scope for F-creativity, meaning that speakers can create novel com-pounds whose originality ranges from hardly noticed to fairly salient. On the other hand, in order to exploit compounds for E-creativity, speakers have to come up with highly unconventional formations that are difficult to interpret even in context. The study by Ryder (1994) provides interesting invented examples such as *cow tree* or *hamburger shrub*. These were used as stimuli for an interpretation task designed to find out whether participants applied recurrent semantic sche-mas when confronted with novel N+N compounds. The situation is similar for the highly flexible process of conversion (see Clark and Clark 1979).

Regarding the use of derivational patterns such as *un-* + V + *-able*, onomasio-logical creativity can take place in the way that novel fillers are inserted in the variable slot of the pattern. As long as these fillers are in line with the regular characteristics of the pattern, such innovations would be considered as cases of F-creativity. Many of the hapaxes listed on the right of Table 1 can be seen as ex-amples, e.g. *unrhymable, unscalable, unscrewable* or *unshavable*. Formations that deviate from the specifications more strongly and perhaps in a more salient way mark the gradual transition to E-creativity. The forms *unsleepable, unliveable* and *unhaveable* come to mind as examples, since they violate the regularity – usually formulated as a productivity restriction or constraint – that the verbal slot is more conventionally filled by a transitive verb than by other types of verbs (Bauer, Lieber and Plag 2013: 307). The intransitive verbs *sleep* and *live* and the weakly transitive verb *have* are less frequently found in the pattern and therefore

less regular and conventional. More extreme cases of deviation found in the BNC include, for example, the noun-based forms *unmarriageable, unknowledgeable, unstereotypeable* and *uncommonable* (used to encode the intention 'cattle which cannot be sent to the commons for grazing'). To extend the envelope of originality even further towards E-creativity, one could coin the adjective-based form *unhappyable* to convey the intention 'impossible to be, become or make happy'. As is shown in Figure 4, the pattern *what the N* is extended in an F-creative manner by formal derivatives of *fuck* such as *fudge* or *frick*. Omission of the *wh*-pronoun (e.g. *The fuck are right wing conservative Republicans doing on a gossip site*, COCA) and adjectival insertions (*What the fuckety fuck?*, COCA) can count as being more E-creative, as would the invented variant *what the Nancy Pelosi*. Readers should be reminded that due to the absence of clear-cut boundaries, all these examples find their place in a continuum of onomasiological regularity ranging from highly frequent via rare tokens to novel variants that show little connection to existing exemplars of the pattern.

From a mathematical point of view, the way in which I conceptualize and visualize E-creativity may well be considered slightly misleading. Figure 4 suggests that there could be regularity scores below 0, i.e. negative ones, which is mathematically weird. Conceptually, the visualization is also somewhat shaky, as it would indicate that a form such as *unhappyable* does not tap into any known symbolic regularity, which does not make an awful lot of sense either. One could indeed argue that any linguistic innovation that makes use of conventional linguistic raw material such as phonemes, graphemes or morphemes inevitably stays within the regularity space delimited by previous utterances (for a discussion of this issue in a rule-based framework, see Bergs and Kompa 2020). Keeping this in mind, the point of the illustration is indeed to suggest that at the time when they are uttered, E-creative uses relying on an existing pattern are characterized by the fact that they extend the realm that can be considered regular in the symbolic dimension. However, by virtue of the very fact that they have been uttered, they shift the boundaries of the regularity space and become part of the available regularity potential. If the word *unhappyable* caught on, i.e. if it got used by speakers and gradually became usualized and diffused, the bubble representing its vector would embark on a trajectory leading from the top-right to the bottom-left corner of the cube.

5.1.2 Creativity in the Semasiological Dimension

When speakers use existing forms with new meanings and in order to denote new referents, they stretch conventionality in the *semasiological* dimension of

regularity. Consider the lexeme *uncomfortable*, which is conventionally associated with the meanings 'not able to give comfort' (e.g. *He was on a very uncomfortable bed*, BNC) and 'not able to feel comfort, feel uneasy' (e.g. *He looked faintly uncomfortable*, BNC). A semasiological extension of these regularities could be to use the adjective in an utterance like *Little Anna is totally uncomfortable today* with the novel meaning 'it is impossible to comfort little Anna' rather than 'little Anna feels uneasy'. Speakers using this type of extension in an F-creative manner run a greater risk of compromising effectiveness, because it is uncertain whether the hearer actually gets the new meaning. In the long run, such acts of F-creativity can lead to a shift or extension of the semantic scope of a pattern, which we can also see in formations of the type *unsleepable, unhaveable* and also *unmarriageable*. These also extend existing semasiological regularities, as they loosen constraints on what the pattern *un-* + X + *-able* can mean. Semasiological creativity can be exploited in an E-creative manner with the intention to be recognized as a salient metaphor or deliberate pun or other form of humour.

5.2 Creativity in the Syntagmatic Dimension

All the examples of symbolic creativity provided so far – except for conversion and the semantic extension of *uncomfortable* – also involve creativity in the syntagmatic dimension, because existing morphological and lexical material is combined in a novel way. Regarding the pattern *un-* + X + *-able*, the most striking and isolated example is the invented *unhappyable* (see Figure 5). As a highly innovative combination of base, prefix and suffix, it lies outside the existing pool of syntagmatic regularities. Other interesting examples of syntagmatic E-creativity that are attested in the BNC include formations based on prepositional verbs such as *unputdownable* or *unturndownable*, as well as formations using complex nouns such as *unmarriageable, unknowledgeable, unpigeonholeable* and *unstereotypeable*. Such formations stand out because the existing record of syntagmatic regularity includes few examples of multi-word verbs and complex nouns as fillers of the variable slot.

5.3 Creativity in the Contextual Dimension

Essentially, what I have discussed so far under the label of symbolic and syntagmatic creativity has been part and parcel of research on productivity and creativity in word-formation. In contrast, creativity in the remaining two types of regularity has received less attention. Consider first creativity in the contextual

Figure 5: Syntagmatic F-creativity and E-creativity relative to three-dimensional regularity space (bubbles stand for differently regular utterance types; the graph is not based on real data, but only serves illustrative purposes).

dimension. As the pattern *un- + V + -able* is not sensitive to contextual and socio-pragmatic regularities, it cannot provide good examples here. The phrase *what the fuck* can, however. In the large, multi-million-word news section of COCA, this taboo phrase does not occur a single time, indicating a strong regularity in the contextual dimension. If a news anchor or reporter actually decided to use it, or if it was inadvertently recorded and broadcast, this would clearly be innovative and thus E-creative, since it would considerably extend the contextual regularity profile of this phrase. To add some more examples from the field of phraseology, when I use the time-sensitive routine formula *good morning* to greet a student who has just resurfaced from a deep slumber in an afternoon lecture, then I also perform a breach of a contextual regularity. While this would not be very funny, it can count as being creative. Another example is the – by now conventional – recontextualization of the phrase *thanks for coming to my TED talk* from the original TED talk situation to its use as an ironic meme on Tumblr and other social media platforms. Or when someone transfers the phrase *And the winner is* from its usual habitat, the announcement of the winner of an Oscar Academy award, to announce the winner of some private contest, then we can also argue that it is a case of creativity in the contextual dimension.

5.4 Creativity in the Community-Related Dimensions

Finally, if an underlying pattern is sensitive to the social or community-related dimensions of regularity, these can also be exploited for the full range from F-creativity to E-creativity. The news anchor example mentioned above is in fact a case in point, as it not only goes against contextual conventions regarding formality, register and style, but also conventions regarding social roles and relations. Imagine further an 85-year-old who alludes to a meme originating in youth slang by saying something like *one does not simply beat grandpa at ping pong,* or a speaker of American English mimicking a posh upper-class British accent. Although there is no novelty in form or meaning, such utterances can be considered E-creative, as they extend the contextual and/or socio-pragmatic regularity profiles of the given utterance types (see Figure 6).

As we have seen, utterances are often creative in several dimensions at the same time. The most extreme cases of F-creativity would combine low degrees of regularity in all dimensions, as is indicated in Figure 7.

Mathematically, these cases could be identified by looking for vectors representing very low scores for conditional probabilities in all dimensions. However, such cases are very unlikely to exist, because novelty in all dimensions of regularity would cause a major challenge for mutual understanding and thus run the risk of being ineffective. Multi-dimensional E-creativity is even less likely to occur, because it would essentially mean that someone produces an utterance that is not at all regular in any dimension of conventionality. Such formations do exist, however: they are referred to as *ex-nihilo creations* in the relevant literature and illustrated by examples such as *Kodak* or *Google.* And they are known, or at least said, to be extremely rare. What keeps speakers from indulging in such high degrees of creativity and what may encourage them to do so once in a while will be discussed in the next two sections.

6 How do Effectiveness and Appropriateness Constrain Creativity, and what other Forces Working Against Creativity can be Identified?

According to the definitions provided in the introduction, creativity is mainly constrained by communicative effectiveness and appropriateness. This is shown in Figure 8, where effectiveness is represented as a counter-force to creativity. The more speakers venture to leave the current regularity space, here illustrated for
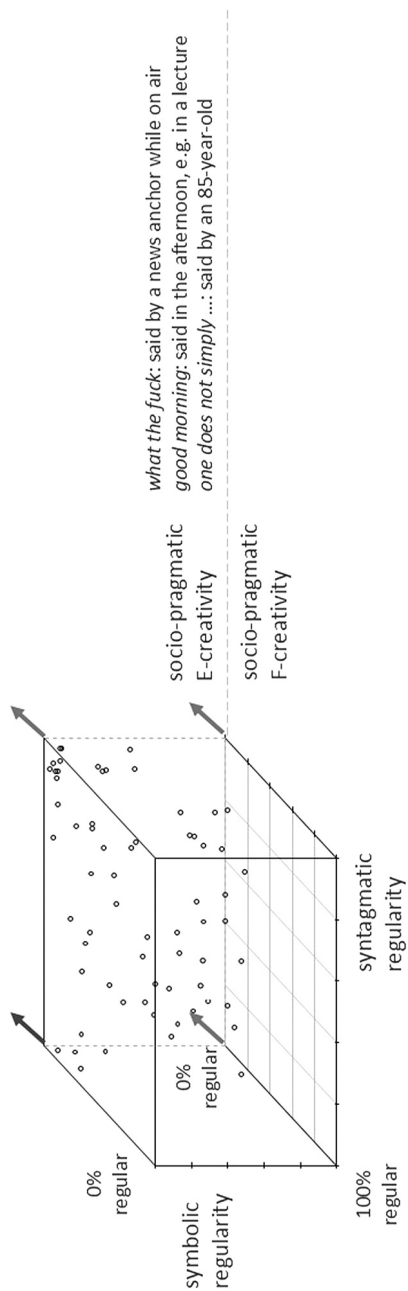
Figure 6: Socio-pragmatic E-creativity relative to three-dimensional regularity space (bubbles stand for differently regular utterance types; the graph is not based on real data, but only serves illustrative purposes).

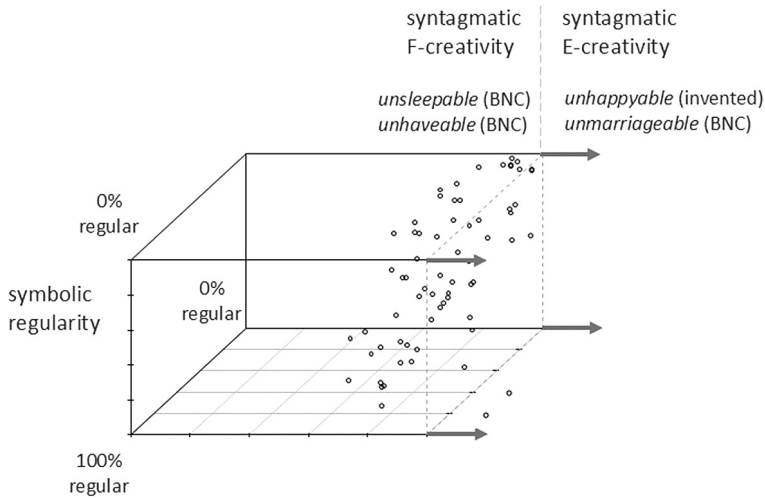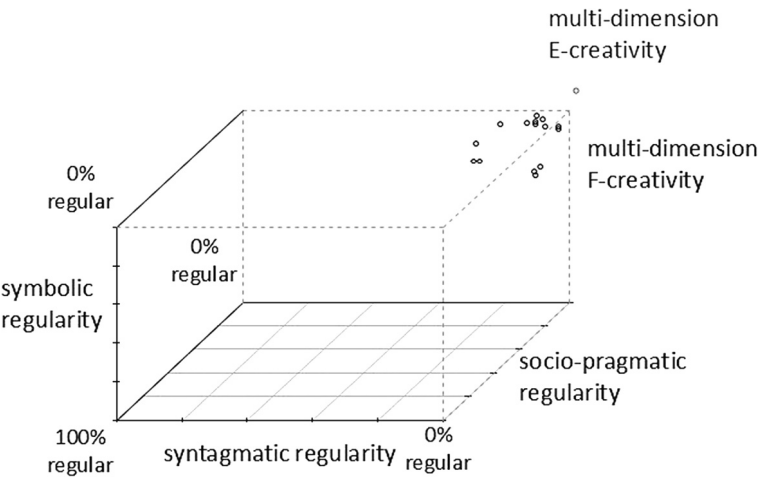**Figure 7:** Extreme cases of multi-dimensional F-creativity and E-creativity relative to three-dimensional regularity space (bubbles stand for differently regular utterance types; the graph is not based on real data, but only serves illustrative purposes).
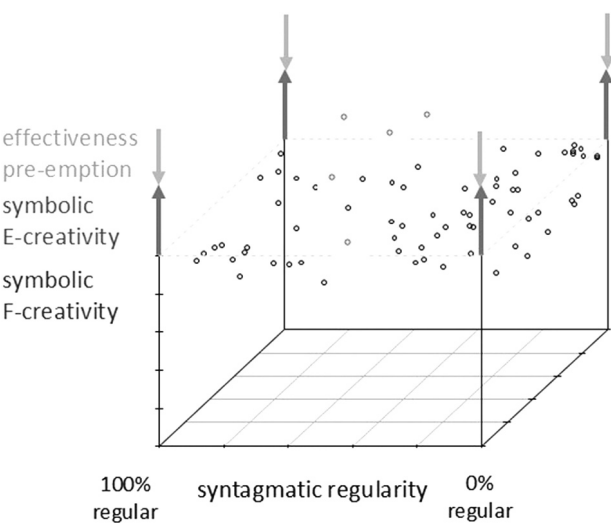
Figure 8: Effectiveness and pre-emption/blocking as forces constraining creativity.

the symbolic dimension, the more they risk compromising co-semiosis, i.e. the mutual belief of mutual understanding, and thus communicative effectiveness.

The notion of *appropriateness*, which was also mentioned in Section 1, is actually built into the system of multi-dimensional regularity by means of the contextual and socio-pragmatic dimensions. These dimensions capture the degree to which an utterance is situationally and socially appropriate in the sense of being correlated with and indexical of certain situational or social factors.

Effectiveness turns out to be an ambiguous notion when we take into account another factor constraining creativity, viz. the degree of unconventionality or distance to the conventional. So far, I have talked about effectiveness in terms of *communicative effectiveness*, associated with the goal of effortless and smooth mutual understanding. It is in this sense that effectiveness is a creativity-constraining force. However, intentional E-creativity tends to target the opposite type of effectiveness, one associated with extravagance, humour, fun or originality for originality's sake. Let's call this type of effectiveness *creative effectiveness*. As these two types of effectiveness produce opposite effects, speakers are forced to balance them out depending on their goals. On the one hand, when speakers aim to maximize mutual understanding in spite of their wish to be creative, they tend to minimize the degree of unconventionality. The data that we have looked at in this paper indicate that speakers do not venture far from the trodden paths. Rather, they tend to rely on strong formal, semantic or larger conceptual, frame-like similarities to reduce the distance to what is highly conventional and make themselves understood.[6] This is not only true of clearly F-creative formations, but also of those that reach out into the domain of E-creativity. On the other hand, when the explicit aim is to produce a strong creative effect, then the distance to the conventional must be larger and salient enough to be recognized. How speakers balance this out depends on their goals, but also on the context and other factors which may actually give more leeway to their creative inclination.

A further factor that constrains creativity is the prior existence of a highly conventional way of solving a given communicative task. This constraint is known as *blocking by synonym* or *pre-emption* (Clark and Clark 1979: 798, Schmid 2016: 116–117). Essentially, both notions capture the onomasiological phenomenon that speakers are much less likely to coin a novel composite, derived or converted form, e.g. *stealer*, if one or more conventionalized simple (*thief*) or complex forms (*robber, shoplifter, pickpocket*) are already available for expressing the target concept. In actual practice, it seems that speakers do not always adhere to the princi-

---

[6] In case the reader may have wondered about *What the Sam Hill* when looking at Figure 1, *Sam Hill* is a conventional euphemism for *devil* based on formal similarity.

ples of blocking or pre-emption. This can have various reasons. First, the speaker may not be aware of the conventionalized solutions, which is frequently the case in first and second language acquisition. Second, the speaker may currently be unable to access and produce this solution, for various reasons. Third, and closest to the present concerns, the speaker may decide to overrule blocking/pre-emption in order to achieve a creative effect.

7 What are the Factors that can Loosen the Constraints on Creativity Imposed by Effectiveness, Appropriateness and Other Constraining Forces?

Speakers can indulge in creativity to achieve communicative effectiveness or creative effectiveness, whichever they strive for. The leeway they have is controlled by various factors. First of all, it is increased by the linguistic, social and cultural experience speakers share with their interlocutors. It is a general principle of communication that I can expect to communicate more effectively and with less effort when I talk to someone I have known for a long time. Therefore, the tolerable distance from 'the cube' increases in correlation with shared experience, common ground and mutual familiarity among the interlocutors (Figure 9). For instance, while perhaps 'over the top' in most everyday situations, a daringly creative utterance like *what the Nancy Pelosi are you doing here* might work perfectly well in a close-knit group of US politicians, in spite of the very low degree of similarity to existing variants of the pattern. In a group of colleagues working together in an administrative unit, even the variant *what the paperwork do you want me to do with this here* may work well enough.

Turning to the effects of context, I should begin by clarifying how the notions of *context* and *contextual regularity* are connected. *Contextual regularity* is a characteristic of conventions. This entails that it emerges cumulatively over repeated usage events which share certain contextual features. *Good morning* is regularly used in the morning, nouns ending in *-ism* are typically used in academic prose and more formal contexts, the word *unsexable* is typically used in contexts revolving around the topic 'fish'. These are contextual regularities. In contrast, *context* is a characteristic of situated usage events.

The situational, social and cultural context of utterances functions as a modulating factor for acts of creativity (see Figure 10). Contexts that deliver strong cues for understanding extend the scope for creativity. The same is true of informal

effectiveness
pre-emption
shared
experience
symbolic
E-creativity

symbolic
F-creativity

100%
regular   syntagmatic regularity   0%
regular
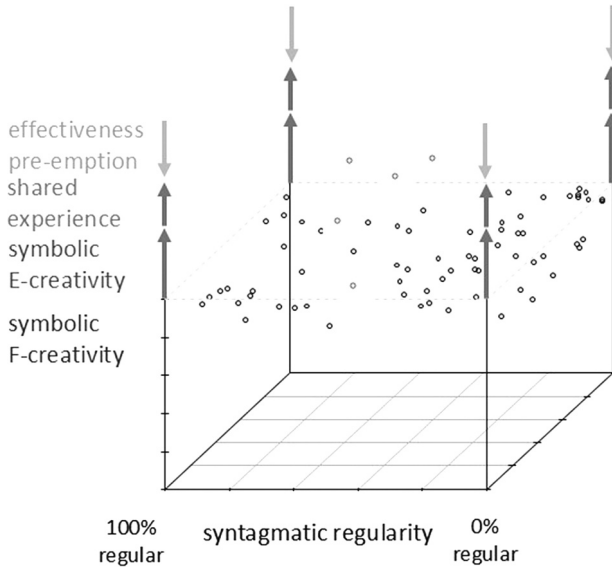
Figure 9: Shared experience as a force extending the scope of creativity.

shared
experience

effectiveness
pre-emption

context

symbolic
E-creativity

symbolic
F-creativity

100%
regular   syntagmatic regularity   0%
regular
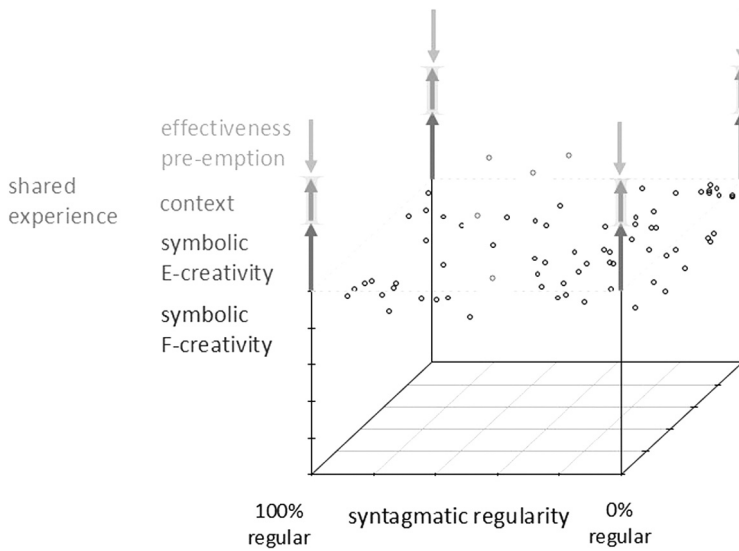
Figure 10: Current context as a factor modulating creativity.

contexts, where precise mutual understanding may not be the top-ranking goal of interlocutors, compared to other interpersonal and social functions of communication. Less constrained and formal contexts that require more precision reduce the leeway that speakers have. A casual conversation among friends in a bar is a better breeding-ground for audacious acts of creativity than a meeting of lawyers who discuss the details of a contract or a meeting of linguists who put the finishing touches on a co-authored paper.

8 What is the Role of the Cognitive Routines of Individual Speakers in the Production and Perception of Creative Utterances?

The notion of *routine* has surfaced in several places in the course of this paper. In the EC-Model, it is assumed that linguistic routines – and indeed linguistic knowledge in general – are subserved by patterns of associations which are extracted from the commonalities of usage events and routinized by repetition (see Section 3 above). Usage events are rife with commonalities – for the simple reason that they rely on the regularities discussed above. Communication requires conventions, conventions are regularities of behaviour manifested in commonalities of utterances, associations become routinized by repeated processing of such commonalities.

I distinguish between four types of associations, which largely correspond to the dimensions of regularity. *Symbolic associations* are activated and become routinized when processing symbolic (onomasiological and semasiological) regularities. *Syntagmatic associations* correspond to syntagmatic regularities. *Socio-pragmatic associations* are activated and become routinized while processing situational and social regularities. The fourth type of association are *paradigmatic associations*, which are active and become routinized as links to semantic and/or formal alternatives that are co-activated as competitors during processing.

All four types of associations become active in cooperation and competition during processing, and eventually routinized by repetition. This is what I mean when I talk about *patterns of associations*. Figure 11 provides illustrations of such patterns for the specific lexemes *uncomfortable* and *unstereotypeable*. Starting with the former, we see strong syntagmatic associations (marked by bold horizontal arrows) linking the constituents *un-*, *comfort* and *-able*. We can assume that these associations have become strengthened by the frequent repetition of the lexeme *uncomfortable*. This syntagmatic strengthening has a number of effects

which are summarized in the *syntagmatic-strengthening principle* (Schmid 2020: 236): As the syntagmatic links between the constituents are strengthened, the symbolic associations connecting the forms of constituents to meanings are weakened (indicated by thin broken vertical arrows), while the symbolic association connecting the whole sequence to its holistic meaning is strengthened (represented by the bold vertical arrow connecting the composite form *uncomfortable* with the meaning 'not at ease'). At the same time, and also as a consequence of syntagmatic strengthening, paradigmatic associations to semantic and formal competitors of the constituents are weakened (thin broken diagonal arrows), while paradigmatic associations of the composite form are strengthened (solid diagonal arrow pointing to 'unhappy', 'comfortable, 'sad', etc.). This means, e.g., that speakers do not have strong associations connecting the word *uncomfortable* with formal and semantic competitors of its base *comfort*, reflecting the high degree of lexicalization of the composite form. Socio-pragmatic associations are not rendered in the figure, as the word is not subject to any specific regularities here.
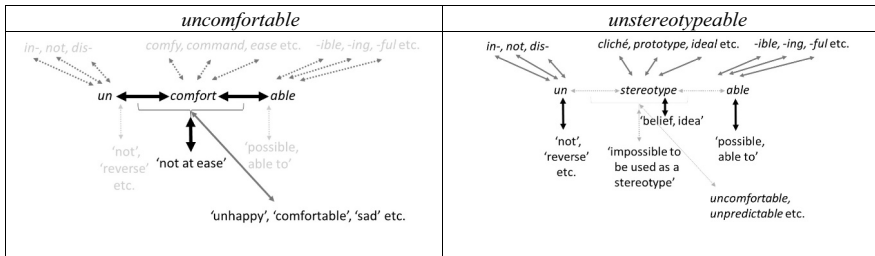
Figure 11: Hypothetical patterns of associations subserving the lexemes *uncomfortable* and *unstereotypeable*.

Shifting our attention to the representation for the lexeme *unstereotypeable*, we can appreciate the contrast between repetition-driven syntagmatic strengthening (on the left, *uncomfortable*) and a novel, or at least highly unconventional formation, which is therefore fully compositional (*unstereotypeable*, on the right). In the graph on the right, syntagmatic associations are weak, as they have not been strengthened by repetition. Symbolic associations of component parts are strong, since these parts are familiar, but the symbolic association of the composite form is weak, too. Paradigmatic associations can be activated to other, more familiar words sharing the internal makeup with *unstereotypeable*; paradigmatic associations that are specific to this word itself are unlikely to become active. Instead, and in contrast to *uncomfortable*, paradigmatic associations of the component parts remain strong.

The processing of the innovative, or at least very rare, word *unstereotypeable* can also rely upon a pattern of associations that has been gleaned from the commonality of all adjectives of the *un-* + V + *-able* variety and their meanings (see the left part of Figure 12). This commonality lies in the formal identity and semantic similarity of the forms *un-* and *-able* and the relational similarity (or analogy) of the verbal bases vis-à-vis this environment (shown on the left, vertical arrows representing paradigmatic associations here). Repetition of different, but in these respects similar forms and meanings leads to strengthening of the pattern of associations depicted on the right in Figure 12: syntagmatic associations involving a variable component; symbolic associations to the meanings shared by the tokens of this pattern; relatively strong paradigmatic associations to competing patterns and also to elements competing with constituents.

uncomfortable
unacceptable
unreasonable
understandable
unsuitable
unpredictable
undesirable
unbelievable
unreliable

in-, not, dis-   *N, Adj, Adv*   *-ible, -ing, -ful* etc.

un   V   able

'not',
'reverse'
etc.

'can be V-ed,
has the quality
of V-ing'

'possible,
able to'

'V-able', un-V etc.

un   V   able
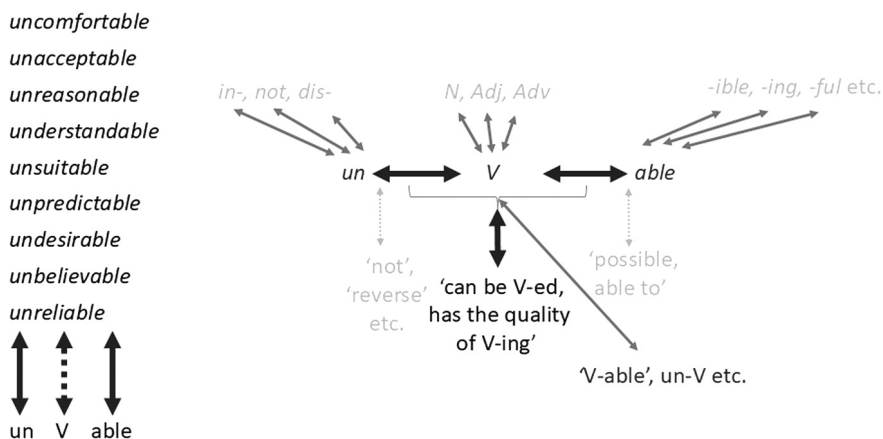
Figure 12: Hypothetical patterns of associations subserving the pattern *un-* + V + *-ing*.

The mental representation that is available in such a form is in many ways and for most practical purposes equivalent to a word-formation schema or morphological construction of the type suggested in network models (Bybee 2010), construction grammar (Diessel 2019, Goldberg 2019), or construction morphology (Booij 2010). One crucial difference, however, is that the associationist account is more flexible and hence less surprised, metaphorically speaking, to see that there are also some formations, like *unstereotypeable*, which use nouns as bases. This account of routines can also explain why *unstereotypeable* will be experienced as being creative by some speakers and can still be communicatively effective: we can assume that it is not in tune with most speakers' routines as a specific item, and it is not densely connected to other forms in the associative network, but still

similar enough to familiar forms and meanings and to a more general pattern of associations to be processed successfully.

Given their very nature as linguistic habits, routines generally stand in the way of creativity. Under normal circumstances, we tend to do what we always do, and what we always do tends to be in line with existing conventions. Therefore, we can treat routines as factors that constrain creativity. This is not to say that speakers are reluctant to stray from the trodden paths in their minds. In fact, they do so for many reasons, not only when they want to achieve creative effectiveness, but also, for instance, when they are facing a new communicative task, when they want to get across an original idea or simply because they enjoy novelty for novelty's sake. As routines are speaker-specific by definition, the perspective on routines also entails that one speaker or a group of speakers, e.g. the members of a family or closely knit community of practice, can have a highly routinized representation of a linguistic element or pattern that they encounter and use frequently. In contrast, another speaker, e.g. an outsider, has no specific representation, but must rely on similarities to other items and more general entrenched routines – which comes down to experiencing a given processing event as being 'creative'. In this way, routines can be considered as sources of speaker-specific assessments of what is creative, both in production and comprehension. If I go out of my entrenched ways of expression and coin something novel, only to find out that others consider my novel formation as perfectly conventional, then I can still feel justified in claiming that I was creative.

That said, it is not the case that speakers only accept as conventional what they themselves have represented as routines. We are perfectly able and happy to accept ways of speaking that lie outside our own box, because we are used to the fact that others speak differently (depending, for example, on their social background and corresponding differences regarding linguistic biographies). Furthermore, entrenchment is not as solipsistic as it often seems (Schmid 2022). Speakers do not entrench routines in a social vacuum, but instead extract all their linguistic knowledge from interpersonal communication, which inevitably comes complete with an interpersonal encounter (if only between writer and reader) and a social event. Part of the repetition-driven knowledge we entrench is information about what other speakers regularly do, what these speakers are like and how they behave in different types of situations. As a result, we are used to not considering utterances as creative, even if they lie outside the scope of what we have routinized.

9 Conclusion

I started out by asking the question what is creative to whom and why. Given the discussion in the preceding pages, the answer to this question might read as follows:

Linguistic creativity is linguistic novelty, defined and operationalized as a low degree of regularity in conventional regularity space charted by multi-dimensional conditional probabilities,
–   which is constrained by
    –   communicative effectiveness, defined as degrees of mutual understanding,
    –   distance from what is highly conventional, defined by location in multi-dimensional regularity space,
    –   pre-emption, defined as onomasiological competition within multi-dimensional regularity space,
    –   and speakers' routines, defined as differentially entrenched patterns of associations,
–   afforded and promoted by
    –   the availability of variable patterns of regularity, defined by multi-dimensional conditional probabilities,
    –   creative effectiveness, defined as distant, intentional and ostentatious novelty,
    –   long-term shared experience,
–   and modulated by current context.

This definition of linguistic creativity is very general. As noted in the introduction, the domains of word-formation and phraseology provide a special breeding-ground for creativity by offering mutual competition and support between lexically specific complex items and expressions, on the one hand, and partly or fully schematic patterns, on the other. Creativity, especially strong forms of creativity labelled as *E-creative,* are achieved and recognized more easily, the more morphologically and lexically specific and fixed its conventional source is. Modifications of highly fixed proverbs or idioms are cases in point. A large part of the raison d'être of partly specific morphological patterns such as derivations and phraseological templates lies in the F-creative potential that speakers can exploit to react to changing communicative demands in a fairly innocuous way. At the same time, these highly productive patterns open up ways of being E-creative in the form of more daring and extreme new formations while keeping up comprehensibility. Due to their open form, the lexically unspecific word-formation patterns of

conversion and N + N compounding or Adj + N compounding offer enormous scope for F-creativity. Speakers who plan to exploit these patterns for E-creativity have to come up with fairly extreme variants, simply because the space of possibly conventional variants is so large.

References

Bauer, Laurie. 2001. *Morphological productivity*, Cambridge: Cambridge University Press.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Bergs, Alexander. 2019. What, if anything, is linguistic creativity? *Gestalt Theory* 41(2). 173–183.

Bergs, Alexander & Nikola Anna Kompa. 2020. Creativity within and outside the linguistic system. *Cognitive Semiotics* 13 (1).

Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.

Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Callies, Marcus. 2016. Of soundscapes, talkathons and shopaholics: On the status of a new type of formative in English (and beyond). *STUF – Language Typology and Universals* 69 (4). 495–516.

Clark, Eve V. & Herbert H. Clark. 1979. When nouns surface as verbs. *Language* 55 (4). 767–811.

Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.

Davies, Mark. 2008-. *The corpus of contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Diessel, Holger. 2019. *The grammar network: how linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.

Evert, Stefan. 2005. The statistics of word cooccurrences: word pairs and collocations. Stuttgart: PhD thesis Universität Stuttgart.

Glynn, Dylan. 2014. The many uses of *run*. in Dylan Glynn & Justyna A. Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 117–144. Amsterdam & Philadelphia: Benjamins.

Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*, Princeton, NJ: Princeton University Press.

Hansen, Barbara, Klaus Hansen, Albrecht Neubert & Manfred Schentke. 1985. *Englische Lexikologie. Einführung in Wortbildung und lexikalische Semantik*. 2. Aufl. Leipzig: VEB Verlag Enzyklopädie.

Hoffmann, Thomas. 2022. Constructionist approaches to creativity. *Yearbook of the German Cognitive Linguistics Association* 10 (1). 259–284.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin & New York: Mouton de Gruyter.

Plag, Ingo. 2004. Syntactic category information and the semantics of derivational morphological rules. *Folia Linguistica* 38 (3–4). 193–225.

Runco, Mark A. & Garrett J. Jaeger. 2012. The standard definition of creativity. *Creativity Research Journal* 24 (1). 92–96.

Ryder, Mary Ellen. 1994. *Ordered chaos. The interpretation of English noun-noun compounds*. Berkeley etc.: University of California Press.

Sampson, Geoffrey. 2016. Two ideas about creativity. In Martin Hinton (ed.), *Evidence. Experiment and argument in linguistics and philosophy of language*, 15–26. Bern: Peter Lang.

Schmid, Hans-Jörg. 2015. A blueprint of the Entrenchment-and-Conventionalization Model. *Yearbook of the German Cognitive Linguistics Association* 3 (1). 3–25.

Schmid, Hans-Jörg. 2016. *English morphology and word-formation: An Introduction*. 3rd edn. Berlin: Erich Schmidt.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system. Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.

Schmid, Hans-Jörg. 2022. Social and cognitive foundations of linguistic variation: towards a unified complex-adaptive framework. Talk presented on *Abralin Ao Vivo*, available at https://aovivo.abra lin.org/lives/hans-jorg-schmid/

Schmid, Hans-Jörg. 2024. Multi-dimensional regularity analysis: How the Entrenchment-and-Conventionalization Model can be applied to corpus data. *Yearbook of the German Cognitive Linguistics Association* 11 (1). 23–56.

Stangl, Joshua. 2021. Investigating the *wh-word the X* construction. Munich: LMU BA Thesis.

Stefanowitsch, Anatol & Susanne Flach. 2017. The corpus-based perspective on entrenchment. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*, 101–127. Boston & Berlin: APA and Walter de Gruyter.

Sternberg, Robert J. & Todd I. Lubart. 1999. The concept of creativity: Prospects and paradigms. In Robert J. Sternberg (ed.), *Handbook of creativity*, 3–15. Cambridge: Cambridge University Press.

Štekauer, Pavol. 1998. *An onomasiological theory of word-formation*, Amsterdam & Philadelphia: John Benjamins.

Štekauer, Pavol. 2005. Onomasiological approach to word-formation. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 207–232. Dordrecht: Springer.

Szymanek, Bogdan. 2005. The latest trends in English word-formation. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 429–448. Dordrecht: Springer.

Tin, Tan Bee 2022. *Unpacking creativity for language teaching*. London & New York: Routledge, Taylor & Francis Group.

Uhrig, Peter. 2018. I don't want to go all Yoko Ono on you. *Zeitschrift für Anglistik und Amerikanistik* 66 (3). 295–308.

Warren, Beatrice. 1978. *Semantic patterns of noun-noun compounds*. Gothenburg: Acta Universitatis Gothoburgensis.

Warren, Beatrice. 1990. The importance of combining forms. In Wolfgang U. Dressler, Hans Christian Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Contemporary Morphology*, 111–132. Berlin & New York: Mouton de Gruyter.

B Dynamics of Creativity and Routine in Diachrony

Fabian Fleissner

How to Get into Containers

The Emergence of German *Come to mind* Constructions

Abstract: Semantic changes of lexical verbs and within multi-word expressions remains a relatively unexplored domain in historical semantics in contrast to studies on grammaticalization. This paper deals with the diachronic development of the syntactic pattern [*in* 'in' N *kommen* 'come']. The most prominent debate within this context revolves around different light verb constructions of the pattern [P N V] which have received much attention in recent literature (see Fleischhauer and Hartmann 2021). Particular attention was given to the semantic content of the verb *kommen* and a presumed desemanticization process. By introducing a German 'Come to mind' construction I will demonstrate that the general development of the structure cannot be uniformly described. Instead, I propose a comprehensive semantic and cognitive viewpoint to explore the evolution of [*in* N *kommen*], which suggests that the emergence of individual constructions depends on the application of several metaphorical schemas that were conventionalized at different points in time.

1 Introduction

The present study investigates multi-word expressions which are formed following the pattern of the VP [*in* 'in' N *kommen* 'come']. This pattern gave rise to different constructions at various points in the history of German and its complexity poses a formidable challenge to understand its evolution comprehensively. This challenge encompasses various other constructions involving other movement and transfer verbs and different prepositions. The aim of this paper is to exemplify the developmental path of the specific construction [*in* 'in' N *kommen* 'come'] from a cognitive perspective.

The history of German has witnessed the emergence and disappearance of different manifestations of [*in* 'in' N *kommen* 'come'], making it appear initially chaotic and disjointed. The structure is encountered in present day German in

several different ways with respect to its semantics and its degree of idiomaticity. The diachronically primary construction (1) is almost completely schematic in the sense that it denotes any change of location of any moving or movable entity: $[P_{loc} \ N_{loc} \ V_{motion}]$. 'Come to mind' constructions as in (2) are semi-specified and more restricted with respect to the verbal and the prepositional slot. They lack a concrete local reading and denote abstract entities like MENTAL CONTENT: $[P_{loc} \ N_{mind} \ V_{motion}]$. Light verb constructions as in (3) are sometimes considered almost fully specified, as the meaning of the whole structure is non-compositional, but their exact categorization and definition remain subject to debate (see e.g. Fillmore 1987, 1997; Fleischhauer and Neisani 2020; Smirnova and Stöber 2022): $[V_{motion} \ P_{loc} \ N_{event}]$.

(1) *Der Schock war so groß, weil damals Millionen von Einwanderern per Schiff in die Neue Welt gekommen waren.*
'The shock was so great because millions of immigrants had come to (lit. 'in') the New World by ship at the time.' (Der Spiegel, 19.09.1983)[1]

(2) *Der erste Gedanke, der mir in den Sinn kam, war: Hunger.*
'The first thought that came to (lit. 'in') my mind was: hunger.' (Moers 1999)

(3) *Denn nur er konnte das Gesetz in Antrag bringen, und gegen seinen Willen kein Vorschlag in Berathung kommen.*
'For only he could propose the law, and no proposal comes into discussion against his will.' (Bluntschli 1875)

Both 'Come to mind' constructions (2) and light verb constructions (3) result from conceptual metaphors in that they create spatialized target domain entities from prelinguistic source domains: ABSTRACT ENTITIES ARE OBJECTS (2), MINDS ARE CONTAINERS (2), STATES / EVENTS ARE LOCATIONS (see Johnson 1987; Lakoff and Johnson 2003; Tendahl and Gibbs 2008). They nevertheless differ strongly with respect to their diachronic development: Light verb constructions consisting of the verb *kommen* are rarely attested until Early New High German and still continue to find high productivity (see Fleischhauer and Hartmann 2021). In contrast, 'Come to mind' constructions are already attested extensively in Old High German. The term 'Light verb constructions' (*Funktionsverbgefüge*, cf. e.g. Von Polenz 1987; Van Pottelberge 2001; Seifert 2016, 2020; Kamber 2008; Heine 2020; Harm 2021) implies that these constructions are primarily described in relation to the status of the

---

1 Examples from the DTA-corpus are cited according to the following pattern: (author/ publication, year).

verb, which, in contrast to its 'heavy' use, is considered restricted or devoid of its 'full' semantic weight. Consequently, they have been repeatedly examined and analyzed from the perspective of grammaticalization theory (see Fleischhauer and Hartmann 2021 for an overview). On the other hand, 'Come to mind' constructions are rarely mentioned in the linguistic literature.

This paper diverges from traditional approaches like grammaticalization theory, which primarily focuses on the linear progression of grammatical elements from lexical to more abstract and functional units. I will argue that there is no coherent and uniform history of the structure [*in* N *kommen*], but rather specific constructions with various functions that feed on different metaphorical image schemas. The interpretation of the individual components of these constructions, especially the verb *kommen*, depends on the holistic constructional meaning and not on grammaticalization and desemanticization.

The paper is structured as follows: Section 2 includes the theoretical argumentation about desemanticization and grammaticalization, which I will challenge in relation to the research subject. In Section 3 I offer a working definition of metaphors emerging in the context of MIND and MENTAL CONTENT. In Section 4, I will explore the possibilities of abstract conceptualization of MIND in the Old Germanic languages. This investigation aims to identify elements that allow us, on the one hand, to reconstruct a metaphorical proto-system and, on the other hand, to serve as a basis for describing 'Come to mind' constructions in German. Section 5 presents the results of the analysis of the semantic space of the structure [*in* N *kommen*] in the history of German. Section 6 summarizes and consolidates the findings of the study and provides an outlook on future points of connection.

2 Theoretical Background – The Case of COME

Studies examining the historical evolution of verbal semantics have always been widely present in linguistic literature. The majority of these studies have primarily focused on the transition from 'full' verbs to auxiliaries (see Barðdal et. al 2019 for an oberview). This transition has been especially prominent in the context of grammaticalization theory, as exemplified by works such as Heine (1993), Bybee et al. (1994), and Krug (2011), which provide cross-linguistic investigations into auxiliaries. Important studies on specific languages include Diewald (1999) and Traugott and Dasher (2001) on modal verbs, as well as Fleischman (1983), Bybee and Thompson (2000), Barðdal (2001), Hilpert (2008), and Diewald and Wischer (2013) on aspectual verbs denoting future events. Haan (2007), Cornillie (2008), Diewald and Smirnova (2010), López-Couso and Méndez-Naya (2015), among

others, have explored evidential verbs in different languages. The vast amount of literature on specific individual phenomena under the label 'grammaticalization' is further exacerbated by the fact that the concept itself has drastically expanded in recent decades. Originally, it referred to a specific result of language change, which involved the transformation of lexical elements into grammatical or functional elements (see Meillet 1912). However, over time, the term has come to encompass almost the entire domain of language change (see Gildea and Barðdal 2023 for an overview). This is primarily due to the established mechanism of 'semantic bleaching' or 'desemantization' which often lead on to label any form of semantic shift of a lexeme or its use in abstract contexts as 'grammaticalized', even though other mechanisms, such as morphological and phonetic reduction or unidirectionality may not be met. To my understanding, the problem is based on one misconception. A common element in grammaticalization, and semantic change in general, is the shift from a concrete domain to an abstract one (see Langacker 1991: 325). It is also assumed that metaphorical processes often act as initiating forces in grammaticalization processes, since metaphors convey abstract or complex ideas by drawing on more concrete or familiar concepts introduce new semantic nuances (see Hopper and Traugott 1993: 87; Heine et. al. 1991: 151; Taverniers 2018). Of course, the reverse inference is not possible: Not every metaphor is an indicator of a grammaticalization process, and not every non-literal use of content words is a result of semantic bleaching. Otherwise, one would have to view every figure of speech and ultimately most forms of linguistic creativity within the context of grammaticalization.

Grammaticalization theory has always faced criticism for being limited in its explanatory power, as some argued that reanalysis or analogy were sufficient to explain the specific processes involved (see Campbell and Janda 2001; Roberts 1993; Kiparsky 2012; Anderson 2015). More recent criticism is less focused on that and more on the aforementioned broadening of the concept of grammaticalization itself. Regardless of these conceptual and ultimately terminological difficulties, there has been limited research devoted to understanding how verbs with full meanings undergo semantic changes to become verbs with different meanings, mostly through metaphorical extension. While some early studies, such as Bréal (1900), Wundt (1904), Sturtevant (1917), and Ullmann (1951, 1962), contain significant theoretical findings and observations related to semantic change, only a few studies have approached this topic from a broad cross-linguistic perspective. A typological study by Viberg (1982) investigates perception verbs related to the domains of SIGHT, HEARING, TOUCH, TASTE, and SMELL. Sweetser (1990) explores the phenomenon where words originally related to the physical domain in Indo-European languages often undergo a shift to represent abstract concepts in the psychological domain. Both Viberg and Sweetser demonstrate a consistent pattern of polysemy for perception verbs, demonstrating

how fully developed lexical verbs can acquire new meanings over time. Wegener (2001) examines how German verbs describing SENSATION diachronically emerge from expressions of PHYSICAL IMPACT. Reznikova, Rakhilina and Bonch-Osmolovskaya (2012) describe the development of PHYSICAL IMPACT verbs into PAIN verbs through metaphor and metonymy. In this research tradition, the work of Barðdal et. al (2019) is particularly noteworthy, who also focus on a specific process of semantic change in which verbs develop new meanings through metaphorical extension rather than grammaticalization. They specifically examine nouns associated with SUCCESS in Germanic and, on a broader scale, within the Indo-European language family. An important insight that this work has provided is that metaphors are dependent on the development of specific multi-word constructions in which they appear. The reconstructed Proto-Germanic metaphor SUCCESS IS MOTION FORWARD (see also Goatly 1997, 2007 and Lakoff 1993) is closely linked to specific patterns of predicate and argument structures. The consequence of this is that apparent desemanticization phenomena can hardly be isolated and generalized into a universal developmental path for individual lexemes without considering the history of the specific constructions in which they appear. Since this paper aims to explore the extent to which COME-verbs serve as a resource for the conceptualization of mental processes in Germanic, this problem becomes particularly evident. In the context of the multi-word constructions with MOTION-verbs, Fleischhauer and Hartmann (2021), within the framework of a loosely defined theory of grammaticalization, identify various signs of the 'desemanticization' of the German verb *kommen* 'come' in diachrony. They assume that *kommen* as an isolated lexical unit gradually reduces its concrete semantics to the extent that it can be used as a suitable element in different abstract contexts. Alongside its frequent appearance in multi-word units such as light verb constructions, they also consider a diachronic increase in the occurrence of abstract entities in subject position as an indicator of a desemanticized and grammaticalized *kommen*. This non-usage-based approach implies that a lexeme can develop its metaphorical potential and its overall more abstract semantics homogeneously across all constructions. On the other hand, Smirnova and Stöber (2022) do not consider the grammaticalization of *kommen* to be construction-independent. They assume a generally broad semantics for COME-verbs, which is a perspective that is well supported from a typological standpoint, since they are considered a rich resource for various grammaticalization processes that are observable cross-linguistically.[2] Since they combine traditional

---

2 They primarily refer to the *World Lexicon of Grammaticalization* (Heine et al. 2002: 68–79) and list different target domains of grammar for lexical expressions of COME-verbs, such as consecutive markers, imperative markers, future tense, hortative markers and semiauxiliary with past participles. From a Germanic perspective, another interesting phenomenon to note is the involvement of COME in different TRANSFER-constructions, the most prominent examples being GET/

grammatical descriptions of *Funktionsverbgefüge* (light verb constructions) with grammaticalization and lexicalization theories, along with more contemporary usage-based constructionist approaches (as seen in works like Traugott and Trousdale 2013; Barðdal et al. 2015; Coussé et al. 2018), they do not necessarily have to provide a unified answer to the question of the semantic content of the examined verb or the syntagmatic pattern in which it appears.

While grammaticalization is undeniably an essential aspect of language change, it is crucial not to overlook the coexistence of other linguistic processes. Different mechanisms may operate simultaneously, interact with each other, or even influence the trajectory of grammaticalization itself. Furthermore, language change can be influenced by a variety of extralinguistic factors, such as cultural, social, and historical developments. Ignoring these contextual influences and solely attributing linguistic change to grammaticalization might lead to an incomplete understanding of the intricacies involved in language evolution. In the following, I will argue that most of the seemingly detectable changes of meaning in German *kommen* 'come' occur as basic metaphorical extensions within specific usage contexts (i.e. constructions). As exemplified by the pattern [*in* 'in' N *kommen* 'come'] it will become evident that the diachronic development of *kommen* in the context of grammaticalization and desemanticization cannot be uniformly traced even for one and the same syntactic structure. Instead, I assume synchronically accessible metaphorical and metonymic patterns that are available to speakers as more or less conventionalized strategies at different points in time.

3 Metaphors in Constructions, Metaphors from Constructions

The essence of meaning lies in how we conceptualize things. Consequently, linguistic semantics needs to engage in the detailed examination and clear elucidation of abstract entities such as thoughts and concepts (see Langacker 1991: 2). The most important source domain for the conceptualization of these abstract entities is embodied experiences, which have image-schematic structure and are

---

BECOME-constructions (see Lenz 2013). Also worth mentioning is the metonymic shift from COME to PUT in Danish and in parts of neighboring North Lower German for their basic COME-verb (see Lenz et al. 2021: 75). This shows that Germanic COME-verbs are generally found in contexts of CHANGE OF LOCATION (like MOVEMENT and TRANSFER), but also in contexts of CHANGE OF STATE. The shared developments of the Germanic languages indicate that these are not phenomena of recent linguistic history.

rooted in recurring and physically perceivable patterns, such as CONTAINMENT, SOURCE-PATH-GOAL, BALANCE, etc. An image schema is a cognitive concept that refers to a mental representation or cognitive structure that emerges from our sensori-motor experiences. These experiences shape the way we understand and conceptualize abstract concepts and image schemas provide the cognitive basis for many metaphors, helping us understand and express abstract ideas in terms of these concrete experiences. Metaphors, in turn, often leverage image schemas to make abstract concepts more comprehensible in language (see Johnson 1987; Lakoff 1987). These image-schematic mappings can be found in both light verb constructions and 'Come to mind' constructions. Reconsider example (2)–(3): Both abstract or metaphorical uses of *kommen* create a spatialized target domain entity for an abstract location (GOAL). In (1), a similar image schema is present, but the domain mapping is more immediate.[3]

One common metaphor that permeate many, but not all languages is the conceptualization of MIND as a CONTAINER (see, for instance, Lakoff and Johnson 1980, 1999; Barnden 1997). In this metaphor, the mind is viewed as a receptacle capable of being filled, storing knowledge, and allowing the retrieval of MENTAL CONTENT. This metaphor, while abstract in its application to mental processes, is fundamentally grounded in our sensorimotor experiences. Consider, for instance, the physicality of our bodies: we possess orifices such as mouths, noses, and ears, through which more or less concrete entities like food, sound or pain can enter. MINDS ARE CONTAINERS is therefore strongly related to the broader conceptual metaphor BODIES ARE CONTAINERS. This becomes the foundation for the conceptual leap that gives rise to the metaphor of the mind as a container:

> We conceptualize the mind metaphorically in terms of a container image schema defining a space that is inside the body and separate from it. Via metaphor, the mind is given an inside and an outside. Ideas and concepts are internal, existing somewhere in the inner space of our minds, while what they refer to are things in the external, physical world. This metaphor is so deeply ingrained that it is hard to think about mind in any other way. (Lakoff and Johnson 1999: 266)

Just as we can fill a container with tangible objects, we metaphorically fill our minds with knowledge, ideas, and thoughts. The act of retrieving mental content

---

3 Often, the terms *metaphor* and *domain mapping* are used synonymously (see Croft 1993). However, strictly speaking, even in different concrete (in this case: spatial-concrete) expressions, the same image schema underlies them. With distance from certain prototypical uses, the degree of metaphoricality gradually increases. The reason for this is more ontological in nature: Even concrete spatial relations are not all semantically equal, even though they are formed according to the same pattern. Therefore, an image schema never has a one-to-one correspondence in linguistic expressions.

parallels the process of taking items out of a physical container. Metaphors do not emerge in isolation but rather form a network with other close representatives within their semantic field. The existence of MINDS ARE CONTAINERS not only necessarily presupposes the existence of a BODIES ARE CONTAINERS metaphor, but also an ABSTRACT ENTITIES ARE OBJECTS metaphor, under which events of mental processes like thinking and imagining are depicted as existing in a physical space outside of a person. Lakoff and Johnson (1999) identify several metaphorical emergences closely related to the CONTAINMENT schema in the context of MIND, such as MINDS ARE MACHINES as in *my mind just isn't operating today* or MINDS ARE BRITTLE as in *his mind snapped.* These individual metaphors function as family members within a broader and likely cognitively prior conceptual framework MINDS ARE PHYSICAL SPACES. The scope and complexity of such metaphorical families can vary from one language to another. Therefore, making universal statements about them is challenging, as they often reflect the unique linguistic and cultural characteristics of a given community.

There is no comprehensive study on MIND metaphors or 'Come to mind' constructions available for German. Occasionally, one can already find observations about this phenomenon in the *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm* such as *der sinn ist der ort und behälter für gedanken, vorstellungen, erinnerungsbilder*.[4] The authors list different 'Come to mind' constructions and synonym expressions, see (4a–i).[5] They also emphasize the diachronically inconsistent nature of these constructions.

(4) a. *in den sinn kommen*
         lit. 'come into the mind'
     b. *zu sinne kommen*
         lit. 'come to the mind'
     c. *einfallen*
         lit. 'fall into'
     d. *in sinn nemmen*
         lit. 'take into the mind'
     e. *aus dem sinne kommen*
         lit. 'come out of the mind'

---

4 'The mind is the place and container for thoughts, ideas, and memories', Sinn, in: Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm, Erstbearbeitung (1854–1960), digitalisierte Version im Digitalen Wörterbuch der deutschen Sprache, <https://www.dwds.de/wb/dwb/Sinn>, last access 26.09.2023.
5 The literal translations are given in the conventional word order of English verb phrases and differ from the German notation in this respect.

  f. *aus dem sinne entfahren*
    lit. 'escape out of the mind'
  g. *im sinne liegen*
    lit. 'lie in(side) the mind'
  h. *inn verstandt kommen*
    lit. 'come into reason'
  i. *[in]gedächtnusz fassen*
    lit. 'contain/grasp into memory'

It seems that the CONTAINER schema for MIND is well established in German. All constructions license dative complements as EXPERIENCERS, which is typical in Indo-European languages for expressions of PERCEPTION, EMOTION and COGNITION (see Barðdal et al. 2012, 2019; Dewey and Arnett 2015). This demonstrates that individual representatives of the metaphorical MINDS ARE PHYSICAL SPACES / CONTAINERS framework are not only semantically but also structurally closely related. To a certain extent, this metaphorical conceptualization is also based on a constructional schema: MIND can be understood as a CONTAINER because the available constructions suggest such an interpretation. Conversely, it's conceivable that the establishment of such a schema also encourages the formation of new constructions that attract the element MIND as a slot filler. Metaphors in constructions are at the same time metaphors from constructions. 'Come to mind' constructions make use of a functor-argument metaphor, which involve the combination of semantically clashing concepts, where an argument forces a metaphorical reading of the functor due to conflicts with the functor's selectional restrictions (see Ellison and Reinöhl 2022). Functor-argument metaphors are analogical or proportional metaphors, where concept A is to B as C is to D. They involve relational terms with one or more slots for arguments, establishing an analogy between different conceptual domains. A 'Come to mind' construction can conceptually only be resolved in such a way that the COME verb takes on an abstract interpretation, as both the movable element MENTAL CONTENT and the spatial metaphorical target MIND in the prepositional phrase (PP) necessitate this abstraction. Ultimately, functor-argument metaphors also enable the conceptualization of schemas like ABSTRACT ENTITIES ARE OBJECTS and MINDS ARE PHYSICAL SPACES / CONTAINERS.

  Given that abstract entities like MENTAL CONTENT are typically not immediately perceptible, it is logical that most languages establish some kind of MIND metaphors. But form and function of specific 'Come to mind' constructions in a specific language cannot be derived from this, as these characteristics also depend on the underlying inventory of motion verbs and spatial prepositions. A comprehensive analysis on all MIND metaphors or 'Come to mind' constructions in Old Germanic languages and cannot be undertaken within the scope of this paper. However,

certain conclusions can be drawn by contrasting Old High German MINDS nouns within [*in* 'in' N *kommen* 'come'] with data from related Germanic languages that are sufficient to answer whether German has always had the prerequisites to use ABSTRACT ENTITIES ARE OBJECTS and MINDS ARE PHYSICAL SPACES / CONTAINERS metaphors in different constructions.

3.1 MIND Metaphors in Old Germanic Languages

The metaphor of locating ABSTRACT ENTITIES in a conceptional PHYSICAL SPACE is present in many ancient Indo-european languages. Both Ancient Greek and Indian, for example, show certain types of MIND metaphors by deriving different MIND nouns from body parts or bodily functions. Marcinkowska-Rosół and Sellmer (2021) list AGr. *φράν* 'diaphragm, lungs', *θυμός* 'vital breath' *στῆθος* 'breast' for Greek, whereas in Indian epics, cognitive items are usually situated in the heart (*hṛd, hṛdaya*).[6] Additionally, there are conceptually less immediately accessible verbal abstracts like AGr. *νόος* 'mind; thinking; thought' (< *νέω* 'I spin' meaning 'to spin the thread of the mind' < PIE. *(s)neh₂-* 'to flow; to swim') or Ved. *manas* (< PIE. *ménos* 'mind' < PIE. *men-* 'to think').

The most commonly encountered word for MIND in Old High German within the structure [*in* 'in' N *kommen* 'come'] is OHG. *muot* 'mind' (< PGmc. *mōdaz* < PIE. *moh₁-*, *meh₁-* 'endeavour, will, temper'). The reconstructed semantics of PGmc. *mōdaz* 'sensation, mood, temper, anger . . .' (< PIE. *moh₁-*, *meh₁-* 'endeavour, will, temper') are mostly preserved in EGmc. as in Go. *mōþs*, whereas WGmc. *mōd* is characterized by a metonymic shift towards the general semantics of MIND as abstract PHYSICAL SPACE while the sensational and emotional semantics of MENTAL CONTENT recede into the background. The representative examples (5) and (6) illustrate this significant difference:

(5)   *nu quimit líhtida imon múat*
        'Now relief comes into their mind' (Otfrid, III, 23, 46)[7]

---

6 In the context of the CONTAINER metaphor, none of the words meaning 'heart' such as *κραδίη*, *κῆρ*, or *ἦτορ* are used. This evidence pertains specifically to classical Greek epics. It will become evident that Biblical Greek, for example, does indeed incorporate such a heart metaphor.
7 Old High German, Middle High German and Early New High German examples from the DDD-corpus are cited according to the original texts. If precise information is not possible due to missing metadata, the reference number is provided.

(6) *jah fullai waurþun allai modis in þizai swnagogein hausjandans þata*
     καὶ ἐπλήσθησαν πάντες θυμοῦ ἐν τῇ συναγωγῇ ἀκούοντες ταῦτα
     'And all they in the synagogue, when they heard these things, were filled
     with wrath' (Wulfila, Luke 4:28)[8]

OHG. *muot* indicates the application of a MINDS ARE CONTAINERS metaphor with *līhtida* as a (not literally) moving entity. *imon* is an amalgamation of the personal pronoun *imo* (Dat. Sg. of *er* 'he') to denote the EXPERIENCER and the preposition *in*. In contrast, Go. *mōþs* (translating AGr. θυμός 'anger, wrath . . .'[9] < PIE. *$d^huh_2mós$ 'smoke') itself enters the human body or is produced within the body. Both constructions make use of the ABSTRACT ENTITIES ARE OBJECTS metaphor. However, the conceptualization of the CONTAINER differs between the two languages. Only Old High German makes an explicit distinction between BODY and MIND (such as Ancient Greek and Vedic), while Gothic metonymically depicts BODY as a unifying CONTAINER for both. Another difference becomes apparent in that the passive structure of the Gothic example does not necessarily imply the involvement of an external force. The idea of an already existing entity that fills the CONTAINER through its swelling would also be conceivable.

The preposition OHG. *in* 'in' in (5) introduces a goal PP which specifies the goal of the abstract motion expressed by the verb *queman* 'come'. In this context, MENTAL CONTENT typically appears as specific emotions or feelings, representing the mental state of the EXPERIENCER, and less frequently as mere thoughts or other forms of consideration. 17 occurrences of OHG. *muot* in [*in* 'in' N *kommen* 'come'] demonstrate that *in muot queman* was well-established in Old High German. Further evidence is provided by numerous uses of *muot* in constructions that may not be classified as 'Come to mind' constructions but nevertheless strongly attest to the frequent application of the MINDS ARE CONTAINERS metaphor. This includes paradigmatically closely related constructions like *in muot habēn* 'to have in mind' and *in muot sīn* 'to be in mind'. Proponents of the desemanticization hypothesis may also perceive a faded semantics in HAVE and BE verbs. However, there is also evidence showing a metaphorical or at least non-literal use of verbs or predicatives that are not inherently suspected to undergo a process of desemanticization, see examples (7)–(9):

(7) *sie gicléiptun sar thaz gúat filu vásto in iro múat*
     'They impressed the joyful message very firmly in their minds' (Otfrid, I, 9, 38)

---

8 For Gothic examples the respective Bible reference is provided.

9 AGr. θυμός has undergone the same metonymic shift towards the general semantics of MIND as WGmc. *mōd.* In this context, the original meaning seems to emerge.

(8)   *gidóugno in themo múate*
     'hidden in the mind' (Otfrid, II, 21, 4)

(9)   *mih io gómman nihein in min múat ni biréin*
     'Never has a man touched my mind' (Otfrid, I, 5, 38)

It should be noted that the mere presence of the preposition *in* in these contexts as evidence for the existence of a container conceptualization of MIND in Old High German without further criteria raises certain concerns, since transferring present-day linguistic intuitions derived from contemporary Germanic languages to historical stages of German is imprecise when it comes to prepositional semantics. OHG. *in* is by far the most frequently used preposition in Old High German and has a diverse range of spatial semantics, which is significantly less restricted than that of its descendent in Modern German (see Schrodt 2004: § 28–29, Schützeichel 2006: 173, Filatkina 2018: 264).[10] However, it can be stated that OHG. *in* in the context of 'Come to mind' constructions in Old High German does not face competition from other prepositions like OHG. *zi* 'to' or adverbs like *ana* 'on'. Infrequently, *in* is alternated with the intensified adverbial form OHG. *innan* (< PGmc. *in 'in' + locative suffix **-nai* 'there'), see (10):

(10)   *Ouh ther wídarwerto thín ni quém er innan múat min*
      'But your enemy may not come into my mind' (Otfrid, I, 2, 29)

Examples like this reinforce the assumption that the use of *in* in OHG. 'Come to mind' constructions is indeed motivated by a CONTAINER schema. This is supported by the fact that it is not just a mental content that occurs to the EXPERIENCER, but the moving entity is conceptualized as an evil spirit that can actively enter the MIND if not prevented. Similar conceptualizations in closely related West Germanic dialects are much harder to identify. The structure [*in* 'in' N *kommen* 'come'] doesn't appear in the Old Saxon corpus. In fact, the preposition *in* is scarcely attested at all (along with OS. *innan*). This doesn't mean that it was impossible to express similar local references in Old Saxon (as in Old English), but the means of expression appear to have shifted towards other prepositions such as OS. *an* or OE. *on* 'on, in, to . . .'. The conceptualization of MIND as PHYSICAL OBJECT is still evident for OS. *mod.* The potential for a PHYSICAL SPACE or a CONTAINER meta-

---

10 A ratio of 8622 occurrences of OHG. *in* to 2944 occurrences of OHG. *zi* indicates the preferred usage of *in* in most concrete and abstract contexts.

phor exists but is not exploited to the same extent as in Old High German. Consider, for instance, example (11):

(11) *uuerðad imu sorga an môde*
       'Sorrow settles on his mind' (Heliand, 3496)

The absence of motion verbs in such constructions does not allow for a spatial dynamic in the sense of a 'Come to mind' construction. While MIND is conceptualized as a PHYSICAL OBJECT that interacts with emotions, (11) suggests an internal interpretation of the entire event structure, rather than the notion of an external force intruding from outside. Such evidence is not attested for Old Saxon, although, of course, the limited extent of the entire preserved text corpus must be taken into account. The Old English text corpus is significantly more heterogeneous and extensive compared to the continental West Germanic one. An exhaustive study on Old English conceptualizations of MIND is currently lacking. Certain tendencies in the usage of the most common constructions can be demonstrated based on dictionaries.[11] OE. *mōd* is the nominal core of the typical 'Come to mind' construction, see (12)–(14):

(12) *Ān wundor iċ wille seċġan þæt mē nū on mōd berann*
       'I want to tell you about one miracle that just occurred to me' (Werferth, Dialogues of Gregory)

(13) *Þā berann mē on mōd þæt iċ þās bōc of Lǣdenum ġereorde tō Englisċre sprǣċe āwende*
       'Then it occurred to me that I should translate this book from Latin into English' (Ælfric, Homilies, preface)

(14) *þā berann him on mōd his ġebrōðra ġemynd*
       'The memory of his brothers came to his mind' (Ælfric, Homilies, XXIII)

As mentioned before, OE. *on* 'on, in, to . . .' does not allow for a specific interpretation in terms of spatial relations. One can speculate whether in the conceptualization of the process is more about an event that takes place directly at the MIND (as in Old Saxon) or if it actually stimulates the idea of an out-of-body external force. The semantics of OE. *beirnan* 'run around / along' (< metathesised form of earlier

---

11 *Spirit, soul, heart*, in: A Thesaurus of Old English (2017), <http://oldenglishthesaurus.arts. gla.ac.uk/category/?id=11240>, last access 26.09.2023.

OE. *berinnan* < PGmc. **birinnaną*) is not always easy to interpret. Clearly defined spatial contexts, however, suggest that with the verbal action expressed by *beirnan*, no clear penetration into a space or object is meant, but rather a washing around or circumventing. This would go against an established CONTAINMENT schema, which is not surprising given the semantically indifferent preposition. In the Old English 'Come to mind' construction, what appears as MENTAL CONTENT are primarily less salient subjects like spontaneous thoughts, not emotions or states of being. In (12) and (13), MENTAL CONTENT is expressed through subordinate clauses, and in (14), the mental content is specified: OE. *gemynd* translates to 'memory' and reflects an archaic meaning, as the original semantics of PGmc. **gamundiz* 'remembrance, memory' (also preserved in Go. *gamunds* used to translate AGr. *μνημόσυνον* 'memory') has been retained without the WGmc. metonymic shift MENTAL CONTENT > MIND. In this usage it is an exception in Old English. Most instances of OE. *gemynd* (as well as OHG. *gimunt*) show a MIND metonymy, see (15):

(15)   *mē cōm swīðe oft on ġemynd hwelċe wiotan iū wǣron ġiond Angelcynn*
        'it very often is in my mind what counselors there were throughout England' (Alfred, Pastoral care, preface)

It also becomes evident that Old English, unlike Old Saxon, occasionally forms 'Come to mind' constructions with *cuman*. The West Germanic metonymic shift MENTAL CONTENT > MIND can also be observed in other Old High German MIND nouns like *sin* (< PGmc. **sinnaz* 'sense, perception' < PIE. **sentnós*, from < **sent-* 'to feel') or *wan.* (< PWGmc. **wāni.* 'hope, expectation' < PGmc. **wēniz* < PIE. **wenh₁-* 'to love'), **hugi* (attested in MHG. *huge* and OS. *hugi* < PGmc. **hugiz* 'thought' < PIE. **ḱk-éy-s*, from **ḱek-* 'to be able, capable') and **sebo* (attested in OS. *seƀo* and OE. *sefa* < PGmc. ** seƀô* 'mind, taste, perception' from PIE. **sep-* 'to taste'). In East Germanic, most these lexemes are not attested,[12] just like cognates of presumably Old High German formations like *gidahti*, *gidrahti* and *gitrahta* or OS. *gihugd* and OE. *gehygd* which are deverbal or paradigmatically related to different verbs of thinking and non prefixed MENTAL CONTENT nouns.

   All the Old High German nouns mentioned so far have in common that they exist as autonomous MIND nouns both within and outside of 'Come to mind' constructions, and are thus entirely confined to the domain of abstract locations. This presupposes a MIND metonymy, which distinguishes modern West Germanic from more archaic East Germanic and Proto-Germanic. A lexical overlap between

---

12 Go. *hugs* and *gahugds* 'thought, memory' are attested, but reflect the Proto Germanic semantics without the metonymic shift.

the two branches can be observed with nouns that have more immediate bodily experiences as the basis for coining MIND nouns. The most common nouns in West Germanic are descendants of PGmc. *hertō* 'heart' (< PIE. *ḱérd* 'heart') and PGmc. *breustą* 'chest' (< PIE. *bʰrews-* 'to swell'), see (16)–(22). The absence of OHG. *brust* in the structure [*in* 'in' N *kommen* 'come'] is likely just a coincidence, since numerous examples from Old High German and Old Saxon attest to the metaphorical use of the noun.

(16) *in hérza imo quámi so iz fora góte zami*
      'It came into his heart, as is fitting before God' (Otfrid, III, 2, 14)

(17) *Ni láz thir innan thina brúst arges wíllen gilúst*
      'Never let the desires of evil will come into your heart' (Otfrid, I, 12, 27)

(18) *Nim nu wórt minaz in herza*
      'Now take my words (in)to heart' (Otfrid, I, 15, 27)

(19) *in hérzen si iz bifángan*
      'They lock it in their heart' (Otfrid, V, 15, 20)

(20) *rúarta sia thiu smérza ínnan ira hérza*
      'The pain touched her in her heart' (Otfrid, I, 22, 30)

(21) *uuas imu unôðo innan breostun*
      'He felt woe in his chest' (Heliand, 3294)

(22) *Thes uuarđ Âdamas hugi innan breostun suîðo an sorogun*
      'There Adam's spirit in his chest became filled with sorrow' (Genesis, 673)

OHG. *herza* (20) and OS. *briost* (21) display the most primitive metaphorical use of the words, as an aching heart or a wounded chest reflect direct bodily experiences. An interesting conceptual intermediate stage is encountered in example (22). Here, it is not the chest or the heart itself that functions as the MIND, but rather the MIND itself is located in the chest. The Old Germanic languages reveal a metaphorical and metonymic developmental path: BODIES ARE PHYSICAL SPACES > BODIES ARE CONTAINERS > BODIES ARE MINDS > MINDS ARE PHYSICAL SPACES LOCATED WITHIN THE BODY.

Biblical Greek displays numerous heart metaphors, AGr. *καρδία* 'heart' is consistently translated with Go. *hairtō* in the Wulfila Bible. The Gospel of Mark best illustrates the conceptualization BODY and MIND in Gothic. In the context of other

body parts, it is entirely possible for abstract entities to enter the heart, suggesting the application of a CONTAINER metaphor, see (23):

(23) *unte ni galeiþiþ imma in hairto, ak in wamba, jah in urrunsa usgaggiþ, gahraineiþ allans matins.*
ὅτι οὐκ εἰσπορεύεται αὐτοῦ εἰς τὴν καρδίαν ἀλλ' εἰς τὴν κοιλίαν, καὶ εἰς τὸν ἀφεδρῶνα ἐκπορεύεται;
'Because it entereth not into his heart, but into the belly, and goeth out into the draught, purging all meats?' (Wulfila, Mark 7:19)

Go. *galeiþan* 'to go, pass (through)' (< PGmc. **galīþaną* 'to go, pass (through)' < PIE. **leyt-* 'to go, pass, die') is used, at least in this context of negation, as a motion verb in a proto 'Come to mind' construction. The idea that the heart or the whole body is a place, producer, and distributor of mental content is supported by the following examples:

(24) *qaþuþ~þan þatei þata us mann usgaggando þata gamaineiþ mannan.*
ἔλεγεν δὲ ὅτι τὸ ἐκ τοῦ ἀνθρώπου ἐκπορευόμενον ἐκεῖνο κοινοῖ τὸν ἄνθρωπον:
'And he said, that what comes out of the man, that defiles the man' (Wulfila, Mark 7:20)

(25) *innaþro auk us hairtin manne mitoneis ubilos usgaggand: kalkinassjus, horinassjus, maurþra,*
ἔσωθεν γὰρ ἐκ τῆς καρδίας τῶν ἀνθρώπων οἱ διαλογισμοὶ οἱ κακοὶ ἐκπορεύονται, πορνεῖαι, κλοπαί, φόνοι,
'For from within, out of the heart of men, come evil thoughts, adulteries, fornications, murders' (Wulfila, Mark 7:21)

(26) *þo alla ubilona innaþro usgaggand jah gagamainjand mannan.*
πάντα ταῦτα τὰ πονηρὰ ἔσωθεν ἐκπορεύεται καὶ κοινοῖ τὸν ἄνθρωπον.
'All these evil things come from within, and defile the man' (Wulfila, Mark 7:23)

The heart, like the human body as a whole, contains various emotions and thoughts. What is mostly lacking in Gothic is the concept of the heart as a pure vessel, especially when the idea is suggested that a CONTAINER would be filled by external action. Instead, it is usually envisioned more as an organ, a machine. The translator's fidelity to the Greek source makes it challenging to find contrasting examples. In (27), a more genuine Germanic structure seems to emerge:

(27) *akei unte þata rodida izwis, gauriþa gadaubida izwar hairto.*
      ἀλλ' ὅτι ταῦτα λελάληκα ὑμῖν ἡ λύπη πεπλήρωκεν ὑμῶν τὴν καρδίαν.
      'But because I have said these things unto you, sorrow has hardened your heart.' (Wulfila, John 16:6)

The author clearly feels uncomfortable describing the heart as a CONTAINER that is filled. Instead, he chooses to use the verb *gadaubjan* 'to harden' (< PGmc. *(ga)daufs* 'hardened' + denominative suffix *-(i)janą*). The external (possibly located in the body) stimulus of sorrow forces the heart to react. At this point, one could ask whether this is even a metaphor or whether it suggests the idea of an actual physical manipulation or involvement of the organ. It is reasonable to assume that the warlike pastoral cultures of the Iron Age, to which the Germanic peoples undoubtedly belonged, had access to a rich reservoir of knowledge regarding the anatomical features of humans and animals. Apart from individual immediate physical experiences like pain and well-being, the concept of a hardened or pierced heart, a growth, or fluid in the chest and lungs may not exclusively trigger abstract concepts, but rather very concrete and everyday experiences in ancient central and northern european societies. Body parts as abstract locations are primarily constituted by abstract entities acting within this physical space.

Although not exhaustively described, the data allow for some conclusions regarding the conceptualization of MIND in Old Germanic languages: Old High German makes use of an established 'Come to mind' constructions with the structure [*in* 'in' N *kommen* 'come']. Neighboring constructions suggest that the various types of MIND nouns were regularly used in this specific meaning. The picture of closely related West Germanic dialects remains fragmentary for now. What is common to West Germanic languages is the metonymic shift from MENTAL CONTENT to MIND, which hardly appears in East Germanic languages. The same phenomenon can also be observed in Ancient Greek. Both Old English and Old Saxon, as well as Old High German conceptualize MIND as PHYSICAL SPACE. There are also indications in Old Saxon and Old English that suggest the concept of a MINDS ARE CONTAINERS metaphor building upon it, but the opaque relationships regarding the choice of prepositions and the distinct text traditions make a superficial comparison more challenging. The Gothic data appear to exhibit more archaic patterns. The conceptualization of MIND in Gothic is based on more immediate bodily experiences. The body or specific body parts (especially the heart) function as real, existing physical objects and as location of emotional events. It is therefore not surprising that emotions and desires, which are sometimes also physically perceptible, are primarily associated with MIND, while other mental experiences like mere thoughts appear less frequently. In Gothic biblical texts, we can observe a BODIES ARE MINDS metaphor in the sense that the human body and its components are inherently accessible to more

or less concrete entities (BODIES ARE CONTAINERS). However, there is no abstract MIND CONTAINER derived from MENTAL CONTENT nouns as we can observe in Old High German. If we apply these findings to the situation in Proto-Germanic, we can hypothesize a cognitive developmental path that is reflected diachronically in the history of the Germanic languages: BODIES ARE PHYSICAL SPACES > BODIES ARE CONTAINERS (> BODIES ARE MINDS) > MINDS ARE PHYSICAL SPACES > MINDS ARE CONTAINERS. German 'Come to mind' constructions have been formed following the structure [*in* 'in' N *kommen* 'come'] since at least Old High German times. The frequent use of both abstract entities like MENTAL CONTENT and abstract locations like MINDS shows that COME has always been used in constructions denoting metaphorical or abstract motion. So, it is reasonable to assume that different interpretations of its semantics arise from the context, specifically the various constructions, and only in the interplay of their components. A general path of desemanticization of COME cannot be conclusively explained diachronically. Therefore, it is worthwhile asking how the interplay of the moving entity and abstract location within [*in* 'in' N *kommen* 'come'] changes in the further diachronic course.

4 The Semantic Space of [*in* N *kommen*]

In this section, I present the results of the analysis of the semantic space of the structure [*in* N *kommen*] in the history of German. The previous analyses have shown that in the oldest stage of the German language, there was an established 'Come to mind' construction exploiting this structure that drew on two metaphors: ABSTRACT ENTITIES ARE (MOVING) OBJECTS and MINDS ARE CONTAINERS. The term *semantic space* in the context of the structure [*in* N *kommen*] refers to a model that represents the relationship of all respective constructional components to each other and among themselves. The goal is to find out at which point which subjects are available as moving entities for the VP [*in* N *kommen*] and with which prototypical concrete or abstract locations they can be associated.

4.1 Data and Method

The analysis is based on the corpus data from DDD (Deutsch Diachron Digital), the DTA (Deutsches Textarchiv) and the DWDS Kernkorpus. Tokens were extracted from the corpora using different queries: one for the contexts where *kommen* precedes the prepositional phrase with *in* and one for those contexts where *kommen* follows the prepositional phrase with *in*. The entire respective reference

corpus was used as the basis for the older periods Old High German, Middle High German, and Early New High German. Due to the volume of data, a sample of 1000 query hits was used for each of the two more recent periods. he 19th century was chosen because during this time, a significant increase in similar structures following the pattern [P N V] has been observed (see Smirnova and Stöber 2022: 141). The time span 1980–1999 was chosen to contrast historical data with the most modern ones available. Table 1 provides an overview of the data:

Table 1: Overview of the data [*in* N *kommen*].

Period	Corpus size	Query results	Relevant structures
Old High German[13]	500.000	403	116
Middle High German[14]	2.500.000	1216	649
Early New High Geman[15]	3.500.000	2680	722
1800–1899[16]	Sample	1000	502
1980–1999[17]	Sample	1000	526

Due to the incoherent textual basis for these data, the necessary amalgamation of exhaustive historical corpora and samples from more recent periods, a mere presentation of numerical ratios is not informative. Therefor I applied a Correspondence Analysis (Glynn 2014) as an explorative technique to identify patterns and dependencies among the categories and represent them graphically. It's particularly useful when dealing with large datasets with multiple categorical variables, providing insights into the structure and associations within the data. For this analysis, the two variables MOVING ENTITY and GOAL have been chosen. The first one refers to the syntactic position of the subject and is described based on different degrees of animacy. The set of annotation categories for MOVING ENTITY is given in Table 2:

---

13 lemma = "in" . 1,10 lemma = "queman"; lemma = "queman" . 1,10 lemma = "in"

14 lemma = "in" . 1,10 lemma = "quemen"; lemma = "quemen" . 1,10 lemma = "in"

15 lemma = "in" . 1,10 lemma = "kommen"; lemma = "kommen" . 1,10 lemma = "in"

16 (kommen with $p=VV*) #5 (in with $p=APPR) #3 $p=NN*; (In with $p=APPR) #3 $p=NN* #5 (kommen with $p=VV*); (In with $p=APPR) #3 $p=NN* #5 (kommen with $p=VV*), (@ins with $p=APPRART) #3 $p=NN* #5 (kommen with $p=VV*), (kommen with $p=VV*) #5 (@ins with $p=APPRART) #3 $p=NN*

17 (kommen with $p=VV*) #5 (in with $p=APPR) #3 $p=NN*; (In with $p=APPR) #3 $p=NN* #5 (kommen with $p=VV*); (In with $p=APPR) #3 $p=NN* #5 (kommen with $p=VV*), (@ins with $p=APPRART) #3 $p=NN* #5 (kommen with $p=VV*), (kommen with $p=VV*) #5 (@ins with $p=APPRART) #3 $p=NN*

Table 2: Annotation categories (MOVING ENTITY).

Category	Definition	Example
animate	living beings (humans, animals, gods, etc.)	MHG. *ein grozze person* 'a tall person'
concrete	non-living but perceptible objects	MHG. *gewant* 'clothes'
abstract	non-perceptible entities / concepts	OHG. *fruma* 'benefit'

Animate entities refer to those capable of conscious movement, essentially humans and animals, but also supernatural beings. Concrete entities, in my categorization (following Fleischhauer and Hartmann 2021), are restricted to inanimate objects physically experienceable in our world. Abstract entities encompass everything non-living and non-concrete, such as MENTAL CONTENT, EVENTS or CONCEPTS.

The PP-internal nouns (GOAL) were initially defined based on the two most recent periods, considering not only the noun itself but also its function in the context of the entire construction. 'location' refers to a specific place that can be reached through movement. As 'mind', anything capable of serving as an abstract CONTAINER that holds MENTAL CONTENT is considered. Eventive nouns consist of two groups: 'Internal events' refer to processes where the entering change of state, to which the subject is subjected, is not necessarily attributed to an external action presupposed in the noun. 'External events' require a passive interpretation through an inverse argument structure and an external agent of action that affects the subject, see (28).

(28) *Denn nur er konnte das Gesetz in Antrag bringen, und gegen seinen Willen kein Vorschlag in Berathung kommen.*
 'For only he could propose the law, and no proposal comes into discussion against his will.' (Bluntschli, 1875)

This sentence could be transformed into the passive voice: 'no proposal is discussed against his will.' Another category is 'Concepts'. This occurs when a predominantly abstract entity engages with a collective principle, idea, or event. Notably, in this case, the GOAL can undergo a change and the MOVING ENTITY is not affected as in the case of internal and external events, see (29).

(29) *Zum anderen ist Bewegung in die ästhetische Diskussion dadurch gekommen, daß die Literatur- und Kunstsoziologie in den letzten Jahren einen Aufschwung erfahren [. . .] hat.*
 'On the other hand, movement has come into the aesthetic discussion through the fact that the sociology of literature and art has experienced an upswing in recent years [. . .].' (Zimmermann, 1985)

The set of annotation categories for GOAL is given in Table 3:

Table 3: Annotation categories (GOAL).

Category	Definition	Example
location	specific place or space	ENHG. *closter* 'monastery'
mind	application of the MINDS ARE CONTAINERS metaphor	MHG. *sin* 'sense, mind'
event (internal)	application of the STATES / EVENTS ARE CONTAINERS metaphor without an external force	OHG. *kust* 'well being'
event (external	application of the STATES / EVENTS ARE CONTAINERS metaphor with an external force	NHG. *Verwendung* 'usage'
concept	non-perceptible an non-eventive entities or concepts	NHG. *Literatur* 'literature'

4.2 Results

The initial situation in Old High German is as follows: There are exclusively animate and abstract subjects, but no concrete ones. An application of Correspondence Analysis was not possible for this reason, as more than two categories must be available for each variable. Concrete animate entities can come into concrete locations (30). Abstract entities, accounting for almost a quarter of all instances in Old High German, can come into abstract CONTAINERS (31).

(30) *Thô quam ther heilant in Galileu.*
      'Then the savior came to Galilee.' (Tatian, 14, 1)

(31) *thie dáti mir quément in githáhti*
      'The deeds came to my mind' (Otfrid, III, 1, 8)

This binary distribution within the semantic space of [*in* N *kommen*] essentially applies to the oldest texts in German. Only towards the end of the Old High German period three eventive nouns can be found in PP-internal position. All constructions are associated with animate subjects, see (32)–(34):

(32) *chúmet er in ándere chúst*
'He gets into another state (of mind)' (Notker, Boethius, 126)

(33) *Chúmet er in_frêisun. er hilfet ímo dar ûz*
'If he comes in danger, he helps him out again' (Notker, Psalter, 36, 2)

(34) *úbe ih chómen sol in putrefactionem* (Notker, Psalmen)
'Once I am to get into decay'

(32) and (33) are genuine German light verb constructions in the narrower sense. The subject undergoes an emotional change of state. The author makes use of two established eventive nouns, which, synchronically, are not deverbal. Such a noun is obviously not available to him in (34), which is why he instead uses Lat. *putrefactio* 'decay'. In comparison to more recent periods of the German, Old High German has significantly fewer deverbal nouns. This demonstrates that the presence of potential slot fillers plays a crucial role in establishing a syntactic construction. Referring to the research literature, Fleischhauer and Hartmann (2021) mention three additional light verb constructions containing an eventive noun in PP-internal position in Old High German that do not appear in the DDD: *in angest queman* 'frighten' (lit. 'come into fear'), *in forhtūn queman* 'fear' (lit. 'come into fear'), *in scama queman* 'be upset' (lit. 'come into consternation'). These are also combined with animate subjects.

The fundamentally binary structuring of the semantic space is still reflected in Middle High German, for which the results of the analysis are depicted in a contingency table in Figure 1. This archaic system is only disrupted in Early New High German, as Figure 2 illustrates.

The observable patterns in Middle High German reveal a predominant clustering of locations and (internal) events around animate subjects on the left side of the plot. The 'Come to mind' construction is located on the right side of the plot. Abstract entities still avoid goals other than mind. A notable development in Middle High German is the emergence of inanimate subjects, which exclusively appear in combination with animate subjects. Inanimate subjects share the same kind of goals as animate ones, following the conceptual regions already occupied by the animate subjects.

(35) *Swelh man dem andᵉn eín pherít. odˢ eín gewant. odˢ deheín varende guͦt ſetzet. odˢ lihet mít ſínē willen. ſo hat íener reht dar an. ín dez gewalt ez alſo kvmet.*
'Whoever transfers or lends a horse, a clothing or any movable property to another with his will, he has a right to it in whose power it thus comes.'
(Schwabenspiegel, 108vb, 2)
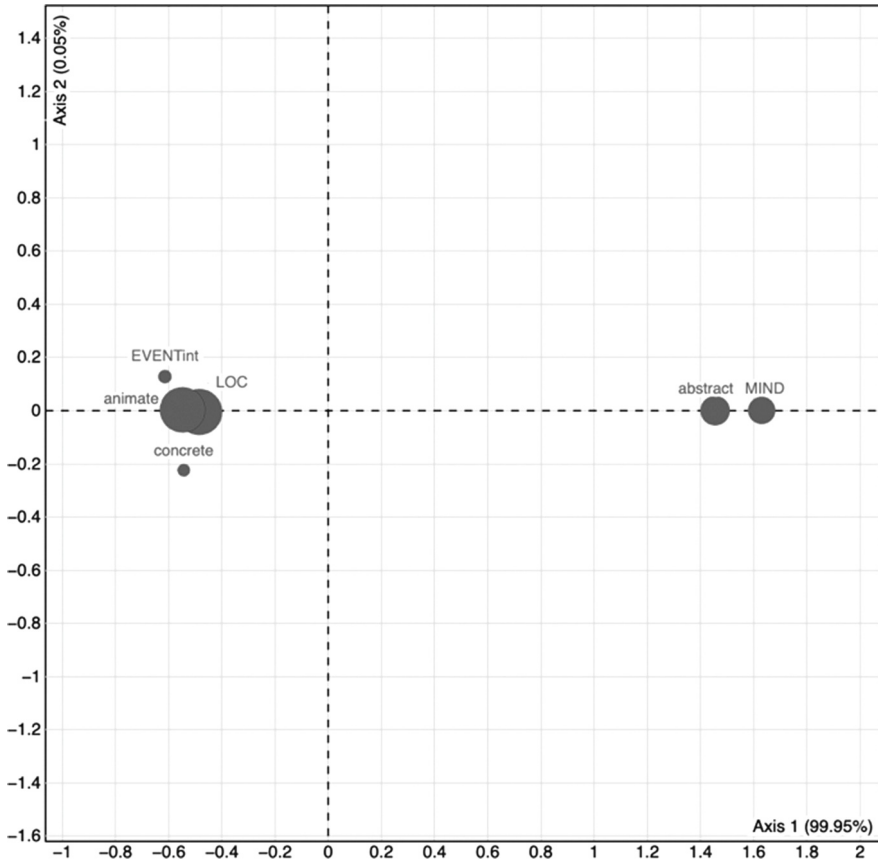
Figure 1: Semantic space of [*in* N *kommen*] (MHG.).

However, inanimate nouns in subject position remain an exception in Middle High German. The axis weighting within the graph indicates that over 99% of the data can still be explained through the Old High German dichotomy. The picture changes drastically in Early New High German. The observed transformation reveals a shift from a two-part system to a triangle. Animated subjects still exhibit fidelity to their established LOCATIONS and internal EVENTS. Abstract entities, predominantly confined to their MIND CONTAINER, exhibit a slight expansion with the introduction of a new category involving abstract CONCEPTS as PP-internal GOALS, see (36):

(36)　*Darnach aber nach langer zytt das die zyt der natur kam in die zytt der gnaden*
　　　'Afterwards, but after a long time, that the time of nature came into the time of grace.' (Antichristdrama, 102va,21)
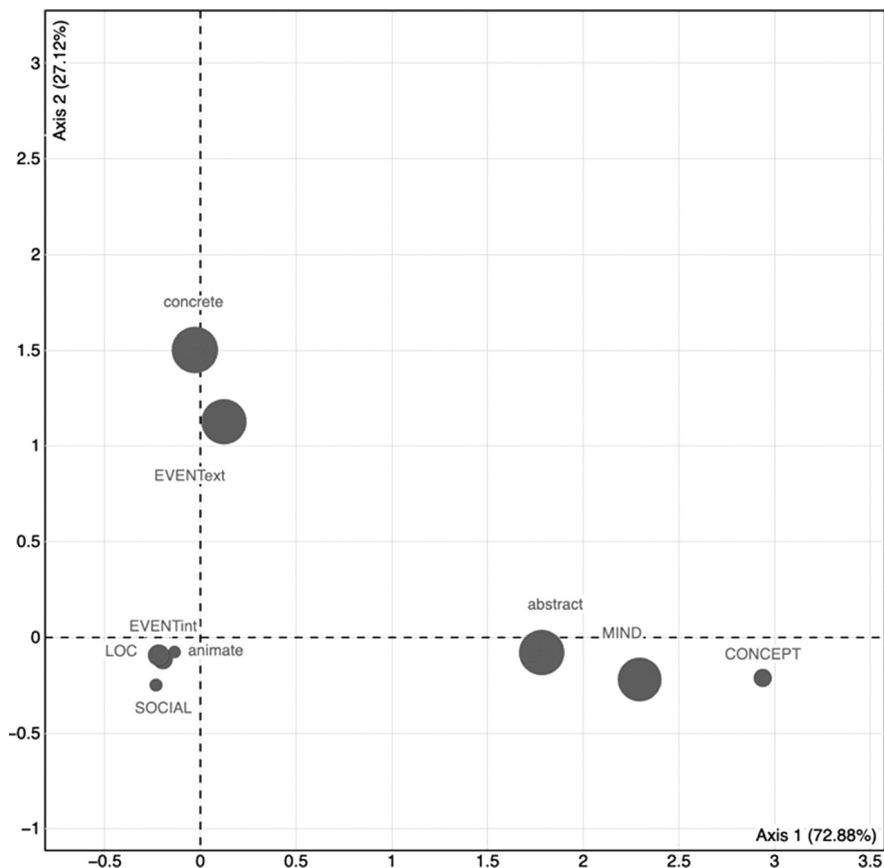
Figure 2: Semantic space of [*in* N *kommen*] (ENHG.).

The most significant change compared to older periods pertains to the substantial increase of inanimate subjects strongly associated with external events.

(37) *vnd hernoch volgende gutere sein Jn die teilunge komen.*
      'And after that the following goods have come into division.' (6. Dresdner Stadbuch, 004va,22)

This increase is associated with the rise in deverbal nouns (or nouns with verbal cognates), such as *Teilung* 'division,' *Abschleif* 'cut-off,' *Aufschlag* 'surcharge,' Schaden 'damage,' *Nutz* 'benefit,' or *Regierung* 'government'. Most of these forma- tions have transitive base verbs that allow for a passive interpretation of the con-

struction, favoring the use of inanimate or abstract subjects that are not capable of conscious movement.

In the 19th century abstract entities gradually become more associated with external EVENTS, converging toward the domain of inanimate subjects. The overall semantic space condenses to some extent and becomes less structured, leading to the dissolution of clear restrictions. The evolving dynamics suggest the formation of a novel cluster, a trend that becomes more apparent in the 20th century. Notably, the role of the MIND container diminishes over time, gradually receding, while abstract entities draw closer to external EVENTS, see Figures 3 and 4.
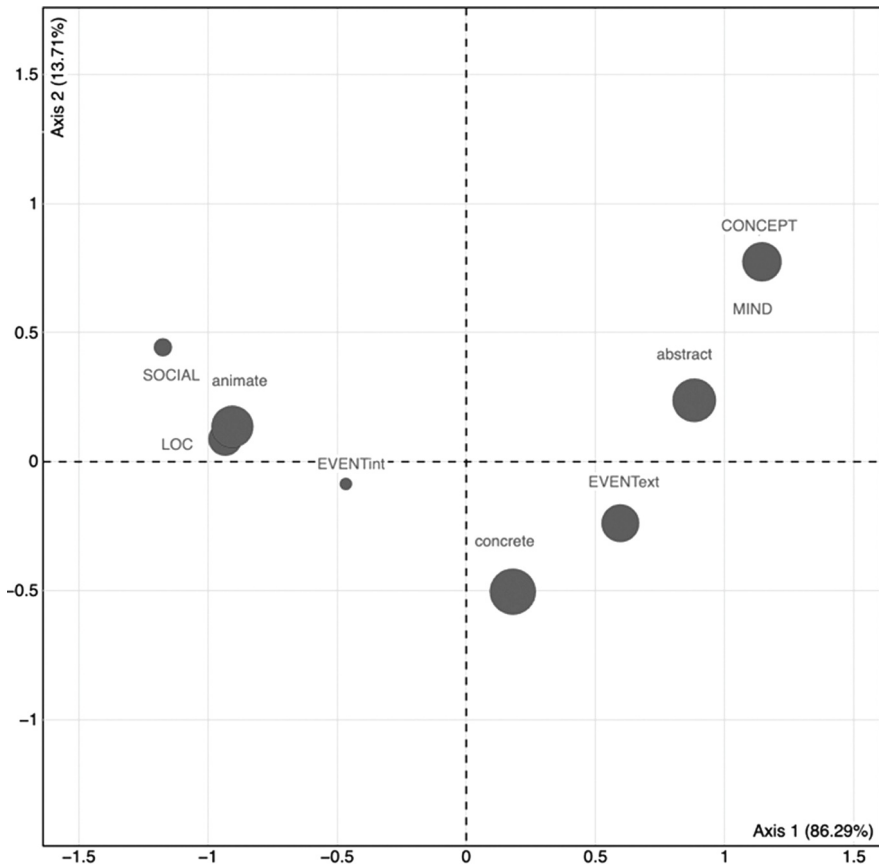
Figure 3: Semantic space of [*in* N *kommen*] (1800–1899).
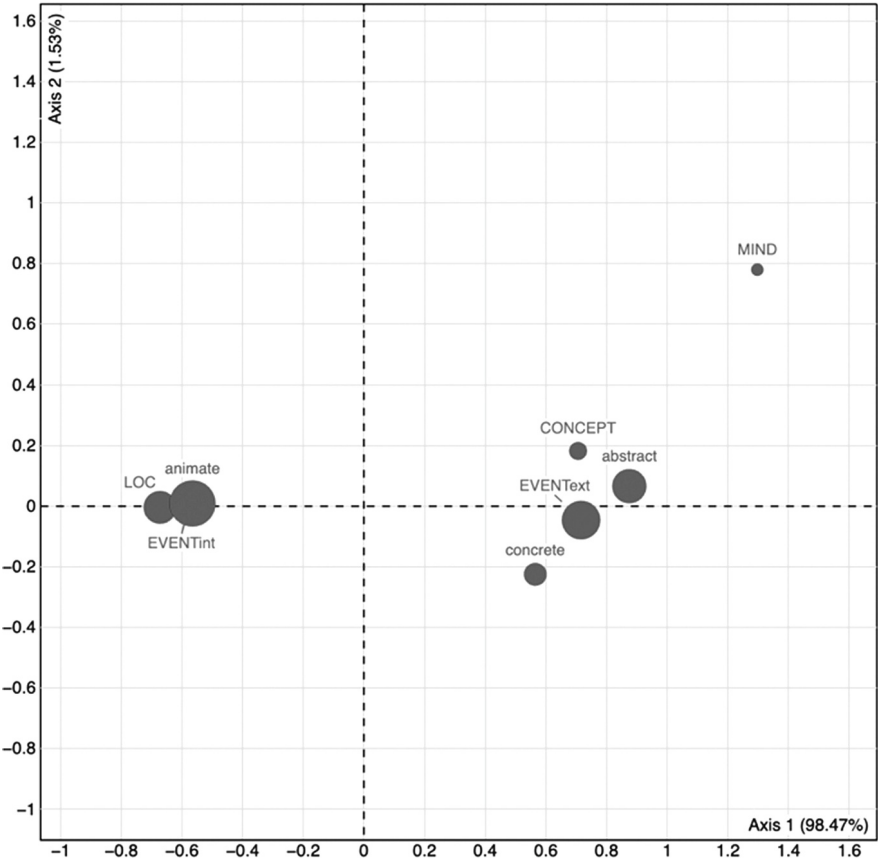
Figure 4: Semantic space of [*in* N *kommen*] (1980–1999).

The semantic space undergoes a rearrangement, returning to a two-part system; however, this time, abstract and inanimate entities constitute a shared associative group with external EVENTS. The most recent data reveal a distinct peripheral position of the 'Come to mind' construction. A brief qualitative examination of 'Come to mind' construction formed following the pattern of the VP [*in* N *kommen*] reveals that the diminishing presence of the construction in the semantic space is not only a matter of frequency. Since Middle High German, there has been a clear downward trend in the number of distinct types of MIND nouns used within the construction, see Table 4:

Table 4: MIND nouns within [*in* N *kommen*] in the history of German.

OHG	MHG	ENHG	1800–1899	1980–1999
sin	*sin*	*sinn*	*Sinn*	*Sinn*
herz	*herz*	*herz*	*Herz*	
	kopf	*haupt*	*Kopf*	
	sel	*sele*		
muot	*muot*	*muat*		
gimuot	*gemuot*	*gemüet*		
gidrahti	*getrecht*	*getrecht*		
gidahti	*gedanc*	*gedenke*		
wan	*wan*			
	busen			
	munt			
	man	*mensch*		
gitrahta				
		wille		
HUMAN BODY	HUMAN BODY	HUMAN BODY		

In modern German, *Sinn* is the sole survivor among lexemes used to denote the concept of MIND within [*in* N *kommen*]. This survival is no coincidence, but rather aligns with the broader pattern observed in West Germanic languages, where MENTAL CONTENT primarily pertains to less salient subjects like thoughts rather than emotions or states of being. The decreasing salience is given in that the nominal pool of MENTAL CONTENT that can be placed in a CONTAINER is potentially almost unlimited. *Sinn* with its basic semantics is at the top of this cognitive development, while competing nouns are more sensitive in terms of possible MOVING ENTITIES. This applies in particular to all types of body part metaphors, which necessarily have a certain affinity to specific emotions or mental states that can also be reflected in concrete physical experiences. As the analysis of the historical data has shown, this group of nouns reflects more archaic concepts that are slowly being overcome by the identified metonymic shift towards the general semantics of MIND as abstract PHYSICAL SPACE in West Germanic.

5 Conclusion

The history of the structure [*in* N *kommen*] exhibits a complex and multifaceted pattern, characterized by the interplay of compatibility between MOVING ENTITY and GOAL, dependent on the overall semantics of each construction. In the case of 'Come to mind' constructions, the question does not arise from the start, as only

abstract entities can naturally enter the MIND container. From the very beginning, abstract entities appear in subject position, albeit restricted in the function of a 'Come to mind' construction. The prerequisite for this is the identified West Germanic metonymic shift from MENTAL CONTENT to MIND, traceable in Old High German and closely related languages. The distinctive feature of Old High German, in contrast to other West Germanic languages, is its consistent metaphorical development from MINDS ARE PHYSICAL SPACES to MINDS ARE CONTAINERS. The evolution of constructions that involve an EVENT as a GOAL diachronically follows a salience hierarchy on the animacy scale: Animated subjects are the innovative pioneers, followed later by concrete entities, and ultimately, abstract entities. This finding is relevant to the development of light verb constructions. The increase in abstract subjects in these constructions may be considered a sign of the 'desemanticization' of the verb *kommen*. But this only applies to this specific group of constructions involving an eventive noun. In this regard, the fate of the 'Come to mind' construction is contradictory. The longstanding potential for the abstract interpretations of *kommen* is evident not only in 'Come to mind' constructions but also in sporadic instances of light verb constructions in Old High German. The diachronically different behavior of the individual constructions could probably be explained on the basis of available lexical items for the favored purpose at a specific time. While the number of MENTAL CONTENT nouns reaches its peak in Middle High German, German light verb constructions do not thrive until new kinds of deverbal word formation were established from Middle High German onwards.

In contrast to light verb constructions, 'Come to mind' constructions cannot maintain their once central position in the semantic space and are progressively marginalized throughout history. The development of the two constructions is, therefore, divergent and cannot be uniformly explained. It must be assumed that different constructions also have different histories. The availability of possible elements within any structure can always be exploited to the maximum unless there are impediments to the process. The creative enrichment of the structure is thus stimulated and demanded by the availability of morphologically and lexically appropriate material. The solidified structure attracts potential elements, which leads to a high variance with respect to the noun slot. But there is obviously a discernible decline in interest as individuals exhaust all possible metaphors, leading to a declining charm of extravagance. The metonymic shift MENTAL CONTENT > MIND could be hypothesized as the reason for this. This cognitive precondition gave the noun *Sinn* the opportunity to assert itself over competing nouns. While the reservoir of new MIND nouns depletes early, other constructions such as light verb constructions, specifically those involving a change of state, benefit from a more extensive supply of new material for specific usage contexts,

particularly deverbal nouns. This could save light verb constructions from an 'evaporation effect' to which constructions with lower lexical potential fall victim.

In conclusion, I propose that the number of abstract nouns in subject position within [*in* N *kommen*] is not solely determined by a uniform two-step process of desemanticization and an increase in eventive nouns in PP-internal position. Instead, the variation observed depends on multiple factors, including the semantics of the specific construction, its underlying metaphor, the availability of suitable lexical material and pragmatic considerations. Taking these factors into account will provide a more nuanced and comprehensive understanding of the emergence and evolution of [*in* N *kommen*] and similar structures in German.

References

Anderson, Stephen R. 2015. Morphological change. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 264–285. London: Routledge.

Barðdal, Jóhanna. 2001. The Role of Thematic Roles in Constructions? Evidence from the Icelandic Inchoative. In Arthur Holmer, Jan-Olof Svantesson & Åke Viberg (eds.), *Proceedings of the 18th Scandinavian Conference of Linguistics 2000*, 127–137. Lund: Department of Linguistics.

Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. Hungering and lusting for women and fleshly delicacies: Reconstructing grammatical relations for Proto-Germanic. *Transactions of the Philological Society* 110 (3). 363–393.

Barðdal, Jóhanna & Spike Gildea. 2015. Diachronic Construction Grammar: Epistemological Context, Basic Assumptions and Historical Implications. In Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.), *Diachronic Construction Grammar*, 1–50. Amsterdam: John Benjamins.

Johnson, Cynthia, Peter Kerkhof, Leonid Kulikov, Esther Le Mair & Jóhanna Barðdal. 2019. Argument Structure, Conceptual Metaphor and Semantic Change: How to Succeed in Indo-European without Really Trying? *Diachronica* 36, 463–508.

Barnden, John. 1997. Consciousness and Common-Sense Metaphors of Mind. In Seán Ó Nualláin, Paul Mc Kevitt & Eoghan Mac Aogáin, (eds.), *Two Sciences of Mind*, 311–340. Amsterdam & Philadelphia: Benjamin.

Bréal, Michel. 1900. *Semantics. Studies in the Science of Meaning*. London: Heinemann.

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

Bybee, Joan & Sandra Thompson. 2000. Three Frequency Effects in Syntax. *Berkeley Linguistic Society* 23. 65–85.

Campbell, Lyle & Richard Janda. 2001. Introduction: Conceptions of grammaticalization and their Problems. *Language Sciences* 23. 2–3, 93–112.

Cornillie, Bert. 2008. On the Grammaticalization and (Inter)Subjectivity of Evidential (Semi-)Auxiliaries in Spanish. In Elena Seoane & María José López-Couso (eds.), *Theoretical and Empirical Issues in Grammaticalization*, 77–110. Amsterdam: John Benjamins.

Coussé, Evie, Peter Andersson & Joel Olofsson. 2018. Grammaticalization meets Construction Grammar: Opportunities, challenges and potential incompatibilities. In Evie Coussé, Peter

Andersson & Joel Olofsson (eds.), *Grammaticalization meets Construction Grammar*, 3–19. Amsterdam: John Benjamins.

Dewey, Tonya Kim & Carlee Arnett. 2015. Motion Verbs in Old Saxon with the Oblique Subject Construction: A Semantic Analysis. *Beiträge Zur Geschichte der Deutschen Sprache und Literatur* 137 (2), 183–220.

Diewald, Gabriele. 1999. *Die Modalverben im Deutschen: Grammatikalisierung und Polyfunktionalität*. Tübingen: Niemeyer.

Diewald, Gabriele & Elena Smirnova. 2010. *Evidentiality in German: Linguistic Realization and Regularities in Grammaticalization*. Berlin: Mouton de Gruyter.

Diewald, Gabriele & Ilse Wischer. 2013. Markers of Futurity in Old High German and Old English: A Comparative Corpus-Based Study. In Gabriele Diewald, Leena Kahlas- Tarkka & Ilse Wischer (eds.), *Comparative Studies in Early Germanic Languages: With a Focus on Verbal Categories*, 195–216. Amsterdam: John Benjamins.

Ellison, T. Mark & Uta Reinöhl. 2022. Compositionality, Metaphor, and the Evolution of Language. *International Journal of Primatology* 23.

Filatkina, Natalia. 2018. *Historische formelhafte Sprache: Theoretische Grundlagen und methodische Herausforderungen*. Berlin & Boston: De Gruyter.

Fillmore, Charles. 1987. How to Know Whether You're Coming or Going. In Gisa Rauh (ed.), *Essays on Deixis*, 219–227. Tübingen: Narr.

Fillmore, Charles. 1997. *Lectures on Deixis*. Stanford: CSLI Publications.

Fleischhauer, Jens & Mozhgan Neisani. 2020. Adverbial and attributive modification of Persian separable light verb constructions. *Journal of Linguistics* 56. 45–85.

Fleischhauer, Jens & Stefan Hartmann. 2021. The emergence of light verb constructions: A case study on German kommen 'come'. *Yearbook of the German Cognitive Linguistics Association* 9/1. 135–156.

Fleischman, Suzanne. 1983. *The Future in Thought and Language: Diachronic Evidence from Romance*. Cambridge: Cambridge University Press.

Gildea, Spike & Jóhanna Barðdal. 2023. From Grammaticalization to Diachronic Construction Grammar: A Natural Evolution of the Paradigm. *Studies in languages* 47 (4). 743–788.

Glynn, Dylan. 2014. Correspondence analysis: Exploring data and identifying patterns. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, 443–485. Amsterdam: John Benjamins.

Goatly, Andrew. 1997. *The Language of Metaphors*. London: Routledge.

Goatly, Andrew. 2007. *Washing the Brain: Metaphor and Hidden Ideology*. Amsterdam: John Benjamins.

Haan de, Ferdinand. 2007. Raising as Grammaticalization: The Case of Germanic SEEM verbs. *Rivista di Linguistica* 19.1. 129–150.

Harm, Volker. 2021. *Funktionsverbgefüge des Deutschen: Untersuchungen zu einer Kategorie zwischen Lexikon und Grammatik*. Berlin & Boston: De Gruyter.

Heine, Bernd, Ulrike Claudi & Friederike Hünnemayer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.

Heine, Bernd. 1993. *Auxiliaries: Cognitive Forces and Grammaticalization*. Oxford: Oxford University Press.

Heine, Antje. 2020. Zwischen Grammatik und Lexikon. Forschungsgeschichtlicher Blick auf Funktionsverbgefüge. In Sabine De Knop & Manon Hermann (eds.), *Funktionsverbgefüge im Fokus*, 15–38. Berlin & Boston: De Gruyter.

Hilpert, Martin. 2008. *Germanic Future Constructions: A Usage-based Approach to Language Change*. Amsterdam: John Benjamins.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

Kamber, Alain. 2008. *Funktionsverbgefüge – empirisch*. Tübingen: Niemeyer.

Kiparsky, Paul. 2012. Grammaticalization as optimization. In Dianne Jonas, John Whitman & Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcome*, 15–51. Oxford: Oxford University Press.

Krug, Manfred. 2011. Auxiliaries and grammaticalization. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of Grammaticalization*, 547–558. Oxford: Oxford University Press.

Lakoff, George. 1993. Contemporary Theory of Metaphor. In Andrew Ortony (ed.), *Metaphor and Thought*, 202–251. Cambridge: Cambridge University Press.

Lakoff, George. 1987. Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. Chicago: University of Chicago Press.

Lakoff, G. & M. Johnson (1999). *Philosophy in the Flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Lakoff, George & Mark Johson. 2003. *Metaphors we live by*. Chicago: Chicago University Press.

Langacker, Ronald W. 1991. *Foundations of cognitive grammar II: Descriptive application*. Stanford, CA: Stanford University Press.

López-Couso, María José & Belén Méndez-Naya. 2015. Epistemic/Evidential Markers of the Type Verb + Complementizer: Some Parallels from English and Romance. In Andrew Smith, Graeme Trousdale & Richard Waltereit (eds.), *New Directions in Grammaticalization Research*, 93–120. Amsterdam: John Benjamins.

Marcinkowska-Rosół, Maria & Sven Sellmer. 2021. The Mind as Container: A Study of a Metaphor in Homer and Hesiod with a Parallel Analysis of the Sanskrit Epics. *Mnemosyne* 75 (6). 873–897.

Meillet, Antoine. 1912. L'evolution des formes grammaticales. *Revue internationale de synthese scientifique* 12. 6.

Polenz, Peter von. 1987. Funktionsverben, Funktionsverbgefüge und Verwandtes: Vorschläge zur satzsemantischen Lexikographie. *Zeitschrift für Germanistische Linguistik* 15 (2).169–189.

Pottelberge, Jeroen Van. 2001. *Verbonominale Konstruktionen, Funktionsverbgefüge: vom Sinn und Unsinn eines Untersuchungsgegenstandes*. Heidelberg: Winter.

Reznikova, Tatiana, Ekaterina Rakhilina & Anastasia A. Bonch-Osmolovskaya. 2012. Towards a Typology of Pain Predicates. *Linguistics* 50 (3). 421–465.

Roberts, Ian. 1993. A formal account of grammaticalization in the history of Romance futures. *Folia Linguistica Historica* 13. 219–258.

Schrodt, Richard. 2004. *Althochdeutsche Grammatik II: Syntax*. Tübingen: Niemeyer.

Schrodt, Richard & Karin Donhauser. 2003. Tempus, Aspekt/Aktionsart und Modus im Deutschen. In Werner Besch (ed.), *Sprachgeschichte, 3. Teilband: Ein Handbuch Zur Geschichte der Deutschen Sprache und ihrer Erforschung*, 2504–2525. Berlin & Boston: De Gruyter.

Schützeichel, Rudolf. 2006. *Althochdeutsches Wörterbuch*. Tübingen: Niemeyer.

Seifert, Jan. 2016. Zu erkanntnuss gebracht: Frühneuhochdeutsche Funktionsverbgefüge – Sprachliche Strukturen in der Diglossiesituation. In Claudia Wich-Reif (ed.), *Historische Sprachkontaktforschung*, 165–187. Berlin & Boston: De Gruyter.

Seifert, Jan. 2020. Funktionsverbgefüge in älteren Texten: Topologische, prosodische und informationsstrukturelle Aspekte. In Delphine Pasques & Claudia Wich-Reif (eds.), *Textkohärenz und Gesamtsatzstrukturen in der Geschichte der deutschen und französischen Sprache vom 8. bis zum 18. Jahrhundert*, 397–415. Berlin: Weidler.

Smirnova, Elena & Vanessa Stöber. 2020. Verbo-nominal constructions with *kommen* 'come' in German. *Constructions and Frames* 14 (1). 121–149.

Sturtevant, Edgar Howard. 1917. *Linguistic Change: An Introduction to the Historical Study of Language*. Chicago: University of Chicago Press.

Sweetser, Eve. 1991. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.

Taverniers, Mirian. 2018. Grammatical metaphor and grammaticalization. The case of metaphors of modality. *Functions of Language* 25 (1). 164–204.

Tendahl, Markus & Raymond W. Gibbs Jr. 2008. Complementary perspectives on metaphor: Cognitive linguistics and relevance theory. *Journal of Pragmatics* 40 (11). 1823–1864.

Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes* (Oxford Studies in Diachronic and Historical Linguistics 6). Oxford: Oxford University Press.

Ullmann, Stephen. 1951. *The Principles of Semantics*. Glasgow: Jackson.

Ullmann, Stephen. 1962. *Semantics: An Introduction to the Science of Meaning*. Blackwell: Oxford.

Viberg, Åke. 1983. The Verbs of Perception: A Typological Study. *Linguistics* 21 (1). 123–162.

Wundt, Wilhelm Max. 1904. *Völkerpsychologie: eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte*. Leipzig: Engelmann.

Wegener, Heide. 2001. Verbs of Affect from a Synchronic and a Diachronic Perspective. In Nicole Dehé & Anja Wanner (eds.), *Structural Aspects of Semantically Complex Verbs*, 219–248. Frankfurt: Peter Lang.

Primary Sources

[Heliand] Heliand und Genesis. Ed. by Otto Behaghel. Bearbeitet von Burkhard Taeger. Tübingen: Niemeyer 1984 (= Altdeutsche Textbibliothek; 4).

[Notker] Die Werke Notkers des Deutschen. Herausgegeben von James C. King und Petrus W. Tax. Neue Ausgabe. Band 1–10A. Tübingen: Niemeyer 1972 (= Altdeutsche Textbibliothek; 73).

[Otfrid] Otfrids Evangelienbuch. Herausgegeben von Oskar Erdmann 1882. Sechste Auflage, besorgt von Ludwig Wolff. Tübingen: Niemeyer 1973 (= Altdeutsche Textbibliothek; 49).

[Tatian] Tatian. Lateinisch und altdeutsch mit ausführlichem Glossar. Herausgegeben von Eduard Sievers. Unveränderter Nachdruck. Paderborn: Schöningh 1966 (= Bibliothek der ältesten deutschen Literatur-Denkmäler; 5).

[Wulfila] Die gotische Bibel. Herausgegeben von Wilhelm Streitberg. 1. Teil. Der gotische Text und seine griechische Vorlage mit Einleitung, Lesarten und Quellennachweisen sowie den kleinen Denkmälern als Anhang. Heidelberg: Winter 1919 (= Germanische Bibliothek; 2/3).

Regina Ruf & Elena Smirnova

Mit Hilfe von Kreativität?

How German Complex Prepositions Are Shaped

Abstract: This paper deals with the formation and the diachronic development of German complex prepositions following the general syntactic pattern [P N P/GEN] and sets to answer two general questions: (i) Do individual complex prepositions follow a general pattern of formation? And, (ii) if yes, how active and productive is this pattern? Traditionally, a diachronic scenario of a reanalysis followed by the gradual conventionalization and routinization of individual elements is assumed to be at work for individual instances of complex prepositions. The present study will show that complex prepositions are often formed directly after a productive schema by filling its open slots with lexical material. In this case, a diachronic scenario involving the creative application of a pattern that gets extended and strengthened over time is more likely. As we will show, the data reveals a differentiated picture, suggesting that both diachronic ways have been taken on the road towards complex prepositions in German.

1 Introduction: Complex Prepositions in German

Complex prepositions (henceforth CPs) in German are usually described as an open class of multi-word expressions (cf. e.g. Di Meola 2000; Szczepaniak 2009; Stefanowitsch et al. 2020). However, it has been notoriously difficult to provide a precise definition of this class, as the boundaries to the simple prepositions on the one hand and to free syntactic combinations on the other hand are often blurred (see e.g. Lehmann and Stolz 1992: 17; Meibauer 1995).

The majority of CPs in present-day German are realizations of the syntactic pattern containing a simple preposition (=P1), an optional determiner (= DET), and a noun (N), that are followed by another preposition (=P2) or a genitive noun phrase (=GEN). This structure is exemplified in (1), and will be referred to as [P1 (Det) N P2/GEN] throughout the article.

(1) in_{P1} $Bezug_N$ auf_{P2}, im_{P1} $Zusammenhang_N$ mit_{P2}, mit_{P1} $Hilfe_N$ von_{P2}/GEN, im_{P1} $Vergleich_N$ zu_{P2}
        'in relation to, in connection with, with the help of, in comparison with'

In present-day German, most "newcomers" into the class of CPs are formed according to this structure. In the following, the principal characteristics of CPs of this syntactic pattern will be shortly introduced (for the complete list of criteria see Beneš 1974: 34–35; Lehmann and Stolz 1992; Lindqvist 1994; Di Meola 2000; Szczepaniak 2009; Hüning et al. 2020; see also Quirk and Mulholland 1964: 65; Huddleston 1984; Quirk 1985: 671–672; Huddleston and Pullum 2002 for English). Most researchers argue, and we agree with them, that none of the characteristics discussed in the literature may be taken as a unique defining property of a CP, as individual members usually display only a bundle of these properties and sometimes to a different degree. Nevertheless, these characteristics are important to delimit the class of CPs from regularly formed syntactic combinations.

First, in typical CPs of the structure [P1 (Det) N P2/GEN], the first preposition P1 is usually invariable for a given multi-word item, cf. (2).

(2)   *im/*mit/*bei Zusammenhang mit; im/*unter/*mit Gegensatz zu*
        'in/*with/*at connection with; in/*under/*with contrast to'

The second, postnominal preposition P2 or the genitive case of the noun phrase are also widely taken as being fixed, see (3).

(3)   *im Zusammenhang mit/*an/*auf/*zu; im Gegensatz zu/*mit/*auf*
        'in connection with/*at/*up/*to; in contrast to/*with/*up'

Normally, the genitive phrase can be substituted by a prepositional phrase with *von* 'of'; this happens not only in contexts of CPs but also in other syntactic contexts outside of CPs.

With respect to the nominal core [(Det) N] of the pattern, most of the literature cited above notes that the noun cannot be modified syntactically, for exam-

ple by adjectives and determiners, or morphologically, for example by plural marking, see e.g. (4) for the CP *im Zuge* 'in the course' (cf. also Di Meola 2000: 105–109).

(4) a. *Im Zuge der Verhandlungen hat sich eine Kompromisslösung abgezeichnet.*
         'In the course of the negotiations, a compromise solution has emerged.'
     b. determiner
         *In dem Zug der Verhandlungen . . .
         'In the course of negotiations . . .'
     c. adjective
         *Im schnellen Zug der Verhandlungen . . .
         'In the quick course of negotiations . . .'
     d. plural marking
         *In den Zügen der Verhandlungen . . .
         'In the courses of negotiations . . .'

Generally, the characteristic properties of CPs introduced above are associated with the formal fixedness of these multi-word expressions. From a diachronic point of view, the formal fixedness is usually seen as resulting from a gradual process of automatization and conventionalization leading from regular syntactic combinations towards fixed expressions, via repetition in usage. During this diachronic process, the individual elements of the previously freely combined syntactic structure successively become less autonomous, gradually coalescing into a holistic and idiosyncratic structure. The mechanism of reanalysis (or rebracketing) is assumed to be mainly at work here, and this can be exemplified in (5).

(5) a. P1 N P2
         [*in* [*Bezug* [*auf* NP]$_{PP}$]$_{NP}$]$_{PP}$ > [[*in Bezug auf*]$_{CP}$ [NP]]$_{PP}$
     b. P N GEN
         [*im* [*Laufe* [NP GEN]$_{NP}$]$_{NP}$]$_{PP}$ > [[*im Laufe*]$_{CP}$ [NP GEN]]$_{PP}$

For German, it is generally believed that most recently emerged CPs follow the pattern [P1 (Det) N P2/GEN] and originate from regular syntactic structures. The postnominal genitive phrase in (5b) or the prepositional phrase in (5a) are traced back to the originally postnominal attributive phrases that are then reanalyzed as a complement of a new CP (see Lehmann 1998 for the detailed account of the reanalysis). However, there are some problems with this traditional assumption, and we will show in the remainder of this paper that only some of the CPs in German have arisen via this diachronic path. Many present-day CPs, on the other

hand, do not follow this directionality and show little or no evidence of the "free syntactic" stage prior to their establishment as CPs.

The paper is structured as follows. In Section 2, we will introduce our research questions. Section 3 presents data and methods. Section 4 summarizes the results of the corpus analysis. In Section 5, the implications of the results obtained in this study for the general diachronic development of CPs in German will be discussed, with particular attention to the concepts routine, routinization, and creativity. Section 6 concludes with some open questions for further research.

2 Research Questions

From a diachronic point of view, the category of CPs in German is usually conceived of as a constantly expanding open class. As mentioned in Section 1 above, it is generally assumed that most recently emerged CPs follow the syntactic pattern [P1 (Det) N P2/GEN] and stem from regular syntactic combinations with compositional semantics. The reanalysis schema introduced in (5) above takes this regular syntactic structure with a postnominal attributive (noun or prepositional) phrase of the core noun as input, and turns it into a fixed output structure where this postnominal attributive phrase functions as a complement of a new CP.

The problem with this view is however that, as demonstrated by Hoffmann (2005) for English, for many CPs there is no convincing empirical evidence of a development from a free combination to a fixed item. In English, most of them appear in the written record abruptly in the target form [P1 (DET) N P2]. This suggests that at least some CPs may have emerged by analogy to existing individual cases, which in turn had emerged previously in the way outlined in (5). Hoffmann called this process 'grammaticalization by analogy' (2005: 86). At another point in his study, he evokes the existence of an 'abstract construct': "the sequence 'preposition–noun–preposition' appears to be available as a grammaticalized yet abstract construct which under certain circumstances can be filled by new lexical entities to form a new complex preposition" (Hoffmann 2005: 171). In other words, what Hoffmann (2005) suggests is what we could call, in constructionist terms, a schematic and productive construction. The existence of such a constructional schema means, among others, that speakers can create new CPs spontaneously by filling the variable slots [P1] and [N] with lexical material. And this is what seems to have happened to many of the CPs in English. Given this situation in English, it is fair to ask whether this scenario may also hold for the development of CPs in German. In particular, the [N]-slot in the schema [P1 (Det) N P2/GEN] seems a very good candidate, as nouns from the same or similar semantic

class may easily enter the pattern and thus expand the range of possible members of the open class of CPs.

In diachronic terms, such a productive schema may be considered a booster of creativity, namely the F-creativity in the sense of Sampson (2016), meaning that speakers access the existing pattern and (re-)use it to form novel expressions, thus expanding the inventory of CPs. In Smirnova & Sommerer (2020), a first sketch has been offered of how the category of CPs in German may be modeled in terms of a constructional network with several levels of abstraction and with different sub-schemas within the general schema [P1 (DET) N P2/GEN] (see Figure 1).
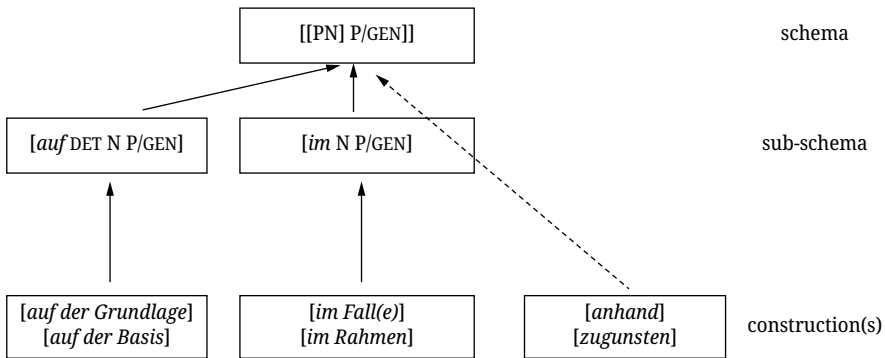
Figure 1: Partial network of German complex prepositions (slightly adapted from Smirnova & Sommerer 2020: 16).

Figure 1 represents the partial network modelled on the basis of a handful of individual CPs which differ from each other with respect to the formal pattern and the semantic content. Smirnova & Sommerer (2020: 17) suggested that synchronically, several constructional sub-schemas in the network can be assumed, which function as generalizations across different individual construction types.

The present study offers a dynamic and diachronic take on the phenomenon of CPs. It is guided by the central research questions concerning the existence of constructional schemas at work in the diachronic development of CPs in German. In other words:

– Do individual CPs follow a general pattern of formation in German?
– If yes, how can this pattern be best modelled in a constructional network?
– If yes, how productive is this pattern in the sense that it regularly produces new CPs?

These research questions are directly linked to the more general issues of creativity and routine in diachronic change. If the traditional view is correct and German CPs predominantly arise via the reanalysis of a previously regular syntactic combination, as it has been sketched in (5) above, one would expect a rather slow diachronic emergence of a small number of CPs in the course of time. This would support a diachronic scenario of a gradual process of *routinization* (other terms found in the diachronic literature are lexicalization, institutionalization, automatization, fossilization, etc.) of individual instances of CPs, as particular syntactic combinations should first achieve a sufficiently high frequency of use in the community of speakers to be reanalyzed and to function as a lexical unit. On the other hand, if there exists an abstract constructional schema after which CPs are formed by directly filling the open slots of the schema with lexical material, we would expect a rather rapid emergence of many different individual CPs, and this does not need to be necessarily associated with a prior gradual increase in usage frequency. This would support a diachronic scenario of a creative application of a schema that gets extended and strengthened every time a new instance is built according to it. As will be shown in the remainder of this paper, the data reveals a more differentiated picture, suggesting that both diachronic ways have been taken on the road towards CPs in German.

3 Data and Method

The present study focuses on the pattern [P1 (Det) N P2/GEN] where the first preposition P1 is represented by *mit* 'with', and the noun is an abstract deverbal noun N_{dev}, i.e. [*mit* (DET) N_{dev} P/GEN], see e.g. (6) and (7) below. This syntagmatic pattern is recurrent in many recent CPs. Moreover, these CPs may be traced back in time (at least) to the 17th century, representing a good test case for the two diachronic scenarios described in Section 2. The study is based on a data sample from the DTA (Deutsches Textarchiv)[1] that covers nine periods and amounts to about 10% of the entire corpus. Table 1 represents the composition of our data sample.

From this data, we extracted all occurrences of deverbal nouns, based on the list that was compiled using the studies by Kamber (2008), Hartmann (2016) and Smirnova & Stöber (2022). The data from the last three periods, i.e. from the 19th century, is not only the largest proportion of our corpus, but it also contains the most occur-

---

1 https://www.deutschestextarchiv.de

Table 1: Investigated periods.

Period	Years	corpus size
1	1605–1607	544,752
2	1643–1646	286,361
3	1683–1685	367,164
4	1719–1722	1,282,914
5	1758–1760	1,154,723
6	1797–1800	2,172,629
7	1834–1836	4,366,178
8	1871–1874	2,256,498
9	1896–1899	4,411,414
	TOTAL	**16,842,633**

Table 2: Most frequent nouns with **mit** in the 19th century.

lemma	N
Hilfe 'help'	338
Ausnahme 'exception'	262
Rücksicht 'regard'	217
Bezug 'reference'	106
Ausdruck 'expression'	28
Berücksichtigung 'consideration'	26
Blick 'view'	25
Zusatz 'addition'	21
Anwendung 'application'	21
Beziehung 'respect'	21

rences of CPs and related syntagmatic patterns. To illustrate this, Table 2 gives the ten most frequent nouns in our data set in the 19th century (= periods 7 to 9).[2]

For the present analysis, structures with 20 deverbal nouns were selected (see Table 3). This selection is motivated by frequency and by the semantics of the nouns, and is based on the three most frequent nouns that occur in the syntagmatic pattern [*mit* (DET) N_{dev} P/GEN] (see Table 2): *Hilfe* 'help' (n=338), *Rücksicht* 'consideration' (n=217), and *Bezug* 'reference' (n=106).[3] In addition to these three

---

2 A complete list of the nouns as well as the data used for this study can be found at https://osf.io/yancd

3 For semantic reasons, the noun *Ausnahme* 'exception' (n=262) will not be considered here. Whereas *Hilfe* and *Bezug/Rücksicht* are in close paradigmatic relation to other semantically simi-

nouns, their contextual synonyms were also considered, i.e. other deverbal nouns that are attested in the pattern [*mit* (DET) N_{dev} P/GEN] in our data and, in these uses, are semantically close to the frequent nouns *Hilfe*, *Bezug*, and *Rücksicht*. Examples (6)–(7) illustrate what is meant by contextual synonymy in the present study.

(6)   *Mit* **Hilfe/Anrufung/Beihilfe** *eines Rechtsanwalts wurde der Prozess entschieden.*
      'With help/invocation/assistance of a lawyer, the case was decided.'

(7)   *Mit* **Bezug/Hinsicht/Rücksicht auf** *das Wetter entschieden sie sich gegen die Wanderung.*
      'With reference/regard/consideration to the weather, they decided against the hike.'

Based on the semantics of the nouns in the usage pattern [*mit* (DET) N_{dev} P/GEN], two groups of structures in the data were distinguished: the so-called *Hilfe*-group on the one hand, with the total of 486 occurrences in the data, and the *Bezug*-group on the other, with the total of 407 occurrences (resulting in a data set of 895 observations in total). Table 3 shows the absolute and the normalized frequency of these nouns in the syntagmatic pattern [*mit* (DET) N_{dev} P/GEN], the periods of attestation, and the total frequency of the nouns in the corpus.

All 12,317 observations were manually annotated with regard to the following variables:
- DETERMINER: presence/absence of a determiner before the noun; if a determiner is present, the type of determiner
- PRENOMINAL MODIFICATION: presence/absence of a modifier before the noun; if a modifier is present, the type of modifier
- NUMBER: number of the noun; singular/plural
- POSTNOMINAL ELEMENT: prepositional phrase or genitive noun phrase.

---

lar nouns and build clusters of semantically related exemplars, *Ausnahme* behaves exceptionally, as it forms a unique construction [*mit Ausnahme von*/GEN].

Table 3: Analysed nouns.

lemma	absolute frequency in [mit (DET) Ndev P/GEN]	normalized frequency (per 100.000 words)	attested in periods[4]	absolute frequency in the corpus
***Hilfe*-group**				
Hilfe 'help'	416	2.47	1–9	1,852
Anbetung 'worship'	1	0.01	9	83
Anrufung 'invocation'	4	0.02	1–9	44
Beihilfe 'assistance'	8	0.05	4–9	121
Beistand 'support'	12	0.07	2–9	313
Beiziehung 'consultation'	1	0.01	6	2
Heranziehung 'reference'	2	0.01	8–9	27
Rat 'advice'	5	0.03	1–4	1,433
Unterstützung 'support'	18	0.11	7–9	410
Zuziehung 'consultation'	19	0.11	3–9	43
***Bezug*-group**				
Bezug 'regard'	104	0.62	7–9	970
Aufsicht 'attendance'	1	0.01	7	350
Berücksichtigung 'consideration'	26	0.15	7–9	158
Beziehung 'relation'	21	0.12	7–9	2,022
Einsicht 'insight'	6	0.04	7–9	768
Ersuchen 'request'	1	0.01	4	58
Hinsicht 'regard'	3	0.02	6–7	612
Rücksicht 'consideration'	223	1.32	6–9	1,639
Schein 'appearance'	5	0.03	5–8	717
Zusatz 'addition'	17	0.1	4–9	591
TOTAL	**895**			**12,317**

---

4 For the sake of lucidity, the numbered periods as indicated above are used here. The periods indicated correspond to the occurrence of the individual lemmas.

4 Results

The presentation of the results of the corpus study will proceed in two steps. We will first present the results for the *Hilfe*-group (Section 4.1), and then for the *Bezug*-group (Section 4.2) separately, before we come to the synthesis. In each case, we will start by giving the frequency numbers, as they shed light on the presence or absence of a schema. A slow and gradual increase can be interpreted in terms of gradual conventionalization, whereas a rather abrupt rise would indicate a direct insertion of the noun into an open slot of an existing schema (see Section 2). We then proceed by reporting the results regarding the internal structure of the pattern [*mit* (DET) N_{dev} P/GEN], focusing on the parameters DETERMINER use and PRENOMINAL MODIFICATION. The interpretation of these parameters is less straightforward, as these characteristics cannot be taken as directly indicating the status of an individual combination as a CP or not (see Section 1 on the definitional problems of CPs). Nevertheless, they may help to delineate more fixed syntagmatic structures with less syntactic regularity from regularly formed syntactic structures. In the last step, the semantic aspects will be taken into account.

4.1 *Hilfe*-Group

Changes in the frequency of use of the individual instances of the syntagmatic pattern [*mit* (DET) N_{dev} P/GEN] with the nouns of the *Hilfe*-group are shown in Figure 2. It can be seen that most of the nouns in this group only rarely occur in the syntagmatic pattern [*mit* (DET) N_{dev} P/GEN], the only exception being the noun *Hilfe* 'help'. Importantly, the structure [*mit* (DET) *Hilfe* P/GEN] displays a relatively high frequency of use right from the start of the analyzed period, and the frequency gradually increases towards the end of the analyzed period, before it picks up in the last decade of the 19th century. For the other nouns, it can be seen that they are attested in the pattern already in the earliest periods at the beginning of the 17th century. However, towards the end of the 19th century, they don't show any signs of frequency increase and are still very rarely used in the relevant structure [*mit* (DET) Ndev P/GEN].

A closer look at the internal structure of the pattern [*mit* (DET) N_{dev} P/GEN], in particular with respect to the use of determiners and prenominal modifiers, reveals that not all nouns of this group behave in the same way in this syntactic configuration. Namely, *Hilfe* and *Beihilfe* are predominantly attested in structures without any determiner or modifier between the preposition *mit* and the noun, as shown in (8)–(9).

Figure 2: *Hilfe*-group (normalized frequencies per 100,000 words).

(8) *Es ist freylich keine Frage, daß man nicht **mit Hůlfe** der guten englischen, französischen und spanischen Stůcke ein so gutes Repertoir zusammen bringen sollte.* (Eckermann, 1836)

'There is of course no question that one should not bring together such a good repertoire **with help** of the good English, French and Spanish pieces.'

(9) *Hiernáchst zeigen auch die Versuche, welche man **mit Beihůlfe** derer Ver-gróßerungsgláser anstellet, daß die Kraft des Herzens noch über diese Grenzen hinauswirke.* (Haller, 1759)

'First of all, the experiments carried out **with aid** of magnifying glasses also show that the power of the heart still works beyond these limits.'

On the other hand, other nouns from the group are often found in structures where they are either preceded by a determiner, see (10), or by an adjectival mod-ifier, see (11).

(10) *besonders **mit der Unterstützung** Papst Gregors brachte er ein Collegium für sie in Douay zu Stande.* (Ranke, 1836)

'especially **with the support** of Pope Gregory, he brought about a college for them in Douay.'

(11) *Darauff ſo ſage ich,* **mit vorhergehender Anruffung** *Gŏttlicher Gnade, daß die Menſchen erſtlich aus Verſehung der Vernunfft, auch in Krafft und vermŏge des natŭrlichen und aller Vŏlcker Rechtens gemeiniglich das, ſo ihnen nutz und zu Erhaltung ihrer Nahrung dienlich ſeyn mag, zu ŭberkommen trachten* [. . .]. (Fleming, 1719)

'Whereupon I say, **with foregoing invocation** of divine grace, that men, first of all, by accident of reason, and also in virtue and by virtue of natural and all peoples' right, commonly strive to obtain that which may be of use to them and may serve for the preservation of their food.'

These two types of syntactic behavior are summarized in Figures 3a and 3b.

The horizontally striped bars represent structures without any intervening material between the preposition *mit* and the noun (= P_N); the diagonally striped bars illustrate structures with a determiner and/or a modifier between the preposition and the noun (= P_DET/Mod_N). In Figure 3a, the structures with *Hilfe* and *Beihilfe* are found, and in Figure 3b, other nouns of the group are represented.[5] It can be seen Figure 3a that *Hilfe* and *Beihilfe* predominantly lack any linguistic material between the preposition *mit* and the noun. In contrast, the other nouns of the *Hilfe*-group show a more heterogenous picture, and there are significantly more observations with an article and/or a modifier.

Two further points deserve attention in this connection. First, as can be seen from Figure 3b, the structures relatively often lack a determiner and/or a modifier when they are attested in the relevant syntagmatic structure [*mit* (DET) N_{dev} P/GEN]. However, if we take a look at the numbers in Table 4, we see that only a very small percentage of the overall uses of these nouns is accounted for by the uses in the relevant structure. In contrast to the noun *Hilfe*, which occurs in 22% in the pattern [*mit* (DET) N_{dev} P/GEN], the other nouns are hardly attracted to this particular syntagmatic structure.[6] As we will see in the next section, this is different for the nouns of the *Bezug*-group.

---

5 As the data is sparse during the first periods, and for better comparability of the two subgroups, only the more recent four periods are taken into account.

6 The figures for the noun *Beiziehung* can be ignored at this point, as it occurs only 2 times in the entire corpus. The noun *Zuziehung* is indeed unusually frequent in the relevant structure, but its affinity to the structure is best explained not by the semantic similarity to *Hilfe* but by other factors specific to this particular noun. Due to the limits of space, this will not be discussed here in greater detail. However, the frequent use of this one noun does not invalidate the claim that no clear schema can be seen that would motivate the syntactic and semantic behavior of the nouns.
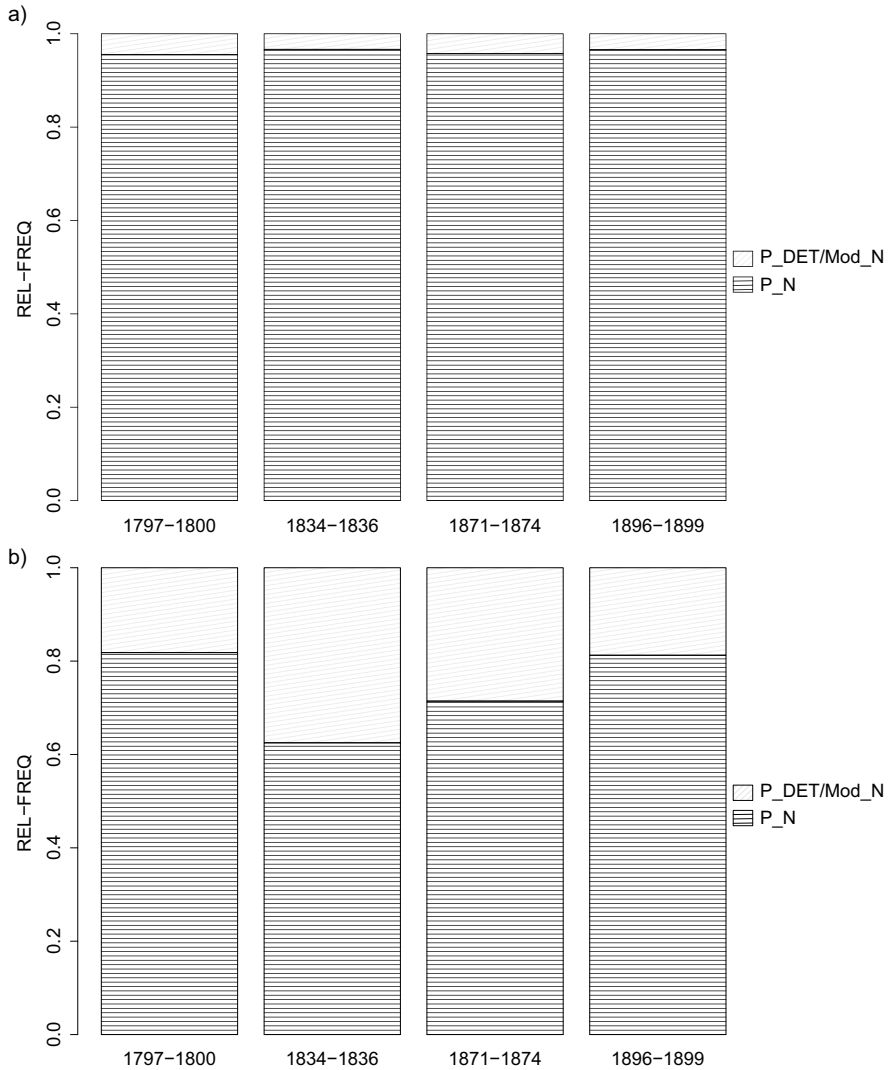
Figure 3a (top) and b (bottom): Intervening material between mit and N top: *Hilfe, Beihilfe*, bottom: *Anbetung, Anrufung, Beistand, Beiziehung, Heranziehung, Rat, Unterstützung, Zuziehung*.

Table 4: Proportion of uses in relevant structures (based on Table 3).

noun	proportion of uses in [*mit* (DET) N$_{dev}$ P/GEN]
Hilfe 'help'	**0.22**
Anbetung 'worship'	0.01
Anrufung 'invocation'	0.09
Beihilfe 'assistance'	0.06
Beistand 'support'	0.03
Beiziehung 'consultation'	0.5
Heranziehung 'reference'	0.07
Rat 'advice'	0.003
Unterstützung 'support'	0.04
Zuziehung 'consultation'	0.44

Second, from the earlier historical periods of German, we know that the structure [*mit* (DET) *Hilfe* P/GEN] had not always been restricted to the uses without any intervening material between *mit* and *Hilfe*. For example, in the data from Middle High German we often find regular syntactic structures with determiners and/or modifiers, cf. (12)–(13).

(12) *ufe deme roste cvm uns armen zu troste **mit der helfe** andir v́wer gesellen* (REM,[7] 13_1-thurhess-PV-X > M107S-N1)
'on the pyre come to comfort us poor **with the help** of other fellows of yours'

(13) *di fvln wír lefchen vnd vertiligen **mít der heiligen hilfe** der tvlt* (REM,[8] 13_1-bair-PV-G > M409-G1)
'we shall extinguish and destroy them **with the holy help** of forbearance'

This indicates that the predominance of the fixed structure without any determiner and/or modifier represented in Figure 3a above has not been there from the start, but should be seen as a result of a long diachronic process of rigidification. That is, what we see in our data is just the final stage of the development of the fixed structure [*mit Hilfe von*/GEN] that was preceded by the stage of variation between the regular syntactic uses of the noun, i.e. with articles and modifiers, and the more fixed structure without them.

---

7 REM, accessed through the ANNIS interface, https://annis.linguistics.rub.de/?id=6f31a09c-418b-42c0-a6eb-b3b1ac2bc55e
8 REM, accessed through the ANNIS interface, https://annis.linguistics.rub.de/?id=da823c93-c8be-4e2c-ba23-e4d88f6647bc

To be able to answer the question whether there is a constructional schema, we need also to look into the semantics of this structural pattern. As exemplified in (14), see also (8) above, *mit Hilfe* has instrumental semantics.

(14) *der Dieb war **mit Hilfe einer Feuerleiter** in das oberſte Giebelfenſter einge-ſtiegen* (Fontane, 1873)
'the thief had climbed into the top gable window **with help of a fire escape**'

Importantly, *Hilfe* is the only noun from the group that is often used with inanimate referents as in (14). Other nouns from the same group are however predominantly used with animate referents and retain their original semantics of helping events as exemplified in (15)–(16).

(15) *Sin-pu fiel, nachdem am 10. Juni Ward **mit Unterstützung der Engländer** seine Truppen unter grossem Verlust aus der Stadt gezogen hatte* (Martens, 1873)
'Sin-pu fell after Ward, **with assistance of the English**, had pulled his troops out of the city with great loss on June 10'

(16) *Man hat Beispiele, dass auf solche Art ein Heger **mit Beihilfe einiger Kinder** täglich 6000 bis 9000 Stämmchen ausgesetzt oder eine Area von 2 bis 3 Joch bepflanzt hat.* (Gerstner, 1834)
'There are examples that in such a way a hunter, **with help of some children**, has released 6000 to 9000 stems a day or planted an area of 2 to 3 yokes.'

This supports the interpretation that [*mit Hilfe von*/GEN] has gone through a long diachronic process during which it has become syntactically more restricted, but at the same time semantically broader, as it developed instrumental semantics in addition to the original meaning of helping.

From that, it may be concluded that nouns that are semantically related to the noun *Hilfe* and that occur in same or similar syntagmatic patterns do not form a homogeneous class and thus cannot be assumed to follow the same formation pattern. Instead, two different sub-groups within the *Hilfe*-group can be distinguished. On the one hand, the noun *Hilfe* (and to a lesser extent the relatively infrequent noun *Beihilfe*) shows a steadily increasing token frequency, reduced syntactic variability, and semantic broadening along with syntactic context expansion. The syntagmatic structure is fixed, articles and modifiers are hardly used in these structures (see Figure 3a). The nouns *Anbetung*, *Anrufung*, *Beistand*, *Beiziehung*, *Heranziehung*, *Rat*, *Unterstützung*, and *Zuziehung*, on the other hand,

are used with a stable and low token frequency in the relevant structures, they are less fixed and more variable as concerns their syntactic behavior, and they maintain their original semantics of helping, which is seen in their contextual preferences for human referents.

4.2 *Bezug*-Group

A similar situation holds for the second group of nouns, the so-called *Bezug*-group. We start by presenting the frequency numbers. Figure 4 shows the changes in the frequency of use of the individual instances of the syntagmatic pattern [*mit* N_{dev} P/GEN] with the nouns from the *Bezug*-group.

As can be seen from Figure 4, the nouns are only attested in the relevant construction starting from the end of the 18th century. The nouns *Bezug* and *Rücksicht* are more frequent than the others, and *Rücksicht* shows a very steep rise in frequency between 1800s and 1870s. Importantly, if we compare Figure 4 with Figure 2 in Section 4.1, we see that in contrast to the *Hilfe*-group, where only *Hilfe* has high frequency of use right from the start of the analyzed period, several nouns of the *Bezug*-group increase in frequency at about the same time. Also, a look at the relative proportions of usages in the relevant structures with respect to the total occurrences in the corpus represented in Table 5 reveals a picture that is very different to the *Hilfe*-group discussed in Section 4.1.

As can be seen from Table 5, there is no clear winner in this group. Instead, several nouns are attested with a relative frequency between 11% and 16% in the structure [*mit* N_{dev} P/GEN]. Notably, their first attestations in the structure occur at about the same time, between 1800 and 1830.

A closer look at the internal structure of the pattern [*mit* N_{dev} P/GEN] with the nouns from the *Bezug*-group reveals a similar bipartite distinction we described for the *Hilfe*-group, but the difference between the two sub-groups is even more pronounced here. There are nouns which clearly prefer syntactic structure without any intervening material between the preposition and the noun, cf. (17)–(18). Other nouns show more variability and more uses of article and/or modifier, cf. (19)–(20).

(17) *Man kann ſie aber **mit Bezug auf** die verſchiedene Behandlungsweiſe in eigentliche Obſtgärten und Weingärten eintheilen* (Baumstark, 1835)
      'However, one can divide them into actual orchards and vineyards **with regard to** the different methods of treatment'
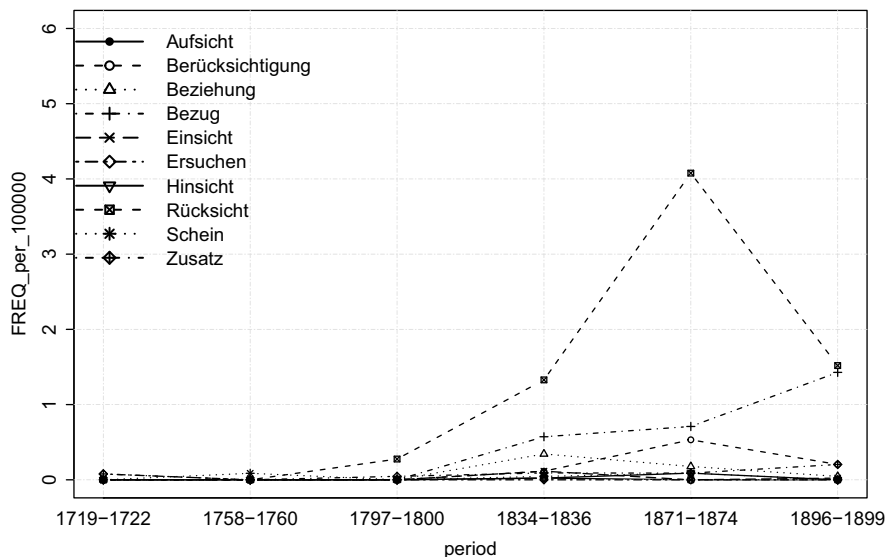
Figure 4: Bezug-group (normalized per 100,000 words).

Table 5: Proportion of uses in relevant structures (based on Table 3).

noun	proportion of uses in [*mit* (DET) N_{dev} P/GEN]
Bezug 'regard'	**0.11**
Aufsicht 'attendance'	0.002
Berücksichtigung 'consideration'	**0.16**
Beziehung 'relation'	0.01
Einsicht 'insight'	0.007
Ersuchen 'request'	0.01
Hinsicht 'regard'	0.004
Rücksicht 'consideration'	**0.14**
Schein 'appearance'	0.007
Zusatz 'addition'	0.03

(18) *Diese sämtlich hier zu besprechen, müssen wir uns* **mit Rücksicht auf** *den knapp bemessenen Raum, der uns noch zur Verfügung steht leider versagen.* (Wrangel, 1898)

'Unfortunately, we have to refrain from discussing all of them here **with regard to** the limited space available to us.'

(19) *Das lateiniſche Exemplar wies ſich offenbar als eine Ueberſetzung des italie-*
*niſchen aus; nur etwas frei, **mit dem Zuſatz** einer leichten Nuance des Ge-*
dankens. (Ranke, 1836)
'The Latin copy evidently proved to be a translation of the Italian; only
somewhat free, **with the addition** of a slight nuance of thought.'

(20) *Angeriebene Farben waren damals noch nicht zu haben; es wurde ein glä-*
*serner Reibestein mit einem Läufer angeschafft, um die rohen Pigmente **mit***
***gehörigem Zusatz** von aufgelöstem Gummi anwendbar zu machen.* (Par-
they, 1871)
'Coated colours were not yet available at that time; a glass rubbing stone
with a runner was purchased to make the raw pigments applicable **with**
proper addition of dissolved rubber.'

Figures 5a and 5b show the distribution of the two types of syntactic behavior in
our data. Similar to Figures 3a and 3b, the horizontally striped bars represent
structures without any intervening material between the preposition *mit* and the
noun (= P_N); the diagonally striped bars illustrate structures with a determiner
and/or a modifier between the preposition and the noun (= P_DET/Mod_N). Figure
5a depicts the aggregated results for five nouns *Bezug*, *Berücksichtigung*, *Bezie-*
hung, *Hinsicht*,[9] and *Rücksicht*. In Figure 5b, the nouns *Aufsicht*, *Einsicht*, *Ersu-*
chen, *Schein*, and *Zusatz* are represented.[10]

What we clearly see in Figure 5a, is a pronounced diachronic tendency towards
a relatively fixed syntagmatic pattern [*mit* N_{dev} P/GEN] without any intervening ma-
terial between the preposition *mit* and the noun, as the use of articles and/or modi-
fiers constantly declines (cf. Figure 5a). The nouns *Aufsicht*, *Einsicht*, *Ersuchen*, *Schein*,
and *Zusatz* are predominantly used in structures with determiners and/or modifiers,
and there is no clear tendency in the data (cf. Figure 5b). Though the use without an
article or modifier increases in the 1890s, it is nowhere near the same level as the
sub-group in Figure 5a. That is, five nouns *Bezug*, *Berücksichtigung*, *Hinsicht*, *Bezie-*
hung, and *Rücksicht* show an increasing token frequency and reduced syntactic vari-
ability, whereas the other five, i.e. *Aufsicht*, *Einsicht*, *Ersuchen*, *Schein*, and *Zusatz* are
associated with a stable but low token frequency and are more variable syntactically.

9 The nouns *Hinsicht* and *Beziehung* are included here into the group on the left-hand side for
reasons of their syntactic behavior. Namely, they behave more similarly to the nouns *Bezug* and
Rücksicht than to the nouns *Einsicht* or *Aufsicht*. Although the frequency information from
Table 5 suggests that *Hinsicht* and *Bezieung* do not group together with *Bezug* and *Rücksicht*, the
syntactic behavior strongly indicates that the same pattern is in question here.
10 Here again, as the data is sparse for the first periods, only the most recent four periods are
taken into account.

a)

b)
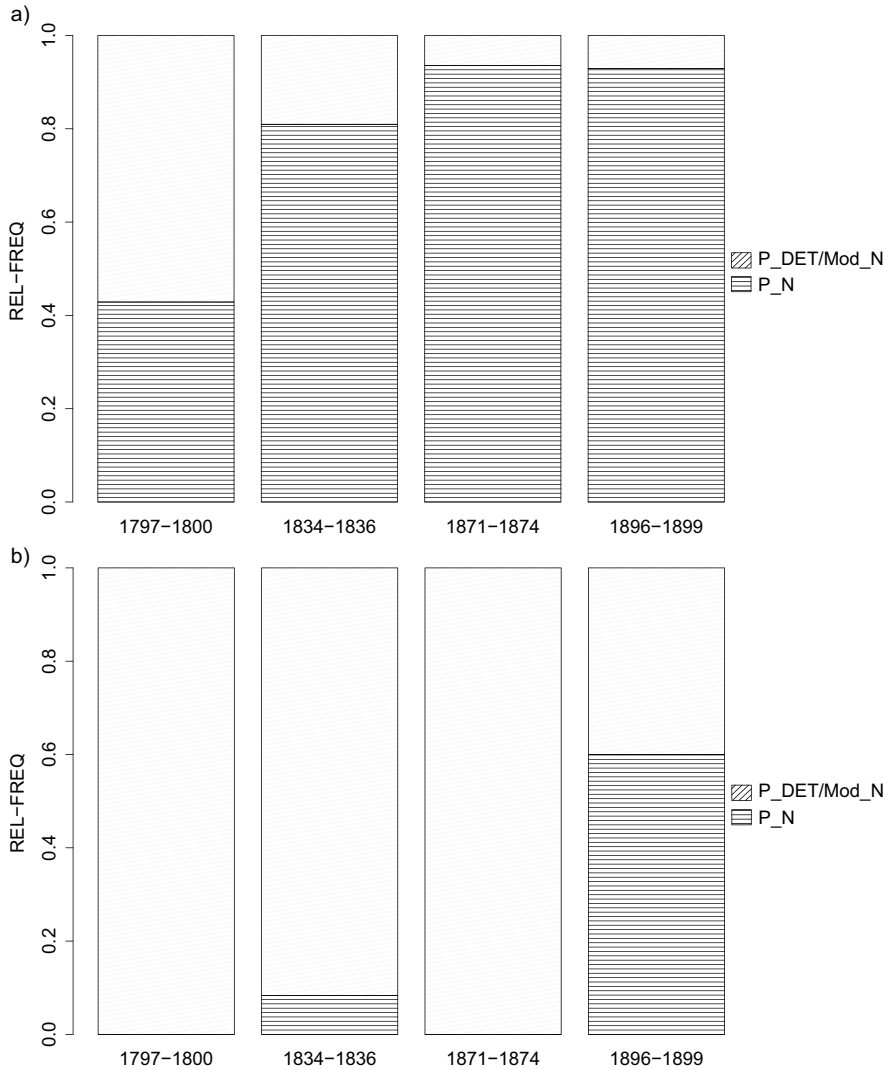
Figure 5a (top) and b (bottom): Intervening material between mit and N top: *Bezug*, *Berücksichtigung*, *Hinsicht, Rücksicht, Beziehung*; bottom: *Aufsicht, Einsicht, Ersuchen, Schein, Zusatz*.

With respect to semantic aspects, the structures with the nouns *Bezug*, *Beziehung*, *Berücksichtigung*, *Hinsicht*, and *Rücksicht* are associated with very similar contextual interpretations and may be best described as evoking the abstract semantics of reference, see (21)–(23).

(21) ***Mit Rücksicht auf** die Sylvesterfeier beginnt die Vorstellung schon um 6 Uhr Abends.* (Mährisches Tagblatt, 1896)
'**With regard to** the New Year's Eve celebration, the performance will begin already at 6 o'clock in the evening.'

(22) *sank [. . .] zu einem Niveau herab, das **mit Bezug auf** die Erkenntniß der Wahrheit nur als Barbarei bezeichnet werden kann.* (Ernst Haeckel, 1899)
'sank [. . .] to a level which, **with regard to** the recognition of truth, can only be described as barbarism.'

(23) [. . .] *und zuletzt sang Klein, **mit Beziehung auf** das warme Wetter sein schönes Frühlingslied* (Parthey, 1871)
'. . . and at the end Klein, **with reference to** the warm weather, sang his beautiful spring song'

From that, it may be concluded that nouns that are semantically related to the noun *Bezug* and that – at first sight – occur in the same or similar syntagmatic patterns, do not form a homogeneous class, similarly to the situation of the *Hilfe*-group discussed in Section 4.1. However, there are some important differences to the *Hilfe*-group, which we would like to emphasize at this point.

First, the five nouns in this group show very similar behavior, with respect to syntactic and semantic characteristics as well as with respect to frequency changes. This strongly supports the idea of the presence of an abstract schema [*mit* N_{dev} P/GEN] with the open slot [N_{dev}] that attracts semantically related nouns. The relatively rapid increase in frequency at about the same time (see Figure 4), coupled with the gradual shift of syntactic variability towards a fixed pattern [*mit* N P/GEN] represented in Figure 5a, corroborate this idea. Additional evidence for the existence of a schema comes from what we know about the time of coining of these deverbal nouns. Indeed, they were coined in different historical periods of German: *Bezug* has existed since Old High German, *Beziehung* was coined in the Early New High German period, and the nouns *Berücksichtigung*, *Hinsicht*, and *Rücksicht* are of rather recent origin between 1750 and 1780. With this chronology in mind, one could imagine that *Bezug* would have been the first to establish the relevant structure, and the others would follow its lead. And still, all of these nouns appear in the pattern [*mit* N_{dev} P/GEN] at about the same time more or less simultaneously. This

again speaks for the plausibility of a schema. The other five nouns from the group, on the other hand, display characteristics that qualify them as regular syntactic combinations, similar to the majority of the nouns from the *Hilfe*-group described in Section 4.1.

In sum, the syntagmatic patterns of the nouns in the *Hilfe*-group are very unlikely to be motivated by a constructional schema, whereas some of the nouns in the *Bezug*-group clearly follow a common pattern of formation.

5 A Constructional Network of the Pattern [*mit* N$_{dev}$ P/GEN]

From what has been said so far, we would like to argue that, in both semantically defined groups of nouns used in the general syntactic pattern [*mit* N$_{dev}$ P/GEN], a very similar general distinction can be observed. On the one hand, we find regular syntactic combinations with compositional semantics, and on the other hand there are more or less fixed syntagmatic patterns with reduced semantic compositionality. Table 6 summarizes these findings.

The crucial difference between the two groups boils down to the different analyses of the structures that do not qualify as regular syntactic combinations with compositional semantics, but rather as fixed syntactic structures with reduced compositionality. They are represented in the middle column in Table 6. These structures display distinct formal characteristics and should be treated as two distinct constructional entities, as we would like to argue in the following.

The results of our corpus study show that *mit Hilfe* can be best characterized as a fixed multi-word lexical item, i.e. as an individual lexically filled construction. It displays high token frequency, reduced compositionality and has instrumental semantics. It does not qualify as a productive schema, as it did not expand further by integrating more and more synonymous nouns into this syntagmatic pattern.

The analyzed data suggests that the pattern [*mit* N$_{dev}$ *auf*] with the nouns *Bezug*, *Berücksichtigung*, *Rücksicht*, *Hinsicht*, and *Beziehung*, on the other hand, does qualify as an abstract and productive constructional schema, as it attracts different nouns with similar semantics and displays its own stable semantic content across different instantiations. This analysis can be visualized in the form of a constructional network (see Figure 6, which is a modified version of Figure 1 from Section 2). Note that the situation modelled in Figure 6 holds for the end of the 19th century, as our data set goes only till 1899. Though intuitively, we think

Table 6: Sub-groups within the general pattern [*mit* (DET) N$_{dev}$ P/GEN].

GROUP	*Hilfe*-group	
NOUN	*Hilfe*	*Anbetung, Anrufung, Beihilfe, Beistand, Beiziehung, Heranziehung, Rat, Unterstützung, Zuziehung*
CHARACTERISTICS	increasing token frequency reduced syntactic variability establishment of fixed structure shift towards instrumental semantics	stable and low token frequency syntactic variability regular syntactic behavior compositional semantics
CONSTRUCTION TYPE	[*mit Hilfe von/*GEN]	no particular construction type regular syntactic combinations
GROUP	*Bezug*-group	
NOUN	*Bezug, Berücksichtigung, Rücksicht, Hinsicht, Beziehung*	*Aufsicht, Einsicht, Ersuchen, Schein, Zusatz*
CHARACTERISTICS	increasing token frequency reduced syntactic variability establishment of fixed structure shared semantics of reference	stable and low token frequency syntactic variability regular syntactic behavior compositional semantics
CONSTRUCTION TYPE	[*mit* N$_{dev}$ *auf*]	no particular construction type regular syntactic combinations

that the present-day situation should be very similar to those represented in Figure 6, it ought to be tested on the basis of present-day corpus data.

The schematic construction [*mit* N$_{dev}$ *auf*/GEN] is located on a middle level of abstraction and is connected via the instantiation link to the more abstract and general schema [P1 N P2/GEN] higher up in the hierarchy. We would like to argue that the most general schema at the highest level of abstraction is neither active nor productive, as only schemas on lower levels of abstraction are connected to semantic content. The general schema is surely a useful generalization for us linguists, but it remains to be investigated whether it can be used by the speakers to produce new CPs in German.

The proposed sub-schema [*mit* N$_{dev}$ *auf*] is associated with the meaning of reference, and this meaning is stably attested for five different individual constructions in our data. The construction [*mit Hilfe von/*GEN], on the other hand, does not qualify as a productive schema. Instead, it is best described as an individual lexically specified construction type with its own instrumental meaning. It is located on the same level of the hierarchy as individual instantiation types of the schema [*mit* N$_{dev}$ *auf*/GEN], but it is not directly connected to any higher-level

```
┌─────────────────────────┐
│    [P1 N P2/GEN]         │                          general schema
└─────────────────────────┘
         │         ╲
         │          ╲
┌─────────────────────────┐                          sub-schema
│   [mit N auf/GEN]        │
│   semantics of reference │
└─────────────────────────┘
         │
         ▼
┌─────────────────────────┐   ┌─────────────────────┐
│   [mit Bezug auf]        │   │  [mit Hilfe von/GEN] │
│   [mit Rücksicht auf]    │   │   instrumental       │   construction type(s)
│   [mit Hinsicht auf]     │   │   semantics          │
│   [mit Beziehung auf]    │   └─────────────────────┘
│   [mit Berücksichtigung GEN] │
└─────────────────────────┘
```
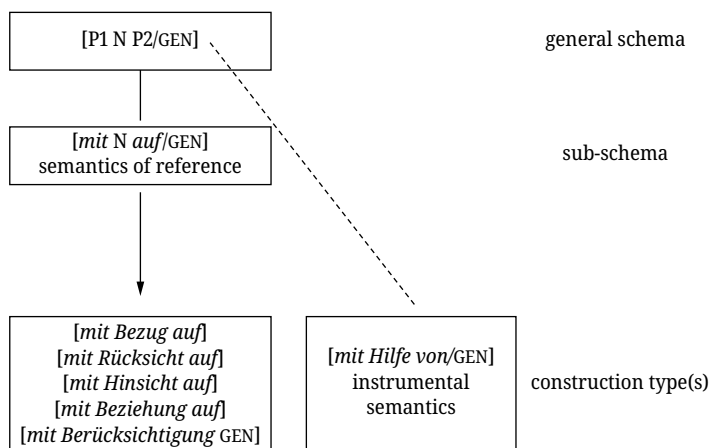
Figure 6: Partial network of German complex prepositions with *mit.*

schema, nor does it instantiate any further constructions lower down in the network. We connected this construction directly to the highest node in the partial network, to the most general schema [P1 N P2/GEN], and the dashed line serves to mark the fact that the connection is very weak, as it is only motivated structurally, but not semantically.

6 Conclusions and Outlook

This paper has offered a diachronic take on the formation of CPs in German, and we would like to conclude this section with some considerations related to the issues of routine and creativity. We started this paper by contrasting the received view of the diachronic development of CPs in German with a more recent hypothesis about their emergence. According to the traditional view, German CPs predominantly arise via reanalysis of previously regular and compositional syntactic combinations. According to the alternative hypothesis, CPs are formed directly after a general schema by filling its open slots with lexical material. We hope to have shown that both diachronic pathways have been taken on the road towards CPs in German and that the open class of CPs has been expanded using different diachronic processes and mechanisms.

The mechanisms at hand in the traditional scenario, along with reanalysis, are entrenchment and conventionalization of a syntagmatic string. Our study has demonstrated that the CP [*mit Hilfe von*/GEN] emerged in accordance with this sce-

nario. The process had started presumably very early in the history of German, as already in the Middle High German period, we find syntactic variation with respect to the internal structure, i.e. the variation in the use of articles and modifiers. In the course of the Early New High German period, the process seems to have consolidated, so that we find a relatively fixed structure towards the end of the 17th century. That is, [*mit Hilfe von*/GEN] developed through a rather slow diachronic process, whereby at some stage in this process, this string was reanalyzed into a CP.

How is the development of this particular complex preposition related to the issues of routine and creativity? We believe that the interpretation in terms of routine and creativity is not straightforward. In the subsequent paragraphs, we will argue that the interpretation is rather contingent on the perspective one adopts towards a phenomenon under scrutiny.

On the one hand, we are dealing with a new multi-word lexical item, a new CP that had not previously existed in the language. Importantly, it is the violation of the original and regular syntactic structure that defines this new item: for example, the noun loses its ability to behave as a regular noun and be accompanied by articles and/or modified by adjectives. In this respect, the development of [*mit Hilfe von*/GEN] can be seen as involving E-creativity (in the sense of Sampson 2016), as original syntactic rules are broken to produce this new item. On the other hand, if we zoom in on the process itself, and not only on the input and the output of this process, a rather different interpretation seems plausible as well. From this perspective, the gradual diachronic development of [*mit Hilfe von*/GEN] represents a good case of routinization of one particular structural variant of the original structure. It got established through repeated use by the members of the speech community and were gradually entrenched as a unit in the minds of individual speakers. Though as a result of this process, a new lexical item is added to a language, and the grammatical system is expanded by a new CP (=E-creativity), the process itself essentially includes many small steps, each of them boiling down to the repetition of the same structure by more and more speakers in more and more communicative situations. It is thus the conformity to some previously existing uses that drives the development, i.e. the routine, and less so a deviation from some rule.

As we have shown, the alternative scenario is also supported by our corpus evidence, in which individual CPs are instantaneously formed after a schematic pattern by filling its open slots. This is the case for the schema [*mit* N_{dev} *auf*/GEN] with the nouns *Bezug, Beziehung, Berücksichtigung, Hinsicht*, and *Rücksicht* in our data. As expected, we observed a rather rapid emergence of five different individual CPs following the same formation pattern and displaying very similar semantics. This diachronic scenario does not rely on gradual entrenchment and conventionalization of

a particular syntagmatic string, instead, it essentially involves the establishment of a schema. Whether this diachronic process is motivated by creativity or by routine, (at least) two different interpretations might be considered here as well, similar to the case of [*mit Hilfe von*/GEN] discussed above. On the one hand, the establishment of a schema itself heavily relies on a repeated use of individual instantiations in individual communicative situations by individual members of a linguistic community. From this point of view, the development of a schema may be seen as a process of routinization that is motivated by repeated use of particular syntagmatic strings. No creative "abuse" of some rule is required to be involved in this process, i.e. no E-creativity. On the other hand, it is due to the productive expansion of this schema that new CPs are coined in German at that time. Productivity is related to creativity, as is well known. Indeed, for Sampson (2016: 19), productivity roughly equals F-creativity, i.e. "activities which characteristically produce examples drawn from a fixed and known". In our view, the formation of any new CP after this productive pattern indeed involves the creative use of the resources available to the speakers of a speech community, and cannot be reduced to the mere reproduction of some earlier syntagmatic strings. It is thus the creative access to the schema [*mit* N_{dev} *auf*/GEN] by speakers/writers that leads to the production of several individual instances of CPs and at the same time serves to strengthen the schema itself.

To conclude, if it comes to the role of routine and creativity in different diachronic processes, it might be rather the question of perspective that we adopt towards the phenomenon, and a change in perspective might well result in very different interpretations. In this connection, we would like to refer to a conclusion presented in Bergs (2019) regarding the relationship between E- and F-creativity:

> In other words, while it seems theoretically absolutely plausible to distinguish the two types, differentiating between the two on the basis of real-life utterances is a lot more difficult. [. . .] So, from a practical point of view, what we ultimately end up with is not so much a clear dichotomy between F-creativity and E-creativity but rather a continuum between these two poles. (Bergs 2019: 181)

In a similar way, it seems to us that the dichotomy between routine and creativity can be conceived of as a matter of degree; additionally, this dichotomy is largely dependent on the vantage point that is adopted in a particular case. It might be the case that a predominantly synchronic perspective will yield totally different interpretation of the role of creativity and routine with regard to the same research object, in our case, German complex prepositions. We have focused on diachronic shifts, but even from a diachronic perspective, one could choose either to take a global view on the phenomenon, or to concentrate on small-scale changes; one could either focus on shifts on the community level (≈ conventionalization) or rather on changes in individual speakers (≈ entrenchment); one could as well

choose to abstract away from individual constructions and look at productive patterns, schemas, and rules. Every time the perspective changes, our interpretation of routine and creativity might change as well. Concerning further research, we believe that it is useful to distinguish between individual conventionalized lexical items on one hand, and schematic patterns of formation on the other. The question that deserves particular attention in future studies is, among others, whether the notions of creativity and routine may be applied in a similar fashion to them.

Of course, the class of CPs in German is much more diverse and consists of many more members following different structural patterns, as for example [*in* N *mit*] as in *in Zusammenhang mit* 'in connection with' or [*im* N *von*/GEN] as in *im Zuge* 'in the course of'. Whether the two scenarios discussed above apply to the diachronic development of other CPs, remains an open question for further research.

By means of conclusion, we would like to briefly touch upon some further perspectives for research on CPs and on complex multi-word expressions in general. If we look at the evolution of these different patterns, we see that in the case of [*mit Hilfe von*/GEN], we have a conventionalized lexical item whose internal structure became more and more fixed, and the meaning became less compositional and shifted towards the instrumental semantics. Some other synonymous nouns which, given their semantic content, had the potential to follow the analogical model of [*mit Hilfe von*/GEN], such as *Beistand* or *Unterstützung*, did not succeed in this way. That is, the establishment of [*mit Hilfe von*/GEN] did not trigger further analogical extensions. The schematic pattern of formation [*mit* N$_{dev}$ *auf*/GEN], on the other hand, appears to have evolved spontaneously as a generalization over several individual structures. These structures, leading to the establishment of a more general schema through generalization, did lead to further coinings of CPs. From the more recent history of German, especially from the 20th century, we know that it resulted in further creations like e.g. *mit Blick auf* 'in view of'. Furthermore, a related pattern with the first preposition *in* [*in* N$_{dev}$ P/GEN] existed simultaneously and produced many instances of CPs. This raises the question as to which of the developments is more characteristic in general and especially for the category of CPs. In other words, are the pathways of change somehow determined or motivated by the target category the structures develop into?

References

Beneš, Eduard. 1974. Präpositionswertige Präpositionalfügungen. In: Engel, Ulrich, Paul Grebe &
    Hugo Moser (eds.), *Sprachsystem und Sprachgebrauch: Festschrift f. Hugo Moser*, 33–52.
    Düsseldorf: Pädagogischer Verlag Schwann. (= Sprache der Gegenwart Bd.33).

Bergs, Alexander.2019. What, If Anything, Is Linguistic Creativity? *Gestalt Theory* 41 (2). 173–183.
    https://doi.org/10.2478/gth-2019-0017.

Di Meola, Claudio. 2000. *Die Grammatikalisierung deutscher Präpositionen*. Tübingen: Stauffenburg.
    (= Studien zur deutschen Grammatik Bd. 62).

Hartmann, Stefan. 2016. *Wortbildungswandel: Eine diachrone Studie zu deutschen
    Nominalisierungsmustern*. Berlin & Boston: De Gruyter. (= Studia linguistica Germanica
    Band 125).

Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions: A corpus-based study*.
    London & New York: Routledge. (= Routledge advances in corpus linguistics 7).

Huddleston, Rodney. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University
    Press. https://doi.org/10.1017/CBO9781139165785.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*.
    Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316423530.

Kamber, Alain. 2008. *Funktionsverbgefüge – empirisch: eine korpusbasierte Untersuchung zu den
    nominalen Prädikaten des Deutschen*. Tübingen: Niemeyer. (= Reihe germanistische
    Linguistik 281).

Lehmann, Christian. 1998. German abstract prepositional phrases. In: Iørn Korzen & Michael
    Herslund (eds.): *Clause combining and text structure*, 87–106. Frederiksberg: Samfundslitteratur.
    (= Copenhagen Studies in Language 22).

Lehmann, Christian & Christel Stolz. 1992. *Bildung von Adpositionen im Deutschen*. Erfurt: Seminar für
    Sprachwissenschaft der Universität. (= Arbeitspapiere des Seminars für Sprachwissenschaft der
    Universität Erfurt 6).

Lindqvist, Christer. 1994. *Zur Entstehung von Präpositionen im Deutschen und Schwedischen*. Tübingen:
    Max Niemeyer Verlag. (= Linguistische Arbeiten 311).

Meibauer, Jörg. 1995. Komplexe Präpositionen – Grammatikalisierung, Metapher, Implikatur und
    division of pragmatic labour. In: Frank Liedtke (ed.), *Implikaturen: grammatische und
    pragmatische Analysen*, 47–74. Tübingen: M. Niemeyer (= Linguistische Arbeiten 343).

Quirk, Randolph (ed.). 1985. *A Comprehensive grammar of the English language*. London & New York:
    Longman.

Quirk, Randolph & Joan Mulholland. 1964. Complex Prepositions and Related Sequences. *English
    Studies* 45(1–6). 64–73. https://doi.org/10.1080/00138386408597188.

Sampson, Geoffrey. 2016. Two ideas of creativity. In: Martin Hinton (ed.), *Evidence, Experiment and
    Argument in Linguistics and the Philosophy of Language*, 15–26. Frankfurt u.a.: Peter Lang. https://
    doi.org/10.3726/978-3-653-05840-6.

Smirnova, Elena & Lotte Sommerer. 2020. Introduction: The Nature of the Node and the Network –
    Open Questions in Diachronic Construction Grammar. In: Lotte Sommerer & Elena Smirnova
    (eds.), *Nodes and Networks in Diachronic Construction Grammar. Book edition*, 1–42. Amsterdam &
    Philadelphia: John Benjamins Publishing Company. (= Constructional approaches to language
    volume 27).

Smirnova, Elena & Vanessa Stöber. 2022. Verbo-Nominal Constructions with kommen 'come' in
    German. *Constructions and Frames* 14 (1). 121–149. https://doi.org/10.1075/cf.00060.smi.

Stefanowitsch, Anatol/Smirnova, Elena/Hüning, Matthias (2020): Complex Adpositions in Three West Germanic Languages: German, Dutch, and English. In: Fagard, Benjamin/Lima, José Pinto de/Stosic, Dejan/Smirnova, Elena (Hrg.): *Complex Adpositions in European Languages. A micro-typological approach to complex*. Boston: De Gruyter. S. 65–138. (= Empirical approaches to language typology 65).

Szczepaniak, Renata. 2009. *Grammatikalisierung im Deutschen: eine Einführung*. Tübingen: G. Narr. (= Narr Studienbücher).

Sören Stumpf

On the Dynamics of Constructional Idioms

A Micro-Diachronic Approach to the Entrenchment and
Conventionalization of Lexico-Grammatical Patterns in German

Abstract: Various studies have shown that lexico-grammatical patterns can emerge
through creative modification of fully lexicalized multi-word expressions, espe-
cially through lexical substitution (e.g., X *oder nicht* X/Y*, das ist hier die Frage* [Engl.
'To X, or not to X/Y, that is the question']). Following the latest research, we ask
how the development from lexically fixed idioms via modification to semi-sche-
matic idioms (so-called constructional idioms), a process which happens over a
shorter period, can be empirically determined, and theoretically explained. From a
theoretical point of view, phraseological and constructionist concepts of linguistic
creativity, approaches of Diachronic Construction Grammar to constructionaliza-
tion, and Schmid's (2020) Entrenchment-and-Conventionalization Model are consid-
ered. Methodologically, micro-diachronic corpus analyses (real-time data) and an
online survey (apparent-time data) of German constructional idioms are combined.
The method is illustrated by two case studies on the multi-word expressions [X *du
noch oder* Y *du schon?*] [Engl. 'Are you still X or do you already Y'] and [*Nach* X *ist
vor* X/Y] [Engl. 'After X is before X/Y']. The case studies show, on the one hand, how
the form and meaning of lexicalized multi-word expressions change over a few
decades. On the other hand, they reveal that constructional idioms are stored very
differently in the minds of individual speakers. Overall, the paper aims to provide
an adequate account of the interplay between routine and creativity, variation and
change, and entrenchment and conventionalization regarding the dynamics and
emergence of constructional idioms in German.

1 Introduction

This paper deals with the dynamics and emergence of lexico-grammatical pat-
terns in German. The focus is on the role of creativity "as a real engine of change"
(Mellado Blanco 2024: 508) and the interaction of creativity and routine in the for-
mation of new patterns. Creative utterances can be defined in the broadest sense
as "novel utterance tokens that do not instantiate a conventional utterance type,
but change such a type" (Schmid 2020: 19). We assume that semi-schematic con-
structions can develop because of frequent creative substitutions of lexical units

within a fully lexicalized multi-word expression. This assumption can be illustrated with examples (1)–(3) taken from the German Reference Corpus.[1]

(1) **Gesichertes Mittelfeld oder Abstiegskampf? Das ist hier die Frage.** Mannheims Trainer Rainer Ulrich blickt zunächst jedoch ausschließlich auf den morgigen Samstag, wenn der VfR vor eigenem Publikum die SpVgg Ludwigsburg empfängt.
(Mannheimer Morgen, March 18, 2005)
[Engl. 'Safe midfield or relegation battle? That is the question here'][2]

(2) **Laden oder nicht laden, das ist hier die Frage.** Es geht um Alkali-Batterien, von denen die Hersteller behaupten, sie seien nicht wieder aufladbar.
(Süddeutsche Zeitung, June 4, 1996)
[Engl. 'To charge or not to charge, that is the question here']

(3) **Sauber oder dreckig? Das ist hier die Frage!** Ungetrübte Badefreuden an den Stränden Niedersachsens verspricht ein EU-Bericht – die Wasserqualität sei 2002 sehr gut geblieben.
(Rhein-Zeitung, February 8, 2005)
[Engl. 'Clean or dirty? That is the question here!']

From a phraseological point of view, the highlighted expressions could be classified as occasional variations of the idiom *Sein oder nicht sein, das ist hier die Frage* [Engl. 'To be, or not to be, that is the question']. The reason for this interpretation is that most people are probably familiar with this catchphrase (Parkinson 2003), which goes back to Shakespeare's "Hamlet". If the origin of a formulaic expression is (still) known in the speech community, such idioms can also be called "geflügelte Worte" (Engl. 'winged words'), according to the classification of German phraseology (Burger 2015: 48–49). In this case, the components *Sein* and *(nicht) sein* are substituted creatively. Such creative variations of familiar and fixed multi-word expressions are quite a common phenomenon called "modification", as it has been described many times in phraseology (Barz 1992; Dobrovol'skij 1999; Ptashnyk 2009; Jaki 2014; cf. also Section 3.1).

---

1 The German Reference Corpus is accessible via the platform COSMAS II: cosmas2.ids-mannheim.de/cosmas2-web/ (March 1, 2024).
2 In examples, the (bold) highlighted multi-word expressions are translated in the following.

However, if we search for this idiom in the German Reference Corpus, we can see that it occurs only 56 times (about 4%) in its original form out of a total of 1,354 hits (100%) (Stumpf 2016: 318). In 96% of all instances, it is "modified", i.e., other words or phrases are used instead of *Sein* and *(nicht) sein*, such as verbs (*laden oder nicht laden*), adjectives (*sauber oder dreckig*), nouns (*Mittelfeld oder Abstiegskampf*) and whole phrases or sentences (*die ganze Nacht für die Klausur lernen oder mit Freunden feiern gehen* [Engl. 'study all night for the exam or go out partying with friends']). Furthermore, it can be observed that in only very few instances the first element is repeated in the second position and combined with the negation *nicht* [Engl. 'not']. The use of different elements is more frequent. Thus, a simple corpus study reveals that the idiom is used much more frequently in a modified form. It is hard to say that these forms are creative variations (modifications) of a fully lexicalized idiom. Instead, the high number of variations suggests that these are fillers of an underlying semi-schematic pattern (X *oder nicht* X/Y, *das ist hier die Frage* [Engl. 'To X, or not to X/Y, that is the question']) which can be described as a "constructional idiom" (Taylor 2002: Chapter 28.2). The example shows that lexicalized idioms can develop into more flexible idioms through modification.

This paper focuses on such dynamic processes at the interface of creativity and routine, and at the interface of phraseology and Construction Grammar. The following questions are addressed:

–   What methods can be used to explore the dynamics and emergence of constructional idioms?
–   What formal and semantic changes take place during the development of (new) constructional idioms?
–   What theoretical approaches can be used to describe such dynamic processes?

Section 2 gives an overview of the research on constructions between lexicon and grammar and defines the term "constructional idiom". In Section 3, we introduce theoretical approaches to the analysis of the dynamics of constructional idioms. We first discuss the state of research on creativity in phraseology and Construction Grammar (Section 3.1); in addition, we address approaches to Diachronic Construction Grammar, particularly to constructionalization (Section 3.2), and we outline the Entrenchment-and-Conventionalization Model by Schmid (2020) (Section 3.3), which we apply from both an empirical and theoretical perspective. Section 4 presents our methodology (Section 4.1) and illustrates the approach with case studies of the constructions [X *du noch oder* Y *du schon?*] [Engl. 'Are you still X or do you already Y'] (Section 4.2) and [*Nach* X *ist vor* X/Y] [Engl. 'After X is before X/Y'] (Section 4.3). A conclusion and a discussion of the interplay between creativity and routinization in the dynamics of constructional idioms are given in Section 5.

2 Constructional Idioms

Usage-based approaches of Construction Grammar assume that linguistic knowledge can be described as a network of conventionalized pairings of form and meaning (function) called constructions (Goldberg 1995, 2006, 2019; Hoffmann and Trousdale 2013). One of the most important assumptions of constructionist approaches is that even completely syntactic and schematic structures (such as passive constructions) have meaning. The extension of form-meaning pairings to lexically unspecified abstract patterns results in the rejection of a strict distinction between lexicon and grammar. Instead, a continuum between lexical units and syntactic structures is considered. At one end of the lexicon-grammar continuum we find atomic and concrete pairings of form and meaning (e.g., simple words), at the other end schematic and complex constructions (e.g., ditransitive constructions). Between lexicon and grammar semi-schematic patterns are situated (e.g., [X *of you*], Goldberg and Herbst 2021), which in constructionist research are known as "constructional idioms" (Taylor 2002), "formal or lexically open idioms" (Fillmore, Kay and O'Connor 1988) or "schematic idioms" (Croft and Cruse 2004). They can be defined as "partially lexically-filled phrasal patterns" (Goldberg 2006: 215) of varying degrees of productivity and schematicity with a (partially or fully) non-compositional (pragmatic) meaning (Booij 2002: 320; Dobrovol'skij 2011a: 114; Ivorra Ordines 2022: 33–35). In phraseology, such patterns are called "open-slot idioms" (Martí Solano 2013), "Phraseoschablonen" (Fleischer 1997) and "Modellbildungen" (Burger 2015). Dobrovol'skij (2011a) brings together phraseological and constructionist approaches and establishes the term "Phrasem-Konstruktionen" ("constructional phrasemes").

Constructional idioms have received greater attention in recent years (e.g., Corpas Pastor 2021, 2022; Mellado Blanco 2022; Mellado Blanco, Mollica and Schafroth 2022). In previous research, they have been analyzed from a product-oriented rather than a process-oriented perspective: both phraseological and constructionist studies have primarily described the lexical (e.g., meaning of the whole construction, formal and semantic properties of the fillers) and grammatical (e.g., syntax of the construction, morphosyntactic restrictions on slot filling) properties of constructional idioms based on synchronic data (for German cf. Staffeldt 2018; Mollica 2020; Stumpf 2021). So far, only a few studies deal with the dynamics of constructional idioms in general and the development of idioms into constructional idioms through creativity in particular as shown in Section 1 (for an overview cf. Mellado Blanco 2022: 9–13). For instance, Mellado Blanco (2018) analyzes the pattern [*Reden ist Silber,* X *ist Gold*] [Engl. 'To talk is silver, to X is gold'] and Stutz and Finkbeiner (2022) explore the pattern [X *kam, sah und* Y]

[Engl. 'X came, saw and Y'].[3] Thus, the emergence of semi-schematic constructions from fully lexicalized idioms is an interesting phenomenon between lexicon and grammar that needs to be studied more intensively.

Also concerned with partly lexicalized patterns is research on linguistic creativity (Bergs 2018: 281–283, 2019: 176–177; Section 3.1) and constructionalization (Traugott and Trousdale 2013: 183–186, 2014: 270–272; Section 3.2) that refers to such patterns as "snowclones". The term goes back to an entry in the linguistics blog Language Log from 2004 (for the history of the term, see in detail Hartmann and Ungerer 2024: 600–602).[4] However, it is used quite vaguely in the literature. Traugott and Trousdale (2013: 150), for instance, consider snowclones as patterns that develop from "fixed micro-constructions that are usually formulae or clichés". What is meant by "formulae and clichés" in this definition is unclear. In a recent study, Hartmann and Ungerer (2024) provide the first theoretical systematization of snowclones and a corpus analysis of the English constructions [*the mother of all* X] and [X *is the new* Y]. They claim that

> snowclones are characterised by the extension of the source construction to new instances via partial lexical substitution. Snowclones can thus be regarded as semi-schematic constructions composed of both fixed elements (e.g. *the mother of all*) and open slots (represented by variables such as X and Y). (Hartmann and Ungerer 2024: 603).

We see that the definition of snowclones is very similar to the definition of constructional idioms, with the difference that snowclones are based on a fixed source construction (a "winged word" in terms of phraseology). For this reason, Hartmann and Ungerer (2024: 626) categorize snowclones as a subtype of constructional idioms, but still "as a class of their own". To some extent, they also consider the dynamics of the analyzed snowclones. However, this is not the focus of their study. This paper therefore presents methodological, empirical, and theoretical perspectives on the dynamics and development of German constructional idioms.

---

3 For analyses of Spanish constructions developing from idioms to semi-schematic constructions, see Mellado Blanco (2020, 2023) and Ivorra Ordines (2022, in press).
4 itre.cis.upenn.edu/~myl/languagelog/archives/000350.html (March 1, 2024).

3 Theoretical Approaches to the Analysis of the Dynamics of Constructional Idioms

3.1 Creativity in Phraseology and Construction Grammar

In both Construction Grammar and phraseology, the concept of creativity plays an important role. In recent years, both fields of research have increasingly dealt with the questions of how creativity can be defined, to what extent creativity and routinization interact with each other, and how linguistic creativity can be explained from a cognitive-linguistic perspective (Langlotz 2006; Dobrovol'skij 2008; Zeschel 2012; Hoffmann 2018b, 2020; Goldberg 2019: Chapter 3; Ungerer and Hartmann 2023: Chapter 5.1).

In phraseology, a distinction is usually made between variation and modification. Variation is defined "as a regular formal change of a pattern licensed by the norms of a given language" (Filatkina 2018a: 26). As a result, phrasemes can have two or more conventionalized forms (e.g., *mit den Achseln zucken / die Achseln zucken* [Engl. 'shrug one's shoulders']). Thus, numerous studies have shown that phrasemes with completely fixed structure are the exception (Fellbaum and Stathi 2006; Fellbaum 2019). Different types of variation can be differentiated. For instance, there is morphological (*seine Hand/Hände im Spiel haben* [Engl. [literally] 'have a hand/hands in the game']) and lexical variation (*bis an/über den Hals in Schulden stecken* [Engl. [literally] 'to be in debt up to/above the neck']), shorter or longer variants (*sich etwas [rot] im Kalender anstreichen* [Engl. 'mark something [red] in your calendar']), or variation in the argument structure of the phraseme (*jmdm./für jmdn. eine Extrawurst braten* [Engl. [literally] 'to fry sb./for sb. an extra sausage']).

While variation refers to the usual (possibly codified in dictionaries) forms of a phraseme, modification, in contrast, is an occasional version of the canonical structure of a phraseme "that has been created ad hoc by a particular user, generally in order to attain specific stylistic effects" (Rodríguez Martín 2014: 4). Modified phrasemes usually have differences in meaning compared to the original phrasemes, while variations show at most minimal semantic differences. In (4) the idiom *jmdm. einen Bären aufbinden* (Engl. [literally] 'to tie a bear on sb.'s back') is modified by an expansion with an adjective attribute (*sozialistischen* [Engl. 'socialist']); in addition, the negation marker *keinen* is used instead of the indefinite article *einen*:

(4)   Dennoch liess Mill sich **keinen sozialistischen Bären aufbinden**.
       (Neue Zürcher Zeitung, July 26, 2018)
       [Engl. [literally] 'Nevertheless, Mill did not let anyone tie him to a socialist
       bear.']⁵

Modified phrasemes are intended by speakers, and they are "always bound to a specific contextual environment" (Gläser 2001: 130). They can be used "for unexpected semantic-pragmatic effects on the part of the hearer" (Filatkina 2018a: 27). The speakers' decision to play with a conventionalized idiom distinguishes modifications from so-called mistakes/mispronunciations (Ptashnyk 2009: 55).⁶ Modification also exhibits certain types (such as substitution, expansion, reduction, and permutation) that do not differ significantly from those of variation (Fiedler 2007: 90–95; Ptashnyk 2009: Chapter 3.2; Dobrovol'skij 2011b). In semantic or contextual modifications, the meaning of the phraseme is "played with", for instance, by realizing its idiomatic as well as literal meaning simultaneously without changing its form (Burger 2015: 164–165), as in (5).

(5)   Meinem Kumpel Benjamin **ist jetzt ein Licht aufgegangen**. Nachdem er mit
       seiner Familie ins Eigenheim gezogen ist, hat er es nach dem anfänglichen
       Stress nun endlich geschafft, in den meisten Zimmern von Nullachtfünfzehn-
       Birnenfassungen auf wesentlich hübschere Deckenleuchten umzurüsten.
       (Rhein-Zeitung, November 7, 2013)
       [Engl. [literally] 'My buddy Benjamin has now seen the light. After moving
       into his own home with his family, he has finally managed, after the initial
       stress, to convert most of the rooms from zero-eight-fifteen bulb sockets to
       much nicer ceiling lights.']⁷

Modification can follow certain patterns, for instance, by preferentially substituting a particular component of an idiom. In this way, patterns of modification emerge that can develop into more schematic constructions when the modified instances far exceed the original phraseme (cf. Section 1). In an earlier study, I suggested a 50%-benchmark for such phenomena (Stumpf 2016): if in corpora the

---

5 The German idiom *jmdm. einen Bären aufbinden* means 'to tell someone something obviously untrue in the hope that he or she will believe it'.
6 The boundary between modifications and variations (and also mistakes/mispronunciations) is not easy to draw. In recent years, however, corpus-analytic approaches have shown how a demarcation between non-intentional errors, occasional modifications, and usual variations can be operationalized (Pfeiffer 2016, 2017).
7 The German idiom *jmdm. geht ein Licht auf* means 'suddenly understand / grasp something'.

original phraseme occurs in less than 50% of all instances, we are dealing with a constructional idiom that has emerged from the original phraseme. The original phraseme in this case can co-exist with the constructional idiom as a fully lexicalized micro-construction (Traugott 2008: 236).

Recent work in Construction Grammar is mostly oriented towards Sampson's (2016) distinction between F-creativity ("fixed creativity") and E-creativity ("enlarging" or "extending creativity") (Hoffmann 2018a, 2019, 2022; Bergs 2018, 2019). F-creativity is the creation of new linguistic units based on existing patterns, which could also be described as productivity (Barðdal 2008). The creation of new constructions by breaking linguistic rules is called E-creativity: "Speakers also have the ability to go beyond their existing constructional possibilities" (Hoffmann 2022: 266). Strictly speaking, it is thus only E-creativity "that enlarges or expands our system(s)" (Bergs 2019: 175). From a constructionist perspective, E-creativity, and the interaction between E-creativity and F-creativity seems to be of particular interest because "F-creativity to a great degree derive from the creative slot filling of schematic constructions" (Hoffmann 2018a: 266).

For instance, Hoffmann (2019: 2) asks how speakers use their cognitive grammar to produce utterances that break existing rules. He refers to the Blending Theory (Fauconnier and Turner 2002) to explain combinations of constructions (Hoffmann 2019, 2022; cf. Herbst 2018, who also argues for this theory). Using the example *Messi is the Mozart of football*, he shows how two input spaces (FOOTBALL and CLASSICAL MUSIC) are blended "into a joint, new space that contains the creative, new meaning" (Hoffmann 2019: 5). Hoffmann (2022: 266) also illustrates E-creativity with a modified idiom (*they kick the proverbial bucket*). Phraseological modification can thus be interpreted as E-creativity. Bergs (2018, 2019) discusses the phenomena of snowcloning (cf. Section 3.1), mismatch/coercion (e.g., *She tried to eat her way out of her clothes*, Bergs 2019: 283), and aberration (e.g., *A fun thing to talk about*, Bergs 2019: 286) and asks whether these phenomena are E-creativity or F-creativity. He points out that from an empirical point of view "differentiating between the two on the basis of real-life utterances is a lot more difficult" and that there "is not so much a clear dichotomy between F-creativity and E-creativity but rather a continuum between these two poles" (Bergs 2018: 181). Overall, however, Construction Grammar is still at the beginning of investigating verbal creativity.[8]

---

8 It can be added that the distinction between F-creativity and E-creativity has similarities with the distinction between "system" and "norm" according to Coseriu (1975, 2007) (for an overview cf. Kabatek 2023: Chapter 3). While "system" refers to the regularities inherent in language, "norm" refers to the social conventions underlying language use. According to Coseriu (2007: 267), the language system (virtually) comprises everything that is realizable in a language based

3.2 Constructional Change and Constructionalization

Since the development of an idiom into a constructional idiom involves language change, we consider diachronic constructionist approaches to its study. Thus, in recent years, numerous articles, monographs, edited volumes, and special issues have been published that deal with language variation and change and belong to the paradigm of Diachronic Construction Grammar (Noël 2007; Hilpert 2008, 2013; Fried 2009, 2013; Bergs and Diewald 2008; Barðdal et al. 2015; Merten 2018; Filatkina 2018b; Noël and Colleman 2021; Hilpert, Cappelle and Depraetere 2021; Huber and Herbst 2022; Sommerer and Smirnova 2022; Lasch and Ziem 2023). The central questions are "how constructions can change on the form or meaning side, or both" (Bergs 2017: 373) and how new constructions emerge. In addition, there is a discussion about how constructionist approaches to language change differ from or can be combined with other usage-based theories such as grammaticalization and lexicalization (Traugott 2003, 2007, 2015; Trousdale 2008, 2010, 2012; Hilpert 2011; Traugott and Trousdale 2013: 30–38).

Within Diachronic Construction Grammar, a distinction is usually made between constructionalization and constructional change. In the case of constructional change, only the form or the meaning of a construction changes (Traugott and Trousdale 2013: 22–26). Since only one side of a construction changes, constructional change does not necessarily "result in conventionalized units in which both morphosyntactic form and semantics are new" (Traugott and Trousdale 2014: 273) (e.g., *going to* → *gonna*, Bergs 2017: 373–374). In the case of constructionalization, in contrast, both the form and the meaning of a construction change, resulting in a new construction (Traugott and Trousdale 2013: 26–27). Examples are the development of the *BE going to* future (Traugott 2015: 65–73) and the development of *all*- and *what*-pseudo-clefts in English (Traugott and Trousdale 2013: Chapter 3.5). It should be mentioned, however, that the line between constructionalization and constructional change cannot be drawn clearly. Both language change processes often are interrelated with each other (Traugott and Trousdale 2013: 27).

Traugott and Trousdale (2013: Chapters 3 and 4) differentiate between grammatical and lexical constructionalization. While the output of the lexical construc-

---

on general rules (e.g., word formation rules/patterns). The norm, on the other hand, is that which constrains speakers and limits their freedom of formulation and the possibilities given by the system to utterances that are conventionalized in the speech community (Coseriu 1975: 88). As a result, there are, first, expressions which conform to the system as well as to the norm (*essbar* [Engl. 'eatable']), second, expressions which conform to the system but do not belong to the norm (*hassbar* [Engl. 'hateable']), and third, expressions which deviate from the system (and therefore also from the norm) of a language (*türbar* [Engl. 'doorable']).

tionalization is contentful, the output of grammatical constructionalization is proce-
dural and at the grammatical end of the lexicon-grammar continuum (Traugott
and Trousdale 2013: 193, 2014: 275–276; Traugott 2019: 127). For the present study,
lexical constructionalization is relevant since it focuses among other things on the
development of productive formulaic patterns out of phrases and clauses (Traugott
and Trousdale 2013: Chapter 4.7). Thus, the change concerning the "Hamlet" quote
(cf. Section 1) is a typical case of lexical constructionalization. It can be described as
the emergence of a semi-schematic pattern through a series of constructional
changes based on a fully lexicalized idiom. Traugott and Trousdale (2013: 183) refer
to such changes as snowcloning (cf. Section 3.1), as "a fixed specific expression be-
comes less fixed by virtue of introducing a variable (a formal change), while the
original meaning of the micro-construction generalizes". The approach of lexical
constructionalization allows us to explain "the role of specific constructions in the
formation of schemas" and "a pattern-based view on changes undergone by con-
tentful constructions" (Traugott and Trousdale 2014: 276–277).

A crucial factor for the development of (semi-schematic) constructions is
analogy (Fischer 2007: Chapter 3.5; Hunston and Francis 2000: Chapter 4.2.2; De
Smet 2013: Chapter 2.2).[9] The emergence of a constructional idiom such as [X *oder
nicht* X/Y, *das ist hier die Frage*] (cf. Section 1) can be seen as the result of the in-
terplay of analogization and modification: certain components of an idiom are
substituted by formal and semantically similar components in a creative way to
achieve certain effects (e.g., to be funny or to arouse interest). In her study of
Spanish constructions, Mellado Blanco (2023: 123) comes to the same conclusion:

> In lexically filled constructions (idioms) with high token frequency and a high degree of se-
> mantic coherence and entrenchment [. . .], the analogical substitution of one or several of
> its constituents by semantically related ones is a frequent process [. . .], which leads to the
> emergence of a schematic construction [. . .].

Furthermore, according to Traugott and Trousdale (2013: Chapter 5), context plays
an important role in constructionalization since "change occurs only in context"
(Traugott 2019: 129). Under "context" they include such factors as linguistic co-text
(i.e., linguistic environment, including syntax, morphology, phonology, semantics,
pragmatic inference, mode) and wider discourse and sociolinguistic contexts
(Traugott and Trousdale 2013: 196). Change usually begins when constructions are
used in unusual contexts that lead to "slight, 'untypical' shifts in the use of exist-
ing constructions" (Traugott 2019: 130). Through repetitive use in other contexts,

---

9 Traugott and Trousdale (2013: 37–38) use the term "analogization".

new pairings of form and meaning can emerge (Diewald and Smirnova 2010: 114). Both entrenchment and conventionalization are involved here.

3.3 The Entrenchment-and-Conventionalization Model

The Entrenchment-and-Conventionalization Model (in the following "EC-model") developed by Schmid (2014, 2015, 2016, 2020) "constitutes a universal and unified theory of how language(s) work(s)" (Schmid 2015: 3). The model combines cognitive with sociopragmatic considerations to explain language use, language variation, and language change. It thus considers both the differences between individual speakers and the patterns of usage within a speech community. Of central importance for the model are therefore the two interrelated processes of entrenchment and conventionalization, which Schmid (2020: 2; emphasis in original) defines as follows:[10]

> *Conventionalization* is the continual process of establishing and readapting regularities of communicative behaviour among the members of a speech community, which is achieved by repeated usage activities in usage events and subject to the exigencies of the entrenchment processes taking place in the minds of speakers.
>
> *Entrenchment* is the continual reorganization of linguistic knowledge in the minds of speakers, which is driven by repeated usage activities in usage events and subject to the exigencies of the conventionalization processes taking place in speech communities.

The EC-model (cf. Figure 1) comprises four central components, which include further concepts (Schmid 2015: 6–9, 2016: 548–549):

(a) Usage and the repeated (motor, sensory, cognitive, and social) activities it consists of
(b) The cognitive processes of association, routinization, and schematization are referred to as entrenchment
(c) The sociopragmatic processes of innovation, co-adaption, diffusion and normation as stages of increasing conventionalization
(d) A set of cognitive, emotive-affective, pragmatic, and social forces influencing usage, entrenchment, and conventionalization and the interaction between them

---

10 See also Langacker (2008: 32): "For ease of discussion, I am conflating two parameters that eventually have to be distinguished: entrenchment or unit status (pertaining to a particular speaker) and conventionality (pertaining to a speech community)". It should also be emphasized that although the term "entrenchment" is rarely used in phraseology, "it displays many theoretical and practical similarities with the notions of reproducibility, fixedness and even idiomaticity" (Colson 2021: 28).
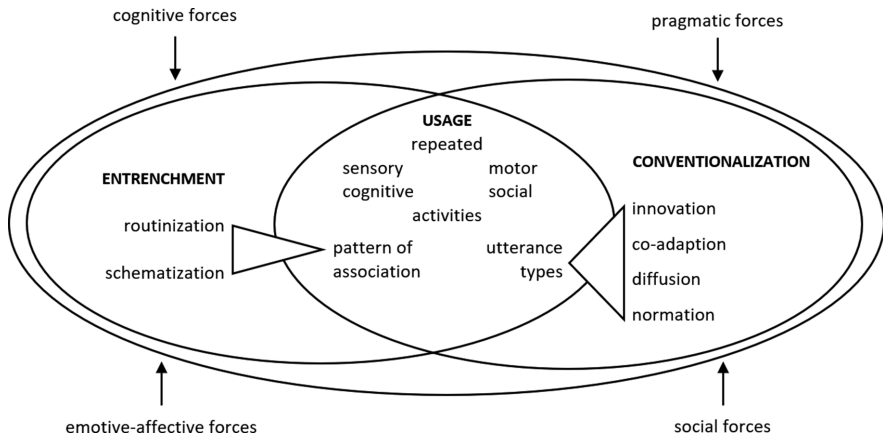
Figure 1: General outline of the Entrenchment-and-Conventionalization Model (taken from Schmid 2015: 7).

Figure 1 also shows that entrenchment and conventionalization interact in language use, whereby frequency plays a key role:

> [T]he usage frequency of a conventional utterance type supports the routinization of patterns of associations, which increases the likelihood of their activation and in turn the likelihood of repetition, which contributes to increasing conventionality. (Schmid 2015: 21)

The EC-model can be applied to describe the structure of (formulaic) language more precisely, since the model integrates different categories necessary for analyzing the linguistic system as

> a multidimensional dynamic contingency space populated by multidimensionally competing co-semiotic potentialities afforded by the interaction of speakers' usage activities and social and cognitive processes under the influence of a wide range of forces. (Schmid 2020: 348; in original with emphasis)

Schmid (2015: 11–16; 2020: 45–48) discusses various types of association and entrenchment effects. Of particular interest for this paper is, first, that Schmid (2015: 14; emphasis in original) assumes "routinization of syntagmatic associations" defined as "the repeated processing of *sequences of identical or similar linguistic elements*". He calls this process "cotextual entrenchment" (Schmid 2015: 15) and explicitly refers to cooccurrence patterns such as "emergent idioms, and other types of semi-fixed expressions" (Schmid 2015: 14). Second, Schmid (2015: 15) argues for the "routinization of paradigmatic association" called "emergent schematization", which interacts with syntagmatic associations in the emergence of

schemas. Thus, paradigmatic association refers to "the repetition of *different elements in an identical or similar cotextual or contextual environment*" (Schmid 2015: 14; emphasis in original). Looking at the example in Section 1, the frequent creative substitution of the components *Sein* and *(nicht) sein* can be seen as a paradigmatic association leading to the emergence of a partly lexicalized syntagmatic pattern [X *oder nicht* X/Y, *das ist hier die Frage*]. The dynamics of the multi-word expressions analyzed in Chapter 4 can be interpreted in the same way.

Regarding the conventionalization processes, the aspect of innovation is particularly important for this study. For Schmid (2015: 19) linguistic innovations are new utterances that are not instances of a conventionalized pattern, but that change such a pattern. Conventionalization and innovation are two sides of the same coin, as innovations are only innovative against the background of what is considered conventional in a speech community and not against the background of the mind of individual speakers (Schmid 2015: 19).

As an extension to Diachronic Construction Grammar, the EC-model is applied in this study, because it "distinguishes systematically between speakers and communities to provide a clearer picture of the way in which a given change unfolds" and it "defines conventions as dynamic regularities of behaviour and mental representations as dynamic patterns of associations" (Schmid 2021: 317). Thus, the benefit of the EC-model is that rather than considering and modeling cognitive and socio-pragmatic processes separately, both entrenchment and conventionalization and the relations between them are considered. The EC-model enables the description of "the systematicity and stability, on the one hand, and the flexibility and variability of languages, on the other" (Schmid 2015: 22). Furthermore, it is not just a theoretical model of language, but a model useful for empirical research. In particular, it can be used for the analysis of lexico-grammatical patterns (constructional idioms) and for the examination of change that take place over a few years, as shown by several studies (Schmid and Mantlik 2015; Mantlik and Schmid 2018; Schmid et al. 2020; Schmid et al. 2021). Since the model captures both the linguistic knowledge of individual speakers and the language use in the speech community, a differentiated methodological approach is needed to explore both sides equivalently.

4 Empirical Insights

4.1 Data and Methods

Besides a theoretical purpose, the paper also aims to propose a methodology for studying the dynamics of constructional idioms. Following the EC-model, two methods are chosen to examine both conventionalization and entrenchment.[11]

(a) Corpus Analysis: Conventionalization
Corpus analysis is used to investigate the variation and change of phrasemes in the speech community. To determine the dynamics of constructional idioms, especially regarding the process from a lexicalized multi-word expression, via the modification of this expression, to a semi-schematic pattern, it is not adequate to use static, synchronous corpus data. To explore micro-diachronies on selected constructions, we manually created six sub-corpora (each covering 5 years) based on the four archives (W, W2, W3, W4) of the German Reference Corpus. It is important to note that the German Reference Corpus consists mainly of these four archives, each containing approximately 10 to 12 billion tokens. However, the archives can only be used separately from each other. To find, therefore, a sufficient number of constructs for the selected constructional idioms, we carried out corpus analyses in the four archives in the different sub-corpora. Then we combined the lists of constructs covering the same periods. This is a complex procedure, since 24 search queries are carried out for each constructional idiom and then the lists covering the same periods must be merged. The result is six lists of constructs (for a particular constructional idiom) covering six different time periods (cf. Table 1).

For the analysis of the constructs, we use the tool "Lexical Pattern Analyzer" (lexpan),[12] which was developed in the IDS project "Usuelle Wortverbindungen" (Engl, 'multi-word expressions which are common in usage').[13] The tool identifies and calculates the fillers of the slots, which are presented in so-called filler tables. These tables can be used to gain results about preferred fillers and about the productivity of a construction. The analysis of sub-corpora can reveal changes in the productivity of the constructions as well as in the fillers of the slots. We generated

---

12 uwv.ids-mannheim.de/lexpan/ (March 1, 2024).
13 www.ids-mannheim.de/lexik/uwv/ (March 1, 2024).

Table 1: Compilation of the sub-corpora from the German Reference Corpus (in brackets the approximate number of tokens).

	Sub-corpus I	Sub-corpus II	Sub-corpus III	Sub-corpus IV	Sub-corpus V	Sub-corpus VI
Archive W	1995–1999 (1.2 billion)	2000–2004 (1.5 billion)	2005–2009 (2 billion)	2010–2014 (2 billion)	2015–2019 (2 billion)	2020–2022 (1.2 billion)
Archive W2	1995–1999 (47 million)	2000–2004 (1.6 billion)	2005–2009 (2.6 billion)	2010–2014 (2.8 billion)	2015–2019 (3.4 billion)	2020–2022 (1.9 billion)
Archive W3	1995–1999 (45 million)	2000–2004 (1.6 billion)	2005–2009 (2.2 billion)	2010–2014 (2.6 billion)	2015–2019 (3.4 billion)	2020–2022 (2.2 billion)
Archive W4	1995–1999 (34 million)	2000–2004 (1.8 billion)	2005–2009 (2.4 billion)	2010–2014 (2.9 billion)	2015–2019 (3 billion)	2020–2022 (1.4 billion)
Size in tokens	**approx. 1.33 billion**	**approx. 6.48 billion**	**approx. 9.34 billion**	**approx. 10.31 billion**	**approx. 11.83 billion**	**approx. 6.65 billion**

filler tables for the two constructions of the case study, covering five (cf. Section 4.2.1) respective six (cf. Section 4.3.1) time periods. In addition, we calculated the productivity of the constructions in the different sub-corpora (time periods), using the type-token ratio and the hapax-token ratio.

The sub-corpora cover a total of 28 years of contemporary German. Thus, the study asks about changes within a few decades and not about changes that take place over centuries. We follow Buerki (2019: 8), who has shown that formulaic language "can undergo perceptible shifts in usage over relatively short periods". Changes in the structure and semantics of idioms can thus be explored even over short periods of time. This is mainly because change in the field of formulaic language "proceeds very much faster than lexical change" (Buerki 2019: 29).[14] Overall, the micro-diachronic analysis of the sub-corpora provides insights into the variation and conventionalization of (constructional) idioms.

(b) Survey Study: Entrenchment
In previous research, entrenchment in the framework of the EC-model is mainly investigated (like conventionalization) using corpus studies ("from-corpus-to-cognition-

---

14 Buerki (2019: 29) concludes that change in formulaic language "is about a third faster than lexical change" and "appears not to be a (special) case of lexical change, but a different type of change altogether".

principle",[15] Schmid 2000: 38–40). However, in my opinion, it is also fruitful to complement corpus analysis with further empirical methods, especially when focusing on the mental consolidation of patterns (cf. also Van Lancker Sidtis et al. 2015). We therefore conducted an online survey of 21 formulaic expressions that can be traced back to a specific source.[16] Participants had to indicate the following for each expression (e.g., *Sein oder nicht sein, das ist hier die Frage*, cf. Section 1):

(a) I do not know the expression.
(b) I know the expression, but I do not know where it comes from.
(c) I know the expression and where it comes from.

If people indicated (c), they had to write in a free text field where they think the expression comes from. We evaluated and recoded the answers to determine whether the source was named correctly. A distinction was made between correct naming of the source (e.g., Shakespeare, Hamlet), partially correct naming of the source (e.g., from a literary work) and incorrect naming of the source (e.g., quote from a politician). In this way, it is possible to distinguish whether participants who think they know the source of an idiom know the original context.

Formulaic expressions were selected for the survey study where it can be assumed that they have developed or are developing into more schematic constructions over time due to extensive modification. In addition, formulaic expressions were chosen whose original usage comes from different domains (such as politics, literature, movies, advertising, fashion, sports, social media, religion).[17] We assume that especially sentence-like expressions are modified by substituting components, resulting in open slots in the structure. Sentence-like phrasemes include, above all, proverbs, and catchphrases. The latter are the focus of this study, as we are interested in whether speakers know and can name the source of such phrasemes.

---

15 The principle states the following: "Frequency in text instantiates entrenchment in the cognitive system" (Schmid 2000: 39).

16 The online survey was conducted between 9 July and 11 August 2023 using the tool SoSci Survey (www.soscisurvey.de/ March 1, 2024).

17 The survey consists of German catch phrases such as *Am Anfang war das Wort, Es ist besser, nicht zu regieren, als falsch zu regieren, Ich habe heute leider kein Foto für dich, Nach dem Spiel ist vor dem Spiel, Schlaflos in Seattle, Sein oder nicht sein, das ist hier die Frage, Wohnst du noch oder lebst du schon?*. The compilation is oriented on a list of German snowclones that can be found online: emmanuel.dammerer.at/snowclonerey (March 1, 2024). Other collections of (German) catchphrases (winged words) that are referred to include Duden (2021), the compilation of famous sentences from politicians by Klein (2013, among others), and various lists of funny and popular quotes from celebrities (e.g., footballers, actors) that can be found on the Internet.

We collected metadata (such as age, education level, media consumption, and interests) on the 270 participants of the online survey to find out sociodemographic differences in knowledge of idioms.[18] Table 2 presents the frequency distributions of the knowledge of the idioms that we will focus on in this paper (*Wohnst du noch oder lebst du schon?* [Engl. 'Are you still residing or are you already living?'] and *Nach dem Spiel ist vor dem Spiel* [Engl. 'After the game is before the game']) as well as the frequency distributions of the independent variables. It must be emphasized that some participants did not give any statements for some idioms and metadata queries (for this reason, the numbers of participants in the two case studies are somewhat lower, cf. Sections 4.2.2 and 4.3.2). The central question is whether speakers know certain idioms and whether they can indicate the source of these idioms. If speakers know the source, this suggests that they have more knowledge of the construction (regarding discourse-functional [etymological] attributes) than if speakers do not know the source. If speakers do not know the idiom at all, this logically suggests that they have no entry in the mental lexicon for the fully lexicalized (original) expression.

The online survey method can also be seen as an apparent-time approach if we focus on the variable 'age' (Labov 1963; Bailey et al. 1991; Cukor-Avila and Bailey 2013). This sociolinguistic approach assumes that "differences among generations of adults mirror actual diachronic developments in a language" (Bailey et al. 1991: 242). Against this background, by comparing younger and older speakers, differences in knowledge of lexico-grammatical constructions can be determined that indicate language change.

For this paper, results of the corpus analysis and the survey on two phrasemes are presented. The two phrasemes originate from different domains and decades. The first case study examines the idiom *Wohnst du noch oder lebst du schon?*, which goes back to a slogan of the furniture company IKEA (cf. Section 4.2). The second case study focuses on the idiom *Nach dem Spiel ist vor dem Spiel*, which

---

18 Media consumption was queried using the following scale: (1) never, (2), once a month or less often, (3) two or three times a month, (4) about once a week, (5) several times a week, (6) (almost) daily, (7) several times a day. Interest was asked with the following scale: (1) do not agree at all, (2) disagree, (3) undecided, (4) agree, (5) fully agree. Education is divided into the following items (they are partly listed in German, since there are no equivalents in English due to the different school systems): (1) Volkshoch-/Hauptschulabschluss (school leaving qualification), (2) Mittlere Reife, Realschulabschluss (intermediate school-leaving certificate) or equivalent education, (3) Fachabitur, Fachhochschulreife (subject-related entrance qualification, specialized A-levels), (4) Abitur, Hochschulreife (A-levels, university entrance qualification). Age is composed as follows: (1) 11–29 years, (2) 20–39 years, (3) 40–49 years, (4) 50–59 years, (5) 60–69 years, (6) 70 and older.

Table 2: Frequency distributions of the variables in the data set of the online survey.

Variable	M	SD	Min	Max
Knowledge of the expression *Wohnst Du noch oder lebst Du schon?*	2.73	.52	1.00	3.00
Answer recoded by the researcher: knowledge of the expression *Wohnst du noch oder lebst Du schon?*	2.94	.24	2.00	3.00
Knowledge of the expression *Nach dem Spiel ist vor dem Spiel*	2.30	.58	1.00	3.00
Answer recoded by the researcher: knowledge of the expression *Nach dem Spiel ist vor dem Spiel*	2.34	.24	1.00	3.00
Media consumption television	5.26	1.40	1.00	7.00
Media consumption newspaper	4.55	1.95	1.00	7.00
Media consumption social media	4.64	2.49	1.00	7.00
Media consumption radio	4.97	1.91	1.00	7.00
Media consumption magazines	2.50	1.47	1.00	7.00
Interest in movies	3.84	.96	1.00	5.00
Interest in literature	3.75	1.10	1.00	5.00
Interest in music	4.15	.87	1.00	5.00
Interest in social media	3.13	1.28	1.00	5.00
Interest in politics	3.69	1.45	1.00	5.00
Interest in football (soccer)	2.39	1.45	1.00	5.00
Interest in advertising	1.84	.97	1.00	5.00
Interest in fashion	2.90	1.18	1.00	5.00
Interest in technology	3.01	1.14	1.00	5.00
Interest in religion	2.45	1.23	1.00	5.00
Age	2.88	1.58	1.00	6.00
Education	3.46	.93	1.00	4.00

comes from a former German football coach (cf. Section 4.3). The two examples were chosen because, although they undergo a similar constructionalization, the changes are quite different regarding conventionalization and entrenchment.

4.2 Case Study 1: [X *du noch oder* Y *du schon?*]

4.2.1 Corpus Analysis

The first case study deals with the multi-word expression *Wohnst du noch oder lebst du schon?,* which was originally a slogan created by the furniture company IKEA in 2002.[19] Examples (6)–(8) show that the two verbs *wohnen* and *leben* are often substituted.

---

19 www.slogans.de/slogans.php?GInput=wohnst&SCheck=1 (March 1, 2024).

(6) **Mietest du noch oder besitzt du schon?** Die Flucht ins „Betongold" treibt auch in Deutschland die Immobilienpreise hoch.
(Nürnberger Zeitung, August 4, 2012)
[Engl. 'Are you still renting or do you already own?']

(7) Dabei geht es nicht um Inhalte, sondern leidenschaftlich wird diskutiert, wie man es sich einverleiben soll, das Buch. **Liest du noch oder kindlest du schon?**
(St. Galler Tagblatt, November 26, 2012)
[Engl. 'Are you still reading or are you already kindling?']

(8) „**Koblenz, planst du noch oder baust du schon?**" Das fragten die Wirtschaftsjunioren Mittelrhein [. . .].
(Rhein-Zeitung, February 8, 2005)
[Engl. 'Koblenz, are you still planning or are you already building?']

We used the five sub-corpora from 2000–2022 for the analysis. Since the slogan was created in 2002, it makes little sense to use the sub-corpus covering the years 1995–1999.[20] Table 3 illustrates the 10 most frequent fillers for the analyzed periods.[21]

Table 3: The ten most frequent fillers of the multi-word expression [X *du noch oder* Y *du schon?*] in the German Reference Corpus between 2000 and 2022.

2000–2004			2005–2009		
Filler	Freq.	Percentage	Filler	Freq.	Percentage
Wohnst . . . lebst	96	24.74	*Wohnst . . . lebst*	215	18.96
Lernst . . . lebst	26	6.70	*Poppst . . . zeugst*	34	3.00
Lachst . . . denkst	19	4.90	*Lebst . . . wohnst*	21	1.85
Lebst . . . wohnst	16	4.12	*Arbeitest . . . bettelst*	20	1.76
Lernst . . . sparst	11	2.84	*Lebst . . . stirbst*	20	1.76
Schraubst . . . wohnst	8	2.06	*Rennst . . . lebst*	18	1.59
Spielst . . . lebst	7	1.80	*Suchst . . . lebst*	18	1.59

---

20 Among a total of 5,322 hits that can be found for the pattern [X *du noch oder* Y *du schon?*] in the four DeReKo archives, only one example dates from a year before 2002, when IKEA created the slogan: 'Doch der ‚Blindflug' hat seine Tücken: Geht es bergauf oder bergab, **fährst du noch oder stehst du schon?**' (Süddeutsche Zeitung, December 20, 1996) [Engl. 'Are you still driving or are you already standing?']. It can therefore be strongly assumed that the IKEA slogan is the starting point for the conventionalization of this syntactic pattern.
21 Search query in COSMAS II: (((du /+w2 „noch") /+w1 oder) /s0 (du /+w1 schon)).

Table 3 (continued)

2000–2004			2005–2009		
Filler	**Freq.**	**Percentage**	**Filler**	**Freq.**	**Percentage**
Träumst . . . lebst	6	1.55	*Glaubst . . . denkst*	14	1.23
lernst . . . lebst	6	1.55	*Rauchst . . . lebst*	13	1.15
Suchst . . . googlest	5	1.29	*Lachst . . . schwingst*	9	0.79

2010–2014			2015–2019		
Filler	**Freq.**	**Percentage**	**Filler**	**Freq.**	**Percentage**
Wohnst ... lebst	114	9.53	*Wohnst ... lebst*	123	11.93
Sprichst ... kommunizierst	60	5.02	*Versteuerst ... lebst*	48	4.66
Rennst ... lebst	36	3.01	*Chillst ... kletterst*	30	2.91
Mietest ... kaufst	35	2.93	*Träumst ... lebst*	21	2.04
Übersetzt ... verstehst	20	1.67	*Kehrst ... zahlst*	16	1.55
Versteuerst ... lebst	18	1.51	*Schwitzt ... klebst*	15	1.45
Träumst ... gründest	15	1.25	*Fährst ... teilst*	14	1.36
Fragst ... glaubst	14	1.17	*Lachst ... swingst*	13	1.26
Lernst ... lebst	14	1.17	*Lebst ... wohnst*	13	1.26
Lebst ... wohnst	13	1.09	*Lebst ... funktionierst*	12	1.16

2020– 2022		
Filler	**Freq.**	**Percentage**
Wohnst ... lebst	38	12.26
Lebst ... stirbst	20	6.45
Fliegst ... fährst	9	2.90
lebst ... stirbst	9	2.90
Lernst ... verstehst	8	2.58
Grillst ... smokst	7	2.26
Siezt ... duzt	7	2.26
Prüfst ... impfst	6	1.94
Hoffst ... schraubst	5	1.61
Lebst ... wohnst	5	1.61

It is noticeable that in the period from 2000–2004 the original multi-word expression occurs in only about 25% of all hits. This means that the slogan was modified in numerous ways immediately after it was created. In the most recent sub-corpus (2020–2022), the original phraseme is used in only about 12% of all hits. Overall, the frequency of the original expression decreases over time compared to other realizations. It is interesting to note that the type-token ratio as well as the hapax-token ratio do not change much (cf. Table 4). If at all, we can speak of a slight increase in the type-token ratio when comparing the two sub-corpora from 2000–2004 (0.39) and from 2020–2022 (0.46). Thus, comparing only these two sub-corpora marking the beginning

and the end of the period, we can say that the construction becomes more productive. However, regarding the data, it is worth mentioning that these two corpora contain the least number of hits of the construction. Thus, it can be questioned how comparable these two are with the other three corpora (2005–2009, 2010–2014, 2015–2019).

Table 4: Type-token and hapax-token ratio of the multi-word expression [X *du noch oder* Y *du schon?*] in the German Reference Corpus between 2000 and 2022.

Sub-corpus / time period	2000–2004	2005–2009	2010–2014	2015–2019	2020–2022
Types	153	485	496	416	142
Tokens	388	1,134	1,196	1,031	310
Hapax legomenons	107	335	314	268	91
Type-token ratio	**0.39**	**0.43**	**0.41**	**0.4**	**0.46**
Hapax-token ratio	**0.28**	**0.3**	**0.26**	**0.26**	**0.29**

It should be noted, that the two verbs *wohnen* and *leben* are not always replaced completely. On the one hand, there are examples in which the verbs are used in reversed order (cf. [9]).

(9) Der arme Mann musste leiden für einen Möbeltransport. Ich würde ihn jetzt gern fragen: **Lebst du noch oder wohnst du schon?**
     (Braunschweiger Zeitung, March 16, 2006)
     [Engl. 'Are you still alive or are you already residing?']

On the other hand, there are many instances in which one of the two verbs is still used, either in the first or second position. Thus, in these instances, only one of the original verbs is substituted with a new verb. Four variants can be observed:
– [*Wohnst du noch oder* X *du schon?*] (e.g., *Wohnst du noch oder liebst du schon?* [Engl. 'Are you still living or are you already in love?'])
– [X *du noch oder lebst du schon?*] (e.g., *Träumst du noch oder lebst du schon?* [Engl. 'Are you still dreaming or are you already living?'])
– [*Lebst du noch oder* X *du schon?*] (e.g., *Lebst du noch oder guckst du schon?* [Engl. 'Are you still living or are you already looking?'])
– [X *du noch oder wohnst du schon?*] (e.g., *Suchst du noch oder wohnst du schon?* [Engl. 'Are you still searching or are you already living?'])

If we add up the tokens of all instances in which the lexemes *wohnen* and/or *leben* occur, we find an overall decrease of these verbs (cf. Table 5). In the period between 2000–2004, it is still 65% of all hits in which at least one of the verbs is realized. In the period between 2005–2009 this percentage already falls to 50%.

From 2010–2014, the proportion decreases to about 40% and remains at this level until the period between 2020–2022. This means that the original verbs *wohnen* and *leben* are used less frequently within the 20 years examined. Overall, the corpus data show that a process of lexical constructionalization (cf. Section 3.2) from a more lexicalized idiom to a more schematic idiom can be observed.

Table 5: Proportion of the verbs *wohnen* and *leben* within the constructs of the multi-word expression [X *du noch oder* Y *du schon?*] in the German Reference Corpus between 2000 and 2022.

Sub-corpus / time period	2000–2004	2005–2009	2010–2014	2015–2019	2020–2022
Tokens / constructs in the sub-corpus	**388**	**1,134**	**1,196**	**1,031**	**310**
Wohnst du noch oder lebst du schon?	96	215	114	123	38
wohnst du noch oder lebst du schon?	4	3	4	2	1
Lebst du noch oder wohnst du schon?	17	22	14	14	6
Wohnst du noch oder X *du schon?*	21	58	66	65	13
X *du noch oder lebst du schon?*	75	150	164	165	30
Lebst du noch oder X *du schon?*	25	105	64	44	34
X *du noch oder wohnst du schon?*	14	21	41	54	8
Tokens with the lexemes *wohnen* and/or *leben* (in total and in percentage)	**252 (65%)**	**574 (50.6%)**	**467 (39%)**	**467 (45.3%)**	**130 (41.9%)**

The meaning as well as the semantic change of the construction can be described as follows. The starting point is the catchphrase (winged word) *Wohnst du noch oder lebst du schon?* with a specific meaning based on the original advertising context. A paraphrase might be: 'Do you have only functional furniture in your apartment (bed, table, chair) or have you made your apartment your home (i.e. furnished it comfortably, decorated it, etc.) (attitude towards life "hygge")?'. The catchphrase moves in the continuum from lexicon to syntax in a very short time because of the frequent modification and the formation of slots. In other words, the fully lexicalized phraseme with specific meaning and strong contextual binding develops into a partly lexicalized constructional idiom with a more abstract, wider meaning and more flexible function. The use in other contexts entails a strong extension of the meaning.

The construction no longer refers only to the narrow semantic domain of 'living', but to numerous other domains. The semi-schematic construction has the fol-

lowing meaning: 'Are you (still) doing activity X or are you (already) doing activity Y, which has a more positive connotation compared to activity X (e.g., more advanced, healthier, etc.)?' or 'Used when asking whether one state is still desirable or another is preferable'.[22] The variants in which either *wohnen* or *leben* are realized (cf. Table 5) stand between the fully lexicalized catchphrase and the constructional idiom. One thing is clear: with the formal change and the formation of slots, the meaning of the construction also expands. Schematization enables verbs other than *wohnen* and *leben* to appear in the construction.

It should be noted that there are also instances in which the constructional meaning described deviates and in which the second verb does not have a more positive connotation than the first. This variation can be observed in some instances where only one slot exists in the second position and *wohnen* or *leben* is realized in the first position (cf. [10] and [11]).

(10) *Wohnst du noch oder haust/kapitulierst/klagst/stirbst/weinst* etc. *du schon?*
      [Engl. 'Are you still residing or are you already dwelling/capitulating/complaining/dying/crying etc.?']

(11) *Lebst du noch oder arbeitest/dopst/hartzt/herbst/schuftest/stirbst/vegetierst*
      etc. *du schon?*
      [Engl. 'Are you still living or are you already working/doping/being on welfare/harvesting/shuffling/dying/vegging etc.?']

However, there are also some instances in which neither *wohnen* nor *leben* is used and in which the second verb has a more negative connotation than the first. For instance, *Trinkst du noch, oder säufst du schon?* in a report about alcohol addiction (cf. [12]), *Fährst du noch, oder kriechst du schon?* as a quote from a politician on the idea of lowering the speed limit in cities (cf. [13]) and *Arbeitest du noch oder bettelst du schon?* in the context of strikes (cf. [14]).

(12) „**Trinkst du noch, oder säufst du schon?**" Die Grenzen zwischen unschädlichem Alkoholgenuss und Sucht sind fließend. Die gute Nachricht ist: „Man kann zu jedem Zeitpunkt etwas dagegen tun", sagt Suchtexperte Wolf Dietrich Braunwarth.
      (Nürnberger Nachrichten, May 1, 2013)
      [Engl. 'Are you still drinking, or are you already boozing?']

---

22 www.owid.de/artikel/404090 (March 1, 2024).

(13) Die Lobby der Autofahrer läuft trotzdem Sturm. „**Fährst du noch, oder kriechst du schon?**", lästert die FDP und befürchtet, dass die „wirtschaftliche Entwicklung" der Stadt leidet.
(die tageszeitung, November 20, 2010)
[Engl. 'Are you still driving, or are you already crawling?']

(14) Im dichten Schneetreiben wurden Plakate geschwenkt: „**Arbeitest du noch oder bettelst du schon?**" Das Motto einer kleinen, aber wichtigen neuen Streikgruppe unter den Demonstranten.
(Hamburger Morgenpost, March 11, 2006)
[Engl. 'Are you still working or are you already begging?']

Thus, another meaning can be formulated in which the second verb has a more negative connotation than the first, describing a state that is not desirable or preferable to another. The construction thus also undergoes an expansion of meaning. While in the original slogan the first verb had a more negative connotation than the second (in the opinion of the advertisers and the company IKEA), the more schematic construction also allows the slots to be filled with verbs in which the connotation is in the other way around (a verb with a more positive connotation is followed by a verb with a more negative connotation).

Notably, there are examples in which meta-linguistic comments on the origin of the construction can be found. In (15) and (16), the underlined phrases make a direct reference to the original advertising context (*Werbeslogan/-spruch*). In addition, the expressions are referred to as variations (*Abwandlungen*) from the original slogan. Such meta-linguistic corpus comments indicate that knowledge of the origin of the construction is still present in the speech community (*bekannten*).

(15) **Deckst du noch oder dämmst du schon?** <u>So könnte man es in Abwandlung eines bekannten Werbeslogans formulieren</u>. Gemeint sind die Dachdecker, deren Berufstätigkeit sich mehr und mehr den modernen Erfordernissen angepasst hat.
(Berliner Morgenpost, March 4, 2011)
[Engl. 'Are you still covering or are you already insulating? This is how one could put it in a variation of a well-known advertising slogan.']

(16) In Hessen zeigt sich unterdessen, wie schwierig eine von der Lafontaine-Partei tolerierte Regierung werden könnte. Da sagen Linke wie Wolfgang Gehrke bereits lange vorher, sie würden den eher konservativen Sozialdemokraten Jürgen Walter nicht zum Minister wählen. <u>In Abwandlung eines</u>

bekannten Werbespruchs werden ihn wohl manche fragen: **Duldest du noch, oder regierst du schon?**
(Nürnberger Nachrichten, August 13, 2008)
[Engl. 'In a variation of a well-known advertising slogan, some will probably ask him: Do you still tolerate, or do you already rule?']

The following must be mentioned: If we look at the constructs that deviate from the original phraseme *Wohnst du noch oder lebst du schon?* in context, we see that they are quite often titles of, for instance, cabaret events (cf. [17]) or books (cf. [18]). Because of their title character, these are used in different texts, which increases their frequency of use and thus their percentage of all hits.

(17)   Dabei bringen die Kabarettisten aus der Zweiburgenstadt alltägliche Kurio-sitäten so gekonnt auf den Punkt, dass der ganze Saal vor Lachen bebt. **„Lachst Du noch oder denkst Du schon?"** heißt denn auch das neue Pro-gramm.
(Mannheimer Morgen, February 12, 2004)
[Engl. 'Are you still laughing or are you already thinking?']

(18)   Er haut nicht, er schaut den Deutschen aufs Maul: Satiriker Wiglaf Droste stellt heute in Nürnberg sein neues Buch **„Sprichst du noch oder kommu-nizierst du schon?"** vor.
(Nürnberger Zeitung, April 17, 2012)
[Engl. 'Are you still talking or are you already communicating?']

Here, an analogy process can be observed by using the advertising slogan of IKEA in a modified form for the naming of certain titles (Bebermeyer and Bebermeyer 1977). The semi-schematic construction that emerges is thus strongly tied to the context of the use of naming titles. Indeed, such instances are not the productive filling of the slots, but merely the repetition of fixed titles. In other words, an in-stance of a construction is repeated, i.e., the construction is not instantiated with new lexical material. The percentage distribution of the slot fillers (cf. Table 3) should thus be considered with caution. However, it cannot be denied that in the various corpora the use of the original phraseme is decreasing. Thus, the naming of titles by modifying the slogan has itself become a usage pattern.

4.2.2 Online Survey

That the modified slogan is used very frequently to name titles could be explained by the fact that the slogan is still well known in the speech community, as the results of the online survey show. Thus, out of 267 participants, 204 (76.4%) say that they know where the multi-word expression comes from. 53 (19.9%) people know the idiom, but not where it comes from. Only 10 (3.7%) participants said they are not familiar with the construction. Furthermore, after checking the answers to which context the expression is related, we conclude that 192 of the 204 answers are completely correct. The participants therefore know that the expression is an IKEA advertising slogan. 12 people were able to determine the rough context, for instance, by stating only "advertising". No one named a source that is completely wrong or too imprecise.

Since the number of those who know the expression is very high (approx. 96% of the participants), there are very few significant correlations regarding the sociodemographic factors queried (cf. Table 6). For instance, a certain media con-

Table 6: Simple correlations between the knowledge of the multi-word expression *Wohnst du noch oder lebst du schon?* (not recoded by the researcher) and the other variables.[23]

Knowledge of the expression *Wohnst du noch oder lebst du schon?*	Correlation
Media consumption television	.01
Media consumption newspaper	.07
Media consumption social media	−.11
Media consumption radio	.10
Media consumption magazines	.03
Interest in movies	−.05
Interest in literature	.14*
Interest in music	.05
Interest in social media	−.12*
Interest in politics	.13*
Interest in football (soccer)	−.10
Interest in advertising	−.03
Interest in fashion	−.11
Interest in technology	−.02
Interest in religion	.01
Age	−.00
Education	.18**

---

23 *p < .05; **p < .01; ***p < .001. Variables were assumed to be interval scaled. Accordingly, Pearson's r was used.

sumption is non-significant at the 5 percent level. Age also shows no statistically significant correlation between whether someone knows the expression and its context of origin. The only statistically significant results are the following (marked with an asterisk in Table 6): the more interested people are in literature and politics and the higher their level of education, the more likely they are to know the expression. The more interested the participants are in social media, the less familiar the expression is. However, the data should be interpreted with caution.

Considering this result, it makes little sense to statistically analyze the data of those who have indicated a source. This is because there is no completely incorrect naming of the original context. Nevertheless, the answers are very interesting considering the EC-model (cf. Section 3.3), as these are quite different (cf. Table 7). Some people were able to specify the advertising context more precisely than others (e.g., indicating the time when the slogan was born and was common in the media [cf. interview numbers 88, 327, 266]).[24] This suggests that for them the knowledge of its origin is stored with greater precision in the mental lexicon than for other participants who only gave a rather vague or only partly correct description of the source (e.g., some participants did not name IKEA but other furniture companies or home improvement stores [cf. interview numbers 546, 528, 421). Thus, there are interindividual differences in the knowledge of (the origin of) this construction.

It must be emphasized that the results and their interpretation are based on the (in part quite brief) answers of the participants. Some participants may know the exact original context of the expression, but they only provided the broader context in the survey. Reasons for this could be that the participants were in a hurry or that they considered it sufficient to classify the expression in the advertising context in general without indicating that it originates from IKEA.

4.3 Case Study 2: [*Nach* X *ist vor* X/Y]

4.3.1 Corpus Analysis

The second example is the multi-word expression *Nach dem Spiel ist vor dem Spiel* which goes back to legendary German football coach Sepp Herberger (1897–1977).[25]

---

24 The slogan dates to 2002. While some people were quite specific about when the slogan was first used (cf. Table 7), some cited, for instance, the 1990s and others even the 2020s as the origin of the slogan.

25 de.wikipedia.org/wiki/Sepp_Herberger (March 1, 2024).

Table 7: Selected answers about the origin of the multi-word expression *Wohnst du noch oder lebst du schon?*

Accuracy of the answer	Answer	Sociodemographic information about the participant (looking at age and other statistically significant parameters)
Precise indication of the original context	Werbeslogan Ikea seit ca 2005?! [Engl. 'Advertising slogan Ikea since ca 2005?!']	Interview number: 88 Age: 30–39 Education: A-levels, university entrance qualification Interest in literature: agree Interest in social media: agree Interest in politics: agree
	Werbeslogan der schwedischen Möbel-Firma IKEA (2000er-Jahre) [Engl. 'Advertising slogan of the Swedish furniture company IKEA (2000s)']	Interview number: 327 Age: 50–59 Education: A-levels, university entrance qualification Interest in literature: fully agree Interest in social media: do not agree at all Interest in politics: fully agree
	IKEA Werbeslogan, ca 2000–2006 [Engl. 'IKEA advertising slogan, ca 2000–2006']	Interview number: 266 Age: 60–69 Education: A-levels, university entrance qualification Interest in literature: fully agree Interest in social media: undecided Interest in politics: fully agree
Rough or partly correct indication of the original context	Werbung [Engl. 'Advertising']	Interview number: 546 Age: 60–69 Education: Volkshoch-/Hauptschulabschluss (school leaving qualification) Interest in literature: agree Interest in social media: agree Interest in politics: agree
	Werbeslogan Baumarkt [Engl. 'Advertising slogan hardware store']	Interview number: 528 Age: 50–59 Education: A-levels, university entrance qualification Interest in literature: agree Interest in social media: do not agree at all Interest in politics: agree

Table 7 (continued)

Accuracy of the answer	Answer	Sociodemographic information about the participant (looking at age and other statistically significant parameters)
	Werbung eines Möbelhauses (raab?) [Engl. 'Advertising of a furniture store (raab?)']	Interview number: 421 Age: 11–29 Education: A-levels, university entrance qualification Interest in literature: fully agree Interest in social media: undecided Interest in politics: disagree

Articles in newspapers and magazines, as well as the Wikipedia entry on common sayings, point out that this quote has become a catchphrase in German.[26] Unfortunately, it is not documented in which year Sepp Herberger made this statement.[27] In addition, it must be mentioned that from today's perspective, it cannot be said with absolute certainty that it was Sepp Herberger who used the pattern [*Nach* X *ist vor* X/Y] for the first time.[28]

However, there is knowledge in the speech community that the expression *Nach dem Spiel ist vor dem Spiel* comes from Sepp Herberger. Several instances of both the original phraseme (cf. [19]) and modified forms (cf. [20]) are linked to the source utilizing meta-linguistic comments (*diese Weisheit Sepp Herbergers* [Engl. 'This wisdom of Sepp Herberger'; *frei nach Sepp Herberger* [Engl. 'freely after Sepp Herberger'].

---

26 de.wikipedia.org/wiki/Liste_gefl%C3%BCgelter_Worte/N#Nach_dem_Spiel_ist_vor_dem_Spiel (March 1, 2024).

27 de.wikiquote.org/wiki/Diskussion:Sepp_Herberger (March 1, 2024).

28 The question of whether the syntactic pattern already existed before the quote can only be determined by analyzing corpora that contain texts prior to Herberger's lifespan. One problem that cannot be solved is that – as mentioned above – it is not known in which year Herberger said this sentence. However, it is interesting to note that the earliest hits for the semi-schematic pattern from the mid-1970s in the four archives of the German Reference Corpus are Herberger's quote. The earliest example is from 1975: "Eine Generation von Trainern setzte Herbergers scheinbar platte Binsenweisheiten in Erfolge um: 'Nach dem Spiel ist vor dem Spiel' [. . .]." (Der Spiegel, June 16, 1975). In addition, there are no hits for the syntactic pattern in the HIST archive of the German Reference Corpus, which comprises 5,245 texts with a size of 66,582,941 tokens from the second half of the 17th century to 1962.

(19) **Nach dem Spiel ist vor dem Spiel**: <u>Diese Weisheit von Sepp Herberger</u> gilt auch für die Sicherheitskräfte, die alle drei Spiele der Mini-WM in Nürnberg als Testlauf für die Weltmeisterschaft in einem Jahr verstanden haben.
(Nürnberger Nachrichten, June 27, 2005)
[Engl. 'After the game is before the game']

(20) <u>Frei nach Sepp Herberger</u> gilt in der Rentengesetzgebung aber unverändert weiter: **Nach der Reform ist vor der Reform**.
(Protokoll der Sitzung des Parlaments Deutscher Bundestag, October 1, 2004)
[Engl. 'After the reform is before the reform']

The corpus study also shows that the catchphrase has an enormous influence on the total number of constructs of the pattern [*Nach* X *ist vor* X/Y] (cf. Table 8).[29]

Table 8: The ten most frequent fillers of the multi-word expression [*Nach* X *ist vor* X/Y] in the German Reference Corpus between 1995 and 2022.

1995–1999			2000–2004		
Filler	Freq.	Percentage	Filler	Freq.	Percentage
Spiel . . . Spiel	104	40.47	*Wahl . . . Wahl*	766	19.19
Wahl . . . Wahl	39	15.18	*Spiel . . . Spiel*	661	16.56
Krieg . . . Krieg	18	7.00	*Reform . . . Reform*	227	5.69
Hochwasser . . . Hochwasser	10	3.89	*Fest . . . Fest*	139	3.48
Rennen . . . Rennen	6	2.33	*Saison . . . Saison*	136	3.41
Saison . . . Saison	4	1.56	*Flut . . . Flut*	80	2.00
Schlacht . . . Schlacht	3	1.17	*Krieg . . . Krieg*	79	1.98
Demo . . . Demo	2	0.78	*Rennen . . . Rennen*	42	1.05
Fest . . . Fest	2	0.78	*Festival . . . Festival*	38	0.95
Film . . . Film	2	0.78	*Turnier . . . Turnier*	35	0.88
2005–2009			2010–2014		
Filler	Freq.	Percentage	Filler	Freq.	Percentage
Spiel . . . Spiel	1,496	15.41	*Wahl . . . Wahl*	1,516	10.52
Wahl . . . Wahl	1,472	15.16	*Spiel . . . Spiel*	1,369	9.50
Saison . . . Saison	428	4.41	*Saison . . . Saison*	811	5.63
Fest . . . Fest	364	3.75	*Fest . . . Fest*	584	4.05
Reform . . . Reform	207	2.13	*Krise . . . Krise*	295	2.05

29 Search query in COSMAS II: ((nach /+w1 (dem oder (der oder den))) /+w3 ((ist /+w1 vor) /+w2 (dem oder (der oder den)))).

Table 8 (continued)

2005–2009			2010–2014		
Filler	Freq.	Percentage	Filler	Freq.	Percentage
Krise . . . Krise	144	1.48	*Turnier . . . Turnier*	209	1.45
Turnier . . . Turnier	144	1.48	*Reform . . . Reform*	162	1.12
WM . . . WM	137	1.41	*WM . . . WM*	161	1.12
Rennen . . . Rennen	111	1.14	*Festival . . . Festival*	152	1.05
WM . . . EM	106	1.09	*Session . . . Session*	148	1.03
2015–2019			2020–2022		
Filler	Freq.	Percentage	Filler	Freq.	Percentage
Wahl . . . Wahl	1,600	8.60	*Wahl . . . Wahl*	607	7.90
Spiel . . . Spiel	1,204	6.47	*Spiel . . . Spiel*	461	6.00
Saison . . . Saison	1,074	5.77	*Saison . . . Saison*	379	4.93
Fest . . . Fest	710	3.82	*Krise . . . Krise*	189	2.46
Festival . . . Festival	238	1.28	*Fest . . . Fest*	161	2.09
Turnier . . . Turnier	228	1.23	*Pandemie . . . Pandemie*	115	1.50
Baustelle . . . Baustelle	217	1.17	*Streik . . . Streik*	104	1.35
Konzert . . . Konzert	180	0.97	*Derby . . . Derby*	103	1.34
Derby . . . Derby	173	0.93	*Sturm . . . Sturm*	93	1.21
Session . . . Session	166	0.89	*Baustelle . . . Baustelle*	76	0.99

In the first sub-corpus (1995–1999), the filler *Spiel . . . Spiel* represents more than 40% of all instances. *Wahl . . . Wahl* follows on the second rank with about 15%. However, this distribution should be taken with caution, as the total number of constructs is not very high. Overall, a decrease of the filler *Spiel . . . Spiel* can be observed. Remarkably, the filler *Wahl . . . Wahl* is even used more frequently than *Spiel . . . Spiel* in the sub-corpus 2000–2004 and then consistently from sub-corpus 2010–2014 onwards.[30] The construction is therefore very often used in the press after (important) elections, to emphasize that the next election is already soon (cf. [21]). However, the use of *Wahl . . . Wahl* also decreases during this short period.

(21) **Nach der Wahl ist vor der Wahl**: Während in Berlin noch darüber debattiert wird, wer mit wem und unter welchen Umständen die nächste Bundesregierung bildet, richten die kleineren Parteien den Blick bereits auf das kommende Wahljahr.
(Norddeutsche Neueste Nachrichten, November 19, 2013)
[Engl. 'After the election is before the election']

---

30 In sub-corpus 2005–2009 *Spiel . . . Spiel* is used slightly more often than *Wahl . . . Wahl*.

Across the sub-corpora, we find other lexemes that are (almost consistently) among the top ten fillers (*Saison . . . Saison* [Engl. 'season'], *Fest . . . Fest* [Engl. 'party'], *Turnier . . . Turnier* [Engl. 'tournament']). Their use does not change much over time. However, the use of *Saison . . . Saison* is becoming closer to that of *Wahl . . . Wahl* and *Spiel . . . Spiel* (cf. Figure 2).
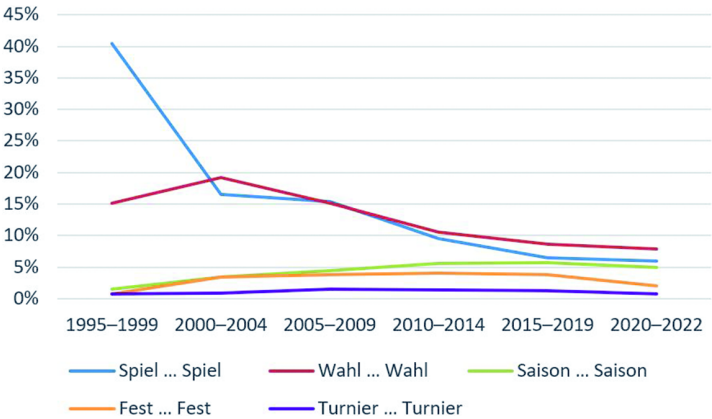
Figure 2: Usage frequency of selected fillers in the multi-word expression [*Nach* X *ist vor* X/Y] in the German Reference Corpus between 1995 and 2022.

If we exclude the first period due to the small number of absolute hits, it is evident that the construction becomes more productive. Between 2000–2004 and 2020–2022, the type-token ratio increases from 0.19 to 0.27 and the hapax-token ratio from 0.12 to 0.18 (cf. Table 9).

Table 9: Type-token and hapax-token ratio of the multi-word expression [*Nach* X *ist vor* X/Y] in the German Reference Corpus between 1995 and 2022.

Sub-corpus / time period	1995–1999	2000–2004	2005–2009	2010–2014	2015–2019	2020–2022
Types	69	740	1,708	3,003	4,293	2,064
Tokens	257	3,992	9,711	14,411	18,598	7,686
Hapax legomenons	51	475	1,088	2,016	2,813	1,366
Type-token ratio	**0.27**	**0.19**	**0.18**	**0.21**	**0.23**	**0.27**
Hapax-token ratio	**0.2**	**0.12**	**0.11**	**0.14**	**0.15**	**0.18**

The first noun is not repeated in the second slot in all instances. There are numerous cases where the two nouns are not identical (cf. [22] and [23]).

(22) **Nach dem Trainingslager ist vor dem UI-Cup**, und daher flog der Hertha-Troß vom Flughafen Wien direkt in die russische Hauptstadt, wo die Berliner heute im UI-Cup-Rückspiel auf FK Moskau treffen.
(Berliner Morgenpost, July 22, 2006)
[Engl. 'After the training camp is before the UI Cup']

(23) **Nach der Fußball-WM ist vor der Fecht-EM**: Die Vorbereitungen auf das internationale Sport-Event in Leipzig laufen auf Hochtouren.
(Leipziger-Volkszeitung, July 1, 2010)
[Engl. 'After the World Cup is before the European Fencing Championships']

The variant [*Nach* X *ist vor* Y] increases over time as we look more closely at the hapax legomenons (cf. Table 10). Its proportion among all hapax legomenons develops from 21.6% (1995–1999) to 63.8% (2020–2022).

Table 10: Two different nouns as slot fillers according to the pattern [*Nach* X *ist vor* Y] in the German Reference Corpus between 1995 and 2022 (only hapax legomenons are counted).

Sub-corpus / time period	1995–1999	2000–2004	2005–2009	2010–2014	2015–2019	2020–2022
Hapax legomenons	51	475	1,088	2,016	2,813	1,366
Nach X *ist vor* Y (in total and in percentage)	11 (21.6%)	167 (35.2%)	528 (48.5%)	1,164 (57.7%)	1,736 (61.7%)	872 (63.8%)

It must be mentioned that in several instances the lexemes *Spiel* or *Wahl* are used as one of the two slot fillers (cf. [24] and [25]).

(24) **Nach dem Jubel ist vor dem Spiel**: Noch sind bei den HCL-Handballerinnen die Glückshormone über das Erreichen des EC-Achtelfinals nicht ganz abgebaut, da wird schon wieder an die Konzentration appelliert.
(Leipziger-Volkszeitung, November 15, 2011)
[Engl. 'After the cheer is before the game']

(25) **Nach der Wahl ist vor der Fastnacht**. Oder war die Wahl schon Fastnacht?
      Das fragten sich die Leser des gestrigen SÜDKURIER.
      (Südkurier, November 30, 2016)
      [Engl. 'After the election is before the carnival']

Comparable to the corpus analysis of the catchphrase *Wohnst du noch oder lebst du schon?* (cf. Section 4.2.1, Table 5), it is therefore interesting to look at those constructs in which neither of the two words occurs. Table 11 shows that the slots are filled less and less often by *Spiel* or *Wahl* during the examined period.

Table 11: Proportion of the nouns *Spiel* and *Wahl* within the constructs of the multi-word expression [*Nach* X *ist vor* X/Y] in the German Reference Corpus between 1995 and 2022.

Sub-corpus / time period	1995–1999	2000–2004	2005–2009	2010–2014	2015–2019	2020–2022
Tokens / constructs in the sub-corpus	**257**	**3,992**	**9,711**	**14,411**	**18,598**	**7,686**
Nach dem Spiel ist vor dem Spiel	104	661	1,496	1,369	1,204	461
Nach dem Spiel ist vor X	2	48	115	156	103	69
Nach X ist vor dem Spiel	1	8	30	17	25	5
Nach der Wahl ist vor der Wahl	39	766	1,472	1,516	1,600	607
Nach der Wahl ist vor X	5	23	67	138	185	139
Nach X ist vor der Wahl	–	18	41	29	25	11
Tokens with the lexemes *Spiel* and/or *Wahl* (in total and in percentage)	**151 (58.8%)**	**1,524 (38.2%)**	**3,221 (33.2%)**	**3,225 (22.4%)**	**3,142 (16.9%)**	**1,292 (16.8%)**

The corpus analysis shows a development from a (more) lexicalized multi-word expression to a (more) semi-schematic construction. This is indicated, for instance, by a constant decrease in the frequency of the word *Spiel* (and the word *Wahl*) and the constant increase in productivity. Like the first example, a semantic change is taking place. The formulaic expression referring to Herberger's quote and thus to a specific football context is used less and less often compared to other realizations of the construction. The meaning of the construction thus becomes more abstract. The expansion of the (abstract) meaning of the construction can also be seen in the formal development. Thus, those instances in which two different lexemes (*Nach* X *ist vor* Y), each with its own meaning, occur in the slots increase over time.

4.3.2 Online Survey

The online survey on knowledge of the source of the formulaic expression *Nach dem Spiel ist vor dem Spiel* provides interesting results. Out of 263 participants, 96 (36.5%) say that they know where the multi-word expression comes from. 150 (57%) people are familiar with the idiom, but do not know where it comes from. And 17 (6.5%) participants said they do not know the construction. Thus, far fewer participants indicated a source than in the case of the idiom *Wohnst du noch oder lebst du schon?* (cf. Section 4.2.2). Table 12 presents the correlations between the expression *Nach dem Spiel ist vor dem Spiel* according to the respondents' indication in the correlations with the other items.

Table 12: Simple correlations between the knowledge of the multi-word expression *Nach dem Spiel ist vor dem Spiel* (not recoded by the researcher) and the other variables.[31]

Knowledge of the expression *Nach dem Spiel ist vor dem Spiel*	Correlation
Media consumption television	.15*
Media consumption newspaper	.23***
Media consumption social media	−.24***
Media consumption radio	.18**
Media consumption magazines	.15*
Interest in movies	.01

---

31 *p < .05; **p < .01; ***p < .001. Variables were assumed to be interval scaled. Accordingly, Pearson's r was used.

Table 12 (continued)

Knowledge of the expression *Nach dem Spiel ist vor dem Spiel*	Correlation
Interest in literature	.15*
Interest in music	−.01
Interest in social media	−.22***
Interest in politics	.25***
Interest in football (soccer)	.18**
Interest in advertising	.09
Interest in fashion	−.05
Interest in technology	.10
Interest in religion	.13*
Age	.34***
Education	−.12

The analysis shows some sociodemographic factors that influence whether people (think they) know where the expression comes from. The higher the consumption of television, newspaper, radio, and magazines and the lower the consumption of social media, as well as the more interest in literature, politics, football, and religion and the less interest in social media, the more pronounced is the knowledge of the source of the expression. Of course, not all statistically significant results can be reasonably explained. We therefore highlight only two of them: interest in football influences knowledge of the catchphrase. This is quite plausible since the expression originates from a quote by a German football coach. In addition, age plays a crucial role. The older the participants are, the more likely they are to believe they know the origin of the expression. This is also plausible, as Sepp Herberger was active between the 1930s and 1960s, winning the 1954 World Cup as Germany's national coach.

However, our evaluation of the answers shows that the given answers of the participants do not always completely correspond to the actual source of the multi-word expression. For instance, some participants do not name Sepp Herberger or remain extremely vague. We recoded answers as quite vague, but partially correct, when it was mentioned that the expression comes from football or from a football coach. We also considered answers as partially correct when people named another German (national) coach. The knowledge that the expression comes from a football coach is therefore present here. We recoded answers as very vague when only *Sport* [Engl. 'sports'] was mentioned as a source. Table 13 presents the correla-

tions between the multi-word expression *Nach dem Spiel ist vor dem Spiel* after the researchers' recoding of the correlations with the other items.[32]

Table 13: Simple correlations between the knowledge of the multi-word expression *Nach dem Spiel ist vor dem Spiel* (recoded by the researcher) and the other variables.[33]

Knowledge of the expression *Nach dem Spiel ist vor dem Spiel* (recoded by the researcher)	Correlation
Media consumption television	.29**
Media consumption newspaper	.27**
Media consumption social media	−.01
Media consumption radio	.12
Media consumption magazines	.11
Interest in movies	.17
Interest in literature	−.12
Interest in music	−.03
Interest in social media	.07
Interest in politics	.21*
Interest in football (soccer)	.36***
Interest in advertising	.20
Interest in fashion	−.06
Interest in technology	.25*
Interest in religion	−.03
Age	.23*
Education	−.16

Table 13 is particularly interesting as it illustrates that interest in football is indeed a very important factor. Among the people who think they know where the expression comes from, the ones who give a very precise answer are those who show a higher interest in football than the others. In other words, the more interested people are in football, the more likely they are to name the formulaic expression as a saying from Sepp Herberger. This example suggests that the origins of winged words are particularly familiar to people who are interested in the domain from which the winged word originates (politics, literature, films, football, etc.). However, this still needs to be proven on a broader empirical basis.

---

32 It must be emphasized that in this analysis the number of cases is significantly lower, since only 96 participants stated that they knew where the phrase comes from. 36 determined the source completely correctly, 57 answered partially correctly, and 3 answered very vaguely.

33 *p < .05; **p < .01; ***p < .001. Variables were assumed to be interval scaled. Accordingly, Pearson's r was used.

It must be emphasized (as already seen in the case of the expression *Wohnst du noch oder lebst du schon?*, Section 4.2.2, Table 7) that the answers are quite diverse. Table 14 gives an insight into different answers.

Table 14: Selected answers about the origin of the multi-word expression *Nach dem Spiel ist vor dem Spiel.*

Accuracy of the answer	Answer	Sociodemographic information about the participant (looking at age and other statistically significant parameters)
Precise indication of the original context	Sepp Herberger, Dt. Fußball-Nationaltrainer, um 1954 [Engl. 'Sepp Herberger, German national football coach, around 1954']	Interview number: 514 Age: 40–49 Media consumption television: several times a week Media consumption newspaper: two or three times a month Interest in football: fully agree
	Zitat von Sepp Herberger nach Fußballspiel [Engl. 'Quote from Sepp Herberger after football match']	Interview number: 295 Age: 60–69 Media consumption television: (almost) daily Media consumption newspaper: several times a day Interest in football: fully agree
Rough or partly correct indication of the original context	Von einem Fußballtrainer [Engl. 'From a football coach']	Interview number: 397 Age: 11–29 Media consumption television: several times a day Media consumption newspaper: two or three times a month Interest in football: do not agree at all
	Zitat Berti Vogts nach Niederlage [Engl. 'Quote from Berti Vogts after defeat']	Interview number: 160 Age: 50–59 Media consumption television: (almost) daily Media consumption newspaper: (almost) daily Interest in football: disagree

Table 14 (continued)

Accuracy of the answer	Answer	Sociodemographic information about the participant (looking at age and other statistically significant parameters)
Very imprecise indication of the source	Sagt man in jedem Sport. [Engl. 'It's said in every sport']	Interview number: 517 Age: 40–49 Media consumption television: several times a week Media consumption newspaper: two or three times a month Interest in football: agree
	Sport [Engl. 'sports']	Interview number: 484[34] Age: 40–49 Media consumption television: about once a week Media consumption newspaper: once a month or less often Interest in football: do not agree at all

Table 14 shows that some people, in the sense of the EC-model, have memorized the expression more strongly in connection with its original context. In some cases, they can even specify when Sepp Herberger was the national coach (cf. interview number 514). Thus, the fully lexicalized expression is presumably a node of its own in the individual's constructicon. Some people can only name the football context or associate the expression with other (famous) German coaches (cf. interview numbers 397, 160).[35] Two participants indicate only the rough context 'sports' (cf. interview numbers 517, 484). In addition, many participants are familiar with the expression but do not know where it comes from. And some people do not know the fully lexicalized unit at all. This means that they do not have an entry for this micro-construction in their constructicon. However, there is the possibility that they have stored the partly lexicalized pattern [*Nach* X *ist vor* X/Y] as a meso-construction (Traugott 2008: 236), which a participant explicitly refers to (cf. [26]).

---

34 The person stated not to be a German but an Austrian native speaker. Where a speaker comes from certainly also plays a role in knowing the source of catchphrases. Austrians and Swiss, for instance, are generally less familiar with German football coaches, politicians, actors, commercials etc. than Germans (and vice versa).
35 The following coaches were named: Franz Beckenbauer (4x), Berti Vogts (3x), Joachim Löw (3x), Lothar Matthäus (2x), Jürgen Klopp (2x) and Erich Ribbeck (1x).

(26)   kenne ich so nicht, aber in der allgemeinen Form "nach dem/der X ist vor
       dem/der X", z.B. "nach dem Konzert ist vor dem Konzert"
       (Response of a participant regarding the multi-word expression *Nach dem
       Spiel ist vor dem Spiel*, Interview number 140)
       [Engl. 'I don't know it like that, but in the general form "after X is before X",
       e.g., "after the concert is before the concert"']

Language knowledge thus varies from speaker to speaker (Dąbrowska 2015), and further research on interindividual variation in lexico-grammatical patterns is needed.

5 Conclusion and Discussion

Following the EC-model and the theory of constructionalization, this paper assumes that "[n]ew patterns come to be entrenched not only in individual minds ("innovations") but come to be shared and entrenched within a community of speakers ("changes")" (Traugott 2019: 129). In contrast to previous studies on (the dynamics of) constructional idioms, this paper, therefore, considers conventionalization (sociopragmatic perspective) and entrenchment (cognitive perspective) by combining corpus analyses and online surveys.

In sum, we can conclude that creative variation defined as modification in phraseology can be a driving force of changes in the form and meaning of idioms. However, it should be emphasized, that the boundary between idiom modification and semi-schematic idioms is fuzzy. It is not possible to determine exactly for every phraseme which is not used in its conventionalized form whether empirically attested variations are creative modifications of a fully lexicalized idiom or rather already fillers of an underlying semi-schematic construction with slots.

In the following, we discuss the interaction between creativity and routinization in the dynamics and emergence of constructional idioms and provide a more general modeling of the development process (cf. Figure 3). For illustration we use the analyzed multi-word expressions *Wohnst du noch oder lebst du schon?* (cf. Chapter 4.2) and *Nach dem Spiel ist vor dem Spiel* (cf. Chapter 4.3).

The starting point of the development is a slogan created by the furniture company IKEA and a quote from Sepp Herberger (cf. step 1). To be precise, the slogan and the quote are not yet phrasemes (not every slogan and quote develop into phraseological expressions). A routinization takes place in the next step (cf. step 2). The expressions are frequently used in everyday language outside the original contexts of IKEA advertising and football. In this way, formulaic expressions emerge which can be classified as catchphrases (winged words). In the next step the formulaic ex-

	Step 1	Step 2	Step 3	Step 4
Expression	*Wohnst du noch oder lebst du schon?*	*Wohnst du noch oder lebst du schon?*	*Mietest/liest* etc. *du noch oder besitzt/kindlest* etc. *du schon?*	*X du noch oder* Y *du schon?*
	"Nach dem Spiel ist vor dem Spiel."	*Nach dem Spiel ist vor dem Spiel.*	*Nach dem Rave/der WM etc. ist vor dem Rave/der EM*	*Nach X ist vor X/Y*
Classification	Slogan (IKEA) / quote (Sepp Herberger)	Catchphrase (winged word)	Frequent modification through substitution	Constructional idiom / pattern-like catchphrase (snowclone)
	Creativity	Routinization	Creativity	Routinization

Figure 3: Interaction between creativity and routinization in the emergence of constructional idioms.

pressions are frequently modified (cf. step 3). It must be emphasized that the modification can already happen or begin during the routinization process. In general, the four steps are not clearly separable from each other and can happen simultaneously. The modification through substitution finally leads to the effect that the modified forms occur much more frequently than the original multi-word expressions. Thus, in the last step, constructional idioms emerge (cf. step 4).

We can therefore assume a "diachronic path from a fully lexicalized phraseme via (occasional) phraseme modification to a full-fledged phraseological schema" (Stutz and Finkbeiner 2022: 289). This diachronic path can be described as lexical constructionalization, since formal and semantic change takes place and a new semischematic pattern with open slots emerges (Traugott and Trousdale 2014: 273). The two case studies in this paper also provide evidence for Buerki's (2019: 30) assumptions that formulaic language "changes at a rapid pace". The phenomenon studied here does not involve processes of change that take place over centuries, but rather changes in the form and meaning of constructions that become conventionalized and entrenched in a short period of time because of creative utterances by speakers.

The constructionalization is characterized by an interaction between creativity and routine. The creation of a slogan by an advertising company and the sentence by Sepp Herberger, which he probably said in an interview, can be classified as creative (cf. step 1 in Figure 3). The development of catchphrases can be described as routinization (cf. step 2). The occasional modifications of these fixed expressions are again creative according to Langlotz (2006: 9) (cf. step 3). Often these modifications are word plays, i.e., they are intentional and funny variations of an idiom. Finally, it is precisely this creative modification that again leads to

the emergence of a fixed, but now semi-schematic construction (cf. step 4). It must be mentioned that the steps cannot be clearly distinguished from one another. Routinization and creativity overlap and can happen simultaneously in the development of constructional idioms.

Following Hoffmann (2022), modification in step 3 can be considered a typical case of E-creativity. However, the more frequently a multi-word expression is modified, the more likely it is that a (productive) pattern will emerge that can be categorized as F-creativity. Thus, the development of lexico-grammatical patterns through repeated modification can be described as a gradual transition from E-creativity (i.e., modification, breaking "rules") to F-creativity (forming a partly schematic construction) (Ungerer and Hartmann 2023: 44).

The emergence of constructional idioms through modification can also be explained by Barðdal's (2008) productivity cline (Ungerer and Hartmann 2023: 45). Initial changes through creative modification represent analogical processes. Through frequent modification, the multi-word expression becomes more and more productive. At a certain point in this process, there is a gray area between creative analogy and productive pattern. If the modifications exceed the original expression in the speech community to a considerable extent (conventionalization), the connection to its source in the linguistic knowledge of the individual speakers may be lost (entrenchment). In this way, semi-schematic constructions can emerge that exist completely independently of their source.

It is evident that for the change of (more lexicalized) idioms to (more schematic) semi-schematic patterns we should refer to phraseology and Construction Grammar. This is because fully lexicalized catchphrases (winged words) are initially situated at the lexicon pole of the lexicon-syntax continuum, which is the core area of phraseology. However, over time they develop more and more in the direction of the syntax pole due to schematization, which Construction Grammar is most interested in. Moreover, phraseological research on the phenomenon of modification can be useful for constructionist research on creativity. So far, however, this connection has not been made (Filatkina 2018a: 27–28). The paper thus contributes to a stronger interaction between phraseological and constructionist approaches.

The constructional idioms analyzed in this study have similarities with snowclones following Hartmann and Ungerer (2024) as they originate from a source. However, it can be questioned if snowclones are really a type of their own within constructional idioms. Instead, we are dealing with a gradual transition or a continuum. The transition ranges from snowclones (speakers know their source) to "typical" constructional idioms, which are not characterized by a source construction. Thus, the online survey of this paper shows that some speakers either cannot name the source or can name it only very imprecise. In other words, speakers are often not even aware of whether a source, and if so what source, underlies a formulaic

expression. This means that knowledge of the source of a catchphrase can be lost in the speech community. The defining criterion that snowclones are based on sources is thus only a feature assumed by linguists. Whether speakers have this knowledge is a completely different question, which can only be determined to a certain extent by corpus studies, but rather by surveys or experimental studies.

References

Bailey, Guy, Tom Wikle, Jan Tillery & Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3. 241–264.

Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.). 2015. *Diachronic Construction Grammar*. Amsterdam & Philadelphia: Benjamins.

Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam & Philadelphia: Benjamins.

Barz, Irmhild. 1992. Phraseologische Varianten. Begriff und Probleme. In Csaba Földes (ed.), *Deutsche Phraseologie in Sprachsystem und Sprachverwendung*, 25–47. Wien: Praesens.

Bebermeyer, Gustav & Renate Bebermeyer. 1977. Abgewandelte Formeln – sprachlicher Ausdruck unserer Zeit. *Muttersprache* 87. 1–42.

Bergs, Alexander. 2017. Diachronic Approaches. In Barbara Dancygier (ed.), *Cambridge Handbook of Cognitive Linguistics*, 347–360. Cambridge: Cambridge University Press.

Bergs, Alexander. 2018. Learn the Rules like a Pro, so you can Break them like an Artist (Picasso): Linguistic Aberrancy from a Constructional Perspective. *Zeitschrift für Anglistik und Amerikanistik* 66. 277–293.

Bergs, Alexander. 2019. What, If Anything, Is Linguistic Creativity? *Gestalt Theory* 41. 173–184.

Bergs, Alexander & Gabriele Diewald (eds.). 2008. *Constructions and Language Change*. Berlin & New York: De Gruyter.

Booij, Geert. 2002. Constructional idioms, morphology, and the Dutch lexicon. *Journal of Germanic Linguistics* 14. 301–329.

Buerki, Andreas. 2019. Furiously fast: On the speed of change in formulaic language. *Yearbook of Phraseology* 10. 5–38.

Burger, Harald. 2015. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 5., neu bearbeitete Auflage. Berlin: Schmidt.

Colson, Jean-Pierre. 2021. Phraseology and Cognitive Entrenchment: Corpus-based Evidence and Applications for Language Teaching and Translation. In Joanna Szerszunowicz (ed.), *Reproducibility of Multiword Expressions in Paremiological and Linguo-cultural Studies*, 15–30. Bialystok: University of Bialystok.

Corpas Pastor, Gloria. 2021. Constructional idioms of 'insanity' in English and Spanish: A corpus-based study. *Lingua* 254. 1–20.

Corpas Pastor, Gloria. 2022. *You are driving me up the wall!* A corpus-based study of a special class of resultative constructions. *Lexis. Journal in English Lexicology* 19. 1–35.

Coseriu, Eugenio. 1975. *Sprachtheorie und allgemeine Sprachwissenschaft. 5 Studien*. München: Fink.

Coseriu, Eugenio. 2007. *Sprachkompetenz. Grundzüge der Theorie des Sprechens*. 2., durchgesehene Auflage. Tübingen: Narr.

Croft, William & Alan D. Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Cukor-Avila, Patricia & Guy Bailey. 2013. Real Time and Apparent Time. In J. K. Chambers & Natalie Schilling (eds.), *The Handbook of Language Variation and Change*. Second Edition, 237–262. Oxford: Wiley-Blackwell.

Dąbrowska, Ewa. 2015. Individual differences in grammatical knowledge. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 650–668. Berlin & Boston: De Gruyter.

De Smet, Hendrik. 2013. *Spreading patterns: Diffusional change in the English system of complementation*. Oxford: Oxford University Press.

Diewald, Gabriele & Elena Smirnova. 2010. *Evidentiality in German: Linguistic realization and regularities in grammaticalization*. Berlin & New York: De Gruyter.

Dobrovol'skij, Dmitrij. 1999. Zu semantischen und pragmatischen Effekten kreativer Idiom-Modifikationen. *Nouveaux Cahiers d'Allemand* 17. 363–374.

Dobrovol'skij, Dmitrij. 2008. Idiom-Modifikationen aus kognitiver Perspektive. In Heidrun Kämper & Ludwig M. Eichinger (eds.), *Sprache – Kognition – Kultur. Sprache zwischen mentaler Struktur und kultureller Prägung*, 302–322. Berlin & New York: De Gruyter.

Dobrovol'skij, Dmitrij. 2011a. Phraseologie und Konstruktionsgrammatik. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, 111–130. Tübingen: Stauffenburg.

Dobrovol'skij, Dmitrij 2011b. Zur Typologie der Idiom-Modifikationen. In Galina I. Fadeeva, Innara A. Gusejnova & Elena I. Karpenko (eds.), *Aktuelle Probleme der modernen Lexikologie und Phraseologie. Festschrift für Professor I. I. Černyševa zum 100. Geburtstag*, 57–88. Moskau: Rema.

Duden. 2021. *Zitate und Aussprüche. Herkunft, Bedeutung und aktueller Gebrauch*. Berlin: Dudenverlag.

Fauconnier, Gilles & Mark Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Fellbaum, Christiane. 2019. How flexible are idioms? A corpus-based study. *Linguistics* 57. 735–767.

Fellbaum, Christiane & Ekaterini Stathi. 2006. Idiome in der Grammatik und im Kontext: Wer brüllt hier die Leviten? In Kristel Proost & Edeltraud Winkler (eds.), *Von Intentionalität zur Bedeutung konventionalisierter Zeichen. Festschrift für Gisela Harras zum 65. Geburtstag*, 125–146. Tübingen: Narr.

Fiedler, Sabine. 2007. *English Phraseology: A Coursebook*. Tübingen: Narr.

Filatkina, Natalia. 2018a. Expanding the Lexicon through Formulaic Patterns: The Emergence of Formulaicity in Language History and Modern Language Use. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin & Esme Winter-Froemel (eds.), *Expanding the Lexicon. Linguistic Innovation, Morphological Productivity, and Ludicity*, 15–42. Berlin & Boston: De Gruyter.

Filatkina, Natalia. 2018b. Historische formelhafte Wendungen als Konstruktionen: Möglichkeiten und Grenzen der diachronen Konstruktionsgrammatik. *Linguistik online* 90 (3), 115–143.

Fillmore, Charles, Paul Kay & Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions. The Case of *let alone*. *Language* 64. 501–538.

Fischer, Olga. 2007. *Morphosyntactic change: Functional and formal perspectives*. Oxford: Oxford University Press.

Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache.* 2. durchgesehene und ergänzte Auflage. Tübingen: Niemeyer.

Fried, Mirjam. 2009. Construction Grammar as a Tool for Diachronic Analysis. *Constructions and Frames* 1. 261–291.

Fried, Mirjam. 2013. Principles of Constructional Change. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 419–437. Oxford: Oxford University Press.

Gläser, Rosemarie. 2001. The Stylistic Potential of Phraseological Units in the Light of Genre Analysis. In Anthony Paul Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*, 125–144. Oxford: Oxford University Press.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2019. *Explain me this. Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.

Goldberg, Adele E. & Thomas Herbst. 2021. The *nice-of-you* construction and its fragments. *Linguistics* 59. 285–318.

Hartmann, Stefan & Tobias Ungerer. 2024. Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns. *Journal of Linguistics* 60. 599–634.

Herbst, Thomas. 2018. Collo-Creativity and Blending: Recognizing Creativity Requires Lexical Storage in Constructional Slots. *Zeitschrift für Anglistik und Amerikanistik* 66. 309–326.

Hilpert, Martin. 2008. *Germanic Future Constructions: A Usage-Based Approach to Language Change*. Amsterdam & Philadelphia: Benjamins.

Hilpert, Martin. 2011. Was ist Konstruktionswandel? In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, 59–75. Tübingen: Stauffenburg.

Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press.

Hilpert, Martin, Bert Cappelle & Ilse Depraetere (eds.). 2021. *Modality and Diachronic Construction Grammar*. Amsterdam & Philadelphia: Benjamins.

Hoffmann, Thomas. 2018a. Creativity and construction grammar: Cognitive and psychological issues. *Zeitschrift für Anglistik und Amerikanistik* 66. 259–276.

Hoffmann, Thomas. 2019. Language and creativity: A construction grammar approach to linguistic creativity. *Linguistics Vanguard* 5. 1–8.

Hoffmann, Thomas. 2022. Constructionist approaches to creativity. *Yearbook of the German Cognitive Linguistics Association* 10. 259–284.

Hoffmann, Thomas (ed.). 2018b. *Linguistic Creativity*. Special Issue of *Zeitschrift für Anglistik und Amerikanistik* 66 (3).

Hoffmann, Thomas (ed.). 2020. *Construction Grammar and Creativity*. Special Issue of *Cognitive Semiotics* 13 (1).

Hoffmann, Thomas & Graeme Trousdale (eds.). 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Huber, Judith & Thomas Herbst (eds.). 2022. *Diachronic Construction Grammar*. Special Issue of *Zeitschrift für Anglistik und Amerikanistik* 70 (3).

Hunston, Susan & Gill Francis. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam & Philadelphia: Benjamins.

Ivorra Ordines, Pedro. 2022. Comparative constructional idioms. A corpus-based study of the creativity of the [*más feo que* X] construction. In Carmen Mellado Blanco (ed.), *Productive Patterns in Phraseology and Construction Grammar. A Multilingual Approach*, 29–52. Berlin & Boston: De Gruyter.

Ivorra Ordines, Pedro. In press. *Dime con quién te juntas y te diré quién eres*. Substantive idioms from Construction Grammar. In Torsten Leuschner, Anaïs Vajnovszki, Gauthier Delaby & Jóhanna Barðdal (eds.), *How to Do Things with Corpora – Methodological Issues and Case Studies*. Berlin: Springer Nature.

Jaki, Sylvia. 2014. *Phraseological Substitutions in Newspaper Headlines. "More than Meats the Eye"*. Amsterdam & Philadelphia: Benjamins.

Kabatek, Johannes. 2023. *Eugenio Coseriu. Beyond Structuralism*. Berlin & Boston: De Gruyter.

Klein, Josef. 2013. Der saliente politische Satz – ein Kristallisationspunkt kollektiven Wissens. In Jörg Kilian & Thomas Niehr (eds.), *Politik als sprachlich gebundenes Wissen. Politische Sprache im lebenslangen Lernen und politischen Handeln*, 137–158. Bremen: Hempen.

Labov, William. 1963. The social motivation of a sound change. *Word* 19. 273–309.

Langacker, Ronald W. 2008. *Cognitive Grammar – A Basic Introduction*. Oxford: Oxford University Press.

Langlotz, Andreas. 2006. *Idiomatic Creativity. A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam & Philadelphia: Benjamins.

Lasch, Alexander & Alexander Ziem (eds.). (2023). *Konstruktionsgrammatik VII. Wandel im Sprachgebrauch*. Tübingen: Stauffenburg.

Mantlik, Annette & Hans-Jörg Schmid. 2018. *That*-complementizer omission in N+BE+*that*-clauses – register variation or constructional change? In Alex Ho-Cheong Leung & Wim van der Wurff (eds.), *The noun phrase in English: past and present*, 187–222. Amsterdam & Philadelphia: Benjamins.

Martí Solano, Ramon. 2013. From Idiom Variants to Open-Slot Idioms: Close-Ended and Open-Ended Variational Paradigms. In Joanna Szerszunowicz, Bogusław Nowowiejski, Takaaki Kanzaki & Katsumasa Yagi (eds.), *Research on Phraseology across Continents*. Volume II, 167–180. Białystok: University of Białystok Publishing House.

Mellado Blanco, Carmen. 2018. Wenn modifizierte Sprichwörter zu Mustern werden. Eine korpusbasierte Studie am Beispiel von *Reden ist Silber, Schweigen ist Gold*. In Martina Nicklaus, Nora Wirtz, Wiebke Langer, Marcella Costa & Karin Ewert-Kling (eds.), *Lexeme, Phraseme, Konstruktionen: Aktuelle Beiträge zur Lexikologie und Phraseologie. Festschrift für Elmar Schafroth*, 183–203. Frankfurt am Main: Lang.

Mellado Blanco, Carmen. 2020. (No) me importa un comino y sus variantes diatópicas: estudio de corpus desde la Gramática de Construcciones. *Estudios de Lingüística. Universidad de Alicante (ELUA)*, VII, 87–109.

Mellado Blanco, Carmen. 2022. Phraseology, patterns and Construction Grammar. In Carmen Mellado Blanco (ed.), *Productive Patterns in Phraseology and Construction Grammar. A Multilingual Approach*, 1–25. Berlin & Boston: De Gruyter.

Mellado Blanco, Carmen. 2023. From idioms to semi-schematic constructions and vice versa. The case of [*a un paso de* X]. In Inga Hennecke & Evelyn Wiesinger (eds.), *Constructions in Spanish*, 103–128. Amsterdam & Philadelphia: Benjamins.

Mellado Blanco, Carmen. 2024. The way of biblical quotations are mysterious: Change of register and snowclones in Spanish and German from a constructivist point of view. In Saša Babič, Fionnuala Carson Williams, Christian Grandl & Anna T. Litovkina (eds.), *"Standing on the shoulders of giants". A Festschrift in honour of Wolfgang Mieder on the occasion of his 80th birthday*, 495–512. Universität Osijek: Faculty of Humanities and Social Sciences.

Mellado Blanco, Carmen (ed.). 2022. *Productive Patterns in Phraseology and Construction Grammar. A Multilingual Approach*. Berlin & Boston: De Gruyter.

Mellado Blanco, Carmen, Fabio Mollica & Elmar Schafroth (eds.). 2022. *Konstruktionen zwischen Lexikon und Grammatik. Phrasem-Konstruktionen monolingual, bilingual und multilingual*. Berlin & Boston: De Gruyter.

Merten, Marie-Luis. 2018. *Literater Sprachausbau kognitiv-funktional. Funktionswort-Konstruktionen in der historischen Rechtsschriftlichkeit*. Berlin & Boston: De Gruyter.

Mollica, Fabio. 2020. Die Phrasem-Konstruktion [X$_{NPnom}$ *sein*$_{Kopula}$ *mir*$_{Exp}$ (Modalpartikel) Det$_{ein(e)}$ Y$_{NP}$]$_{Exkl}$ und ihre Relationen innerhalb der Ethicus-Konstruktion und der Dativ-Familie. *Linguistische Berichte* 261. 47–83.

Noël, Dirk. 2007. Diachronic Construction Grammar and Grammaticalization Theory. *Functions of Language* 14. 177–202.

Noël, Dirk & Timothy Colleman. 2021. Diachronic Construction Grammar. In Wen Xu & John R. Taylor (eds.), *The Routledge Handbook of Cognitive Linguistics*, 662–675. New York & London: Routledge.

Parkinson, Judy. 2003. *Catchphrase, Slogan and Cliché: the origins and meanings of our favourite expressions*. London: Michael O'Mara.

Pfeiffer, Christian. 2016. *Frequenz und Funktionen phraseologischer Wendungen in meinungsbetonten Pressetexten (1911–2011)*. Baltmannsweiler: Schneider.

Pfeiffer, Christian. 2017. Okkasionalität — Zur Operationalisierung eines zentralen definitorischen Merkmals phraseologischer Modifikationen. *Yearbook of Phraseology* 8. 9–30.

Ptashnyk, Stefaniya. 2009. *Phraseologische Modifikationen und ihre Funktionen im Text. Eine Studie am Beispiel der deutschsprachigen Presse*. Baltmannsweiler: Schneider.

Rodríguez Martín, Gustavo A. 2014. Canonical modified phraseological units: Analysis of the paradox. *Yearbook of Phraseology* 5. 3–24.

Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, experiment and argument in linguistics and philosophy of language*, 15–26. Bern: Lang.

Schmid, Hans-Jörg. 2000. *English Abstract Nouns as Conceptual Shells. From corpus to cognition*. Berlin & New York: De Gruyter.

Schmid, Hans-Jörg. 2014. Lexico-grammatical patterns, pragmatic associations and discourse frequency. In Thomas Herbst, Hans-Jörg Schmid & Susen Faulhaber (eds.), *Constructions – collocations – patterns*, 239–293. Berlin & Boston: De Gruyter.

Schmid, Hans-Jörg. 2015. A blueprint of the Entrenchment-and-Conventionalization Model. *Yearbook of the German Cognitive Linguistics Association* 3. 1–27.

Schmid, Hans-Jörg. 2016. Why Cognitive Linguistics must embrace the pragmatic and social dimensions of language and how it could do so more seriously. *Cognitive Linguistics* 27. 543–557.

Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System. Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press.

Schmid, Hans-Jörg. 2021. How the entrenchment-and-conventionalization model might enrich diachronic construction grammar. The case of *(the) thing is (that)*. *Belgian Journal in Linguistics* 34. 312–326.

Schmid, Hans-Jörg & Annette Mantlik. 2015. Entrenchment in historical corpora? Reconstructing dead authors' minds from their usage profiles. *Anglia* 133. 583–623.

Schmid, Hans-Jörg, Quirin Würschinger, Melanie Keller & Ursula Lenker. 2020. Battling for semantic territory across social networks. The case of *Anglo-Saxon* on Twitter. *Yearbook of the German Cognitive Linguistics Association* 8. 3–26.

Schmid, Hans-Jörg, Quirin Würschinger, Sebastian Fischer & Helmut Küchenhoff. 2021. *That's cool*. Computational sociolinguistic methods for investigating individual lexico-grammatical variation. *Frontiers in Artificial Intelligence* 3. 1–16.

Sommerer, Lotte & Elena Smirnova (eds.). 2022. *Nodes and Networks in Diachronic Construction Grammar*. Amsterdam & Philadelphia: Benjamins.

Staffeldt, Sven. 2018. *Gebrauchssemantik von* Hand. *Korpusbasierte Studien zu somatischen Phraseologismen des Deutschen mit der Konstituente Hand*. Tübingen: Stauffenburg.

Stumpf, Sören. 2016. Modifikation oder Modellbildung? Das ist hier die Frage – Abgrenzungsschwierigkeiten zwischen modifizierten und modellartigen Phrasemen am Beispiel formelhafter (Ir-)Regularitäten. *Linguistische Berichte* 247. 317–342.

Stumpf, Sören. 2021. „Star Wars ist eben mehr so ein Männerding". Eine Korpusuntersuchung zur Syntax, Semantik und Pragmatik der Konstruktion [X Kopula *ein* Y-*Ding*]. *Deutsche Sprache* 49. 24–45.

Stutz, Lena & Rita Finkbeiner. 2022. *Veni, vidi, veggie*. A contrastive corpus linguistic analysis of the phraseological construction *Veni, vidi*, X and its German equivalent X *kam, sah und* Y. In Carmen Mellado Blanco (ed.), *Productive Patterns in Phraseology and Construction Grammar. A Multilingual Approach*, 261–287. Berlin & Boston: De Gruyter.

Taylor, John R. 2002. *Cognitive Grammar*. Oxford: Oxford University Press.

Traugott, Elizabeth Closs. 2003. Constructions in Grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, 624–647. Oxford: Blackwell.

Traugott, Elizabeth Closs. 2007. The concepts of constructional mismatch and type-shifting from the perspective of grammaticalization. *Cognitive Linguistics* 18. 523–557.

Traugott, Elizabeth Closs. 2008. Grammaticalization, constructions and the incremental development of language: suggestions for the development of degree modifiers in English. In Regine Eckhardt, Gerhard Jäger & Tonjes Veenstra (eds.), *Variation, selection, development: probing the evolutionary model of language change*, 219–250. Berlin & New York: De Gruyter.

Traugott, Elizabeth Closs. 2015. Toward a coherent account of grammatical constructionalization. In Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.), *Diachronic Construction Grammar*, 51–79. Amsterdam & Philadelphia: Benjamins.

Traugott, Elizabeth Closs. 2019. Constructional pattern-development in language change. In Beatrix Busse & Ruth Möhlig-Falke (eds.), *Patterns in Language and Linguistics*, 125–155. Berlin & Boston: De Gruyter.

Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Traugott, Elizabeth Closs & Graeme Trousdale. 2014. Contentful constructionalization. *Journal of Historical Linguistics* 4. 256–283.

Trousdale, Graeme. 2008. Constructions in grammaticalization and lexicalization: evidence from the history of a composite predicate in English. In Graeme Trousdale & Nikolas Gisborne (eds.), *Constructional approaches to English grammar*, 33–67. Berlin & New York: De Gruyter.

Trousdale, Graeme. 2010. Issues in constructional approaches to grammaticalization. In Katerina Stathi, Elke Gehweiler & Ekkehard König (eds.), *Grammaticalization: current views and issues*, 51–71. Amsterdam & Philadelphia: Benjamins.

Trousdale, Graeme. 2012. Grammaticalization, lexicalization and constructionalization from a cognitive-pragmatic perspective. In Hans-Jörg Schmid (ed.), *Cognitive Pragmatics*, 533–558. Berlin & New York: De Gruyter.

Ungerer, Tobias & Stefan Hartmann. 2023. *Constructionist approaches: Past, present, future*. Cambridge: Cambridge University Press.

Van Lancker Sidtis, Diana, Krista Cameron, Kelly Bridges & John J. Sidtis. 2015. The formulaic schema in the minds of two generations of native speakers. *Ampersand. An Interdisciplinary Journal of General and Applied Linguistics* 2. 39–48.

Zeschel, Arne. 2012. *Incipient productivity: a construction-based approach to linguistic creativity*. Berlin & Boston: De Gruyter.

Monica Vasileanu & Anabella-Gloria Niculescu-Gorpin

Romanian Libfixes in the Making

Abstract: The present study examines whether or not Romanian libfixing is a case of structural borrowing. Libfixes, e.g. *-tastic* (< *fantastic*), are defined here as word parts emerging in blends that are analogically used in coining new words, while still preserving the connection with their source-words. This connection is best attested by the metrical properties of libfixed formations that tend to preserve the prosodic contour of the source-word, making libfixes an intermediate stage on the blending-bound morpheme continuum. By analysing 7 Romanian libfixes based on three corpora, we argue that Romanian libfixing is a case of both material and structural borrowing, relying, most likely, on the increasing productivity of lexical blending in Romanian under the influence of English.

1 Introduction

Mainly discussed in relation to English, libfixes such as *-licious* (< *delicious*) or *-tainment* (< *entertainment*) represent an intermediary stage situated somewhere along a continuum between word parts found in blends and bound morphemes. This study is a first analysis of Romanian libfixes that has sprung from the observation that, with the increased use of lexical blends in Romanian, libfixing has also started to be used to coin new words, most probably under the current English influence. Based on a corpus analysis of 7 Romanian libfixes, out of which 6 are of English origin, we aim to establish whether present-day Romanian libfixing is a case of structural borrowing, i.e. the borrowing of an abstract morphological schema (see Section 2.2).

Libfixes illustrate an interesting instance of the complex creativity-productivity interaction. The two are often viewed in opposition, with productivity being often discussed in relation to predictable, rule-governed word-formation, and creativity conceived as non-rule governed (Bauer 2001: 65; Arnaud 2013: 97; Arndt-Lappe 2018, a.o.). Creative means of coining new words that do not concatenate morphemes have sometimes been outcast from the realm of word-formation processes into those of word creation (Haspelmath 2002: 25; Ronneberger-Sibold 2015: 485–487) or extragrammatical morphology (Dressler 2000). Yet, as shown by Norde and Trousdale (2023: 153), creativity and productivity should also be viewed on a continuum, since a creative non-concatenative process, such as blending, may yield productive morphemes in time. As libfixes arise from one-of-a-kind blends and come to form more

and more lexemes, we believe libfixes represent this very moment, i.e. when creativity becomes productive.

In Section 2, we present an overview of libfixing. First, as we aim to define libfixes, we distinguish them from splinters and bound morphemes such as combining forms and affixes. Secondly, we discuss libfixes and contact-induced language changes, and we explore the conditions in which structural borrowing occurs. Section 3 presents our methodology and details of the queried Romanian corpora. Section 4 comprises the seven case-studies and our data interpretation; tentative conclusions are drawn in Section 5.

2 Libfixing in Language Contact Settings: The Case of Romanian

The term libfix was originally coined by Zwicky (2010) in a blog post to name 'liberated' word-forming elements originating in the reanalysis of an older word, e.g. *-tacular* (< *spectacular*) or *-dar* (< *radar*). While belonging to expressive morphology, i.e. the sum of word-formation patterns with more difficult to-pin-down rules and pragmatic effects, as opposed to plain morphology, i.e. rule-governed word-formation (Zwicky and Pullum 1987), libfixing lacks a unanimously accepted definition (see below), most likely due to the evasive and temporary character of libfixes themselves. Moreover, libfixing has been mostly studied in English, so the situation in other languages is yet to be explored. As the current pervasive influence of English has changed most of the world's languages in various ways, certain subtle changes induced into the word-formation systems of the recipient languages have been analysed as forms of structural borrowing (see below 2.2). In this context, our aim is to observe whether the massive English influence on present-day Romanian has gone over borrowing only lexical items, thus also triggering the borrowing of a word-formation process, namely of libfixing.

2.1 What is a Libfix?

Libfixes are definitely older than their name. For example, the OED currently states that *-licious* and *-tastic* were created in the 19th century, but only in 1928 Jespersen (apud Hamans 2017: 11) first referred to the underlying process itself: he observed that sometimes parts of indivisible words were assigned a grammatical significance. Yet, scholars hesitated to classify them, and tried to place them in pre-existing categories. For instance, Marchand (1960: 160–161) includes *-athon* (< *marathon*) and *-eteria*

(< *cafeteria*) in a list of suffixes. As they are bound morphemes with lexical content, Warren (1990) calls *-aholic* and *-gate* combining forms, and places them in the same category with neoclassical combining forms, e.g. *astro-* or *-drome*. The same label is used by Lalić-Krstin (2014) for *-geddon* (< *Armageddon*), although she calls the resulting new words blends. Lehrer (2007) considers *-aholic*, *-gate* and *-athon* good candidates for future combining forms, but calls them splinters, i.e. non-morphemic word parts preserved in blends. Bauer, Lieber and Plag (2013: 525) use the term *splinters* to name "non-morphemic portions of a word that have been split off and used in the formation of new words with a specific new meaning", not necessarily in connection to blending. Moreover, the OED considers some of these items fully established morphemes: *-gate* and *-athon* are combining forms, whereas *-aholic* has already turned into an affix (although the second edition of the OED registered it as a combining form, in the latest updated online version, it is an affix).[1]

Taking on Zwicky's (2010) term, Hamans (2017) and Norde and Sippach (2019) consider libfixes a category *sui-generis*, but each define them differently. According to Hamans (2017), English libfixes are non-morphemic word parts used as morphemes, 'liberated' as a result of reinterpreting (a) partially transparent loanwords, e.g. Dutch *landschap* > En. *landscape* > *-scape*; (b) autochthonous formations, e.g. *Watergate* > *-gate* 'scandal'; (c) blends, e.g. *television + marathon > telethon* > *-thon*; (d) a series of similar forms, e.g. *Pakistan, Afghanistan, Uzbekistan* etc. > *-istan*. Hamans claims that the liberation of libfixes resembles folk etymology: speakers interpret a word part as an existing morpheme and reinterpret the other part as another morpheme. For instance, in *landscape*, speakers recognize *land-* and interpret the source-word as a compound, and thus also considering *-scape* a morpheme; then, it is added to other bases, e.g. *seascape, mindscape, soundscape* (Hamans 2017: 10–19).

However, we believe that Hamans does not provide a clear-cut definition of libfixes that would differentiate them from affixes or combining forms. In the examples (a) and (d) above, libfixes are detached from loanwords and later combined with native bases, which is also the typical scenario for borrowing affixes or combining forms. This is why Marchand (1960) includes *-scape* and *-eteria* among suf-

---

1 Although some researchers do not draw sharp boundaries between combining forms and affixes, the two types of bound morphemes differ on a number of features. Though both categories combine with lexemes, which is typically a property of affixes Bauer (1983: 213), two combining forms alone may form a lexeme, whereas two affixes cannot, as they need a root (Bauer 1983: 214; Plag 2018: 154, a.o.). Moreover, affixes cannot form words by attaching to combining forms (Bauer 1983: 214–215). The distinction is also made on semantic grounds, as combining forms have more lexical density, i.e. a more specific meaning, whereas affixes merely modify the meaning of the base (Bauer 1983: 215; Warren 1990: 123, Bauer, Lieber and Plag 2013: 486, a.o.).

fixes. Moreover, it is pretty common for affixes to spring from the reinterpretation of native formations, therefore cases under (b) are not specific for libfixes either. For instance, the Romanian adjectival suffix -*os* 'abunding in . . .' is usually attached to singular nouns, but also to plural nouns such as *dealuri* 'hills' or *colțuri* 'corners', leading to adjectives such as *colțuros* 'edgy' or *deluros* 'hilly'; from these words, the form -*uros* was detached and then used as an adjective suffix, e.g. in *buburos* 'pimply'. Thus, for us, only (c) – serial blending – is in fact the only situation specific to libfixes. For instance, in the case of -*gate*, we believe that the first formations were blends of *Watergate* and another word, and that only afterwards the splinter was reinterpreted and then combined with other bases.

Norde and Sippach (2019) consider libfixes specific to playful word-formation and analyse two types of libfixes. The first category includes items originating in blends, e.g. -*licious* < *delicious*, -*fection* < *perfection*, and the second one comprises neoclassical combining forms, e.g. -*meter*, -*cracy*. Norde and Sippach's (2019) corpus analysis shows that libfixes have specific morpho-phonological features: libfixes are usually attached to whole words in an affix-like manner, but also to word parts, acronyms or even phrases, sometimes with a linking vowel. More importantly, the resulting libfixed formations tend to preserve the metrical structure of the source-word, e.g. the prosodic contour of *entertainment* is preserved in most words formed with -*tainment*, just like in blending (see below). Moreover, in Norde and Sippach's (2019) dataset, blend-originating libfixes, e.g. -*tastic* (< *fantastic*), display these features more often than neoclassical combining forms attached to native stems, such as -*cracy* (in *thugocracy*) and -*meter* (in *crapometer*). In our opinion, this finding suggests that in fact the two categories are rather distinct. Neoclassical combining forms may be used in playful formations just like any other regular word-formation process: see, for instance, playful formations with affixes such as -*ee* or *non*- in Bauer (1983). Thus, the distinctive feature of libfixes should not be playfulness, although libfixed formations are somehow marked and informal.

Although we agree that the 'liberation' of a libfix implies reanalysis, we believe that this reanalysis occurs only after the source-word has undergone blending. Therefore, in line with Norde and Sippach (2019), we consider libfixing a category on its own, situated on a gradient between blending and affixation. In fact, many researchers have noticed that blending often triggers the creation of morphemes in time (Fischer 1998; Fradin 2000; Kemmer 2003; Lehrer 2007; Mattiello 2018; Barrena Jurado 2019; Beliaeva 2019), the typical scenario displaying important stages: first, an iconic blend becomes a model for other blends, thus creating a series that models other analogical formations that can further lead to the creation of a new bound morpheme (Kemmer 2003; Mattiello 2018). Mattiello (2023: 187–203) calls the splinters used in analogical formations 'combining forms in the making', and we call them libfixes.

Thus, we characterize libfixes as an intermediary phase between blending and bound morphemes (Figure 1). In order to offer a fine-grained definition, there should be clear points that would mark the beginning and the end of the libfix phase. Nevertheless, as language is continuously changing, it is pretty much impossible to find clear cut-off points. The main criterion is the productivity of the form (see below), but yet again, it is almost impossible to state clearly how many occurrences formed with a liberated part are needed for it to pass the border between libfix and bound morpheme.
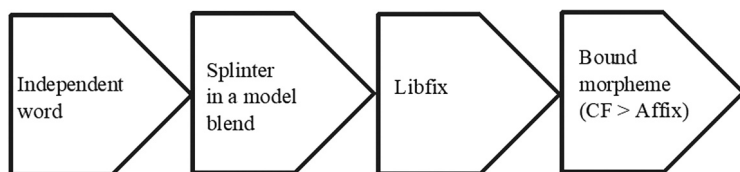
Figure 1: Libfixes on the blending – morpheme continuum.

In our opinion, when a splinter in a model-blend is used again to coin a new word, the creation of a libfix begins. When a blend is coined, it is modelled according to the properties of the two source-words; a subsequent formation takes into account not only the two source-words, but also the previous blend. Thus, the new word is shaped by taking into account more factors than in the case of the original blend. For instance, the libfix -*aholic* (< *alcoholic*) has various forms, such as -*holic*, -*oholic*, but -*aholic* appears to be dominant most likely due to the influence of *workaholic*, with the /a/ occurring whenever the root ends in a consonant (Lehrer in Fradin 2000: 17). This suggests that the original blend heavily influences subsequent formations. However, the creation of a libfix is never witnessed with its second occurrence, but can be observed once it has offspring, i.e. when a number of libfixed creations have already been established; this leads to the observed productivity of libfixes (Norde and Sippach 2019: 380).

The end of the libfix phase coincides with the establishment of a new morpheme whose birth is again difficult to establish, and the transformation is strongly linked with the formant's productivity (Lehrer 2007: 121; Beliaeva 2019; Barrena Jurado 2019; Mattiello 2023: 57). Yet, no fixed number of occurrences could be applied to all libfixes and in all languages so as to say that, from that moment on, a particular form has turned into a bound morpheme. Fischer (1998: 65) argues that "at least three neologisms containing the blend element must be found before the existence of a new combining form may be assumed", a number that has not yet reached consensus among researchers. What we retain from Fischer's (1998) claims is that morphemization implies both quantity and quality: the formants are offshoot only

after the creation of neologisms, i.e. new words that are embraced by speakers and become part of the language. Therefore, the creation of nonce words, i.e. words coined with a specific occasion that disappear once the specific occasion is gone, does not suffice (for the distinction between neologisms and nonce words, see Bauer 2001: 38–40; Mattiello 2017: 23–26). Nevertheless, the fact that a particular formant does appear in nonce words suggests that speakers find it easy to use it to create new words (Barrena Jurado 2019).

Lehrer (2007: 121) argues that new morphemes are created when the formants are no longer connected to their source-words, but does not explain how this connection or the loss thereof manifests itself. The connection could be marked by what Norde and Sippach (2019: 360) call "prosodic preservation", i.e. libfixed formations tend to preserve the number of syllables and the place of the main stress of the source-word, e.g. most -*tainment* formations are 4-syllable words stressed on the penultimate syllable, just like *entertainment*. Discussing blends, Beliaeva (2019) observes that prosodic preservation is negatively correlated with type frequency: the more productive the formant (which she calls splinter), the less it maintains a connection to its source-word. Mattiello considers the transition from splinter to combining form to be complete when metrical deviation from the original pattern is present (Mattiello 2023: 57). This idea relies on a salient, but not mandatory characteristic of blending, the source of libfixing: not all blends preserve the metrical contour of (at least) one source-word, although the majority do (Kemmer 2003; Arndt-Lappe and Plag 2013; Renner 2019 for English, Vasileanu and Niculescu-Gorpin 2022 for Romanian).

Another means of evaluating the connection of the libfix to the source-word is psycholinguistic testing. Niculescu-Gorpin and Vasileanu (2025) tested the processing and understanding of Romanian blends, and only marginally of libfixed formations, by administering a questionnaire. Romanian native speakers were presented with various stimuli, including libfixed formations, and they had to explain, in their own words, how they thought the target words were coined. Three libfixes, -*aholic*, -*gate* and -*exit*, were included in the questionnaire. The analysis showed that the stimuli formed with these libfixes were mainly perceived as blends, since subjects indicated two complete source-words in their responses. For example, most subjects indicated *Watergate* as a source-word for *Udreagate*, the stimulus containing -*gate*. -*Aholic* was slightly more often interpreted as an affix than as a part of a blend; nevertheless, when blending was referred to, En. *alcoholic* and even Ro. *alcoolic* 'alcoholic' were mentioned as source-words. There were also answers pointing to analogical models, mostly to *workaholic* which, due to its high frequency, seems to have triggered the analogical series (Mattiello 2018: 10). The analysis in Niculescu-Gorpin and Vasileanu (2025), although based on a small dataset, suggests that libfixes do preserve a connection to their source-words that is

still perceived by speakers, i.e. speakers still relate the libfixed formation to the source-word of the libfix. This connection is maintained even for borrowed libfixes, as, in order to understand words, speakers appeal to their linguistic knowledge, be it first or second language. The study also suggests that model blends may play a more important part than previously thought in libfixing, a hypothesis also supported by our case study on *-pedia* (see Section 4.1). Last but not least, this connection suggests that Romanian libfixing may be a case of structural borrowing (see below).

We therefore define libfixes as blend-emerging word parts that still preserve their connection with their initial source-word and that are used in analogical combinations to create new words. We believe that the libfixing process could involve the following steps. First of all, two source-words combine in a blend. One of the source-words, that from which the libfix will later be clipped, creates further blends with other source-words. It is pretty difficult to actually capture this moment in time as speakers are creative and new blends are born without being recorded. Nevertheless, it is plausible to assume that, analogically, new blends are formed based on a model blend. This process is then repeated until the libfix is completely broken from its source-word. Once the split is complete, i.e. there is no connection to the prosodic structure of the source-word and people do not link the part of the word with that source-word or with the model blend, the libfix turns into a bound morpheme. As word processing and word creation are yet impossible to be observed directly, we believe that the findings in the literature support our hypothetical scenario of how libfixes are created and evolve.

Thus, libfixed formations differ from blends both in their analogical nature and in their productivity, here understood mainly as type frequency, and from combining forms and affixes in their connection to the source-word they have been clipped from, mainly visible in the prosodic structure of outputs.

2.2 Libfixing and Contact-Induced Language Change

As the current *lingua franca*, English has become the donor language for various types of contact-induced language change phenomena, lexical items being the most numerous and thus an eye-catching type of transfer. However, besides enriching other languages, the massive influx of English loanwords may generate some changes in the word-formation system due to structural borrowing, i.e. the borrowing of abstract morphological schemata (Renner 2018). If the contact-induced change of the frequency of a word-formation pattern is understood as structural borrowing, English appears to have impacted the word-formation systems of other languages, e.g. by increasing the scope of clipping in Polish or of noun-noun compounding in French and Polish (Renner 2018). Hamans (2021:

674–676) shows that Dutch libfixing has become much more frequent when the productivity of English libfixing increased, mainly in American English. Although the process may have become more frequent in each language independently, it is more likely that the Dutch frequency increased partly due to the borrowing of certain libfixes, such as *-tainment*.

English libfixes have been borrowed in languages such as Italian (*-gate* and *-nomics*, Vaccarelli 2019), Bulgarian (*-aholic*, Stamenov 2015), Polish (*-aholic* and *-gate*, Konieczna 2012: 65), and Romanian (*-gate*, Stoichițoiu Ichim 2006: 70), where they combine also with native bases. Since only recently libfixing has been conceived as a word-formation process of its own, data in these languages are scarce. However, the premises for structural borrowing are created by the English-induced increased frequency of blending, as lifbixing is genetically related to blending. Moreover, the recent rise of lexical blending in Bulgarian (Stamenov 2015), Italian (Cacchiani 2016), and Dutch (Hamans 2021) has also been connected with the English influence.

In present-day Romanian, lexical blending has increased its frequency due to the pervasive English influence. Experimental results indicate that international blends are understood better than autochthonous ones (Niculescu-Gorpin and Vasileanu 2025), and corpus data show that international blends have gained wider circulation whereas autochthonous ones, though more numerous, tend to remain nonce words (Vasileanu and Niculescu-Gorpin 2022). Moreover, lexical blending is an active word-formation process and it has become the default means of naming hybrid objects at least for the younger generation, as suggested by an elicitation experiment (Vasileanu, Niculescu-Gorpin and Radu-Bejenaru 2024). In this experiment, subjects were asked to name, using only one word, hybrid entities, mostly imaginary. 68.31% of Romanian outputs were blends, whereas only 12.22% were compounds, which was unexpected since compounding is a major word-formation process in Romanian, unlike blending.

The connection between blending and libfixing sets the premises for interpreting libfixing as another case of structural borrowing in Romanian. Ten Hacken and Panocová (2020: 7) observe that speakers borrow words, not rules, and that borrowing of structures implies re-analysis, making structural borrowing a gradual, diachronic process. In order to claim that libfixing is a case of structural borrowing, we need to see: (a) the number of borrowed libfixed formations; (b) the number of libfixed formations with Romanian bases, as they attest the re-analysis of the previous borrowings; (c) the emergence of new, native libfixes, considering also the fact that libfixing has not yet been observed in the literature on Romanian word-formation.

3 Methodology

The current study on libfixes has sprung from a corpus-based study of lexical blending in Romanian (Vasileanu and Niculescu-Gorpin 2022). To identify blends, we manually checked a wordlist resulting from a 60,000,000-word corpus, made of Romanian newspapers, magazines, blogs and satirical websites from the past 15 years. In doing so, we noticed that some words resembling blends actually formed series that had the same final part in common. There were 45 words formed with 7 libfixes. These were the international *-aholic*, *-gate*, *-aton* (= En. *-athon*), *-ghedon* (= En. *-geddon*), *-zilla*, and *-pedia*, alongside a Romanian one, *-izdă* (< *pizdă* '(slang) vulva'). To further investigate their productivity, we checked two larger Romanian corpora: the one-billion word CoRoLa, currently the reference corpus of Romanian (Barbu Mititelu, Tufiș and Irimia 2018) and RoTenTen16, a 2.6-billion-word corpus, part of the TenTen family (Jakubíček et al. 2013). Since Romanian is a language with a rich inflective system and definite articles are bound to the end of words, our queries included the libfix in a medial position, e.g. $^*aholic^*$. We also took into account phonological alternations that might occur, e.g. a word ending in *-aton* might be pluralized as *-atoane*. The resulting lists were manually checked to discard words that happened to contain the same letter strings, e.g. proper names or *-gate* 'door, entrance' in compounds. When necessary, we checked the actual contexts to clarify the meaning.

As we could not establish whether the three corpora contained completely different texts or whether there was some degree of overlap, we could not determine real token frequencies that could be used as an indicator of productivity. Moreover, libfixed formations were mostly nonce words, with low token frequency; frequency per million words resulted in such small fractions that they cannot be a reliable source of actual productivity. In addition, due to the nature of the input texts, there were also huge differences in terms of token frequency in the two larger corpora: CoRoLa is largely based on published, mainly legal and scientific texts, whereas RoTenTen16 is a web-based corpus, comprising more informal language. Therefore, we assessed the productivity of each libfix based only on type frequency (see Section 4).

Each libfixed word was annotated with respect to (a) the base type (whole word, acronym, multi-word unit, combining form, clipped base[2]); (b) the number

---

[2] We counted established clippings as whole words, e.g. we considered *Maimuzilla* 'name of a Romanian shop' < *maimu*, established clipping of *maimuță* 'monkey', as formed from a whole word, and not from a clipped base.

of syllables; (c) the stress position; (d) for international libfixes, whether the word was coined in Romanian or in another language. The analysis was not always straightforward and required additional searches. In some cases, the presence of Romanian inflection markers eased the task. For example, a Romanian forum user, in a Romanian text, used the expression *"view who's online"-oholică* to describe herself, where the final *-ă* is the Romanian feminine singular ending. In the case of brand names, the task was more complicated. Commercial names are sometimes coined from English bases, even for autochthonous products. For instance, a Romanian parenting platform is called *Muffypedia,* but we could not identify a foreign model for the word. In this case, we considered the formation a pseudo-Anglicism (for an overview see Furiassi and Gottlieb 2015) as the word does not exist in English, be it British or American, and it has actually been created in Romanian. The reasons underlying such occurrences are manyfold, such as: English or English-looking names attract attention, they are seen as more prestigious; or, for online language, a word formed from English bits may yield better web-search results. In other cases, the same word was used in several brand names, some Romanian, some international, and we cannot clearly state whether the companies or products were related or not; even for clearly unrelated companies, we cannot be sure the names have been coined independently. Such words were marked ambiguous, and placed in a separate category. In case the meaning of the Romanian word was different from the meaning of an identical international word, we considered that the two were coined independently. For instance, we marked *pornaghedon* 'a sudden end to a politician's career due to video recordings of him having sex outside marriage' as a Romanian creation, since En. *pornageddon* means 'a sudden end of porn films'.

Our results are presented in Section 4, with Sections 4.1–4.7 each analysing one of the following libfixes: *-pedia, -gate, -aholic, -zilla, -aton, -ghedon,* and *-izdă.* Section 4.8 is a final overview of the data and their interpretation.

4 Results: Corpus Findings

We have found 558 libfixed types in the three analysed corpora. Different spellings of the same word were counted together, e.g. we counted *bagoholic* and *bagoholic* as one type. Likewise, adapted spellings of an English loanword were counted together with the original spellings, e.g. *swimathon* and *swimaton* were considered one type. Different forms of the same libfix were also counted together, e.g. *Vunkaholic* and *Vunkoholic* (< *Vunk,* name of a Romanian music group) were considered one single type and counted as one word.

There is a striking difference in frequency between the three most productive international libfixes with type frequency of over 100 each, *-pedia*, *-gate*, *-aholic*, and the other three, *-zilla*, *-aton* and *-ghedon* (see Table 1 below). Interestingly, though originating in a taboo word, the native Romanian libfix *-izdă* is very productive. One possible explanation is that people do like to use taboo words, but prefer to dress them up. Moreover, such uses can either be funny or derogatory or both, but they are definitely attracting the readers' / hearers' attention.

Table 1: Raw type frequency for the 7 libfixes.

Libfix	Total type frequency
-pedia	178
-gate	125
-aholic	120
-zilla	33
-at(h)on	20
-g(h)ed(d)on	12
-izdă	70

4.1 *-pedia*

With 178 types, the most productive libfix is *-pedia*, extracted initially from *encyclopedia* and preserving its core meaning, that of a comprehensive, highly informed data collection. The new *-pedia* is different from the neoclassical combining form *-pedia* 'education, learning' (< Gr. *παιδεία*), therefore, we discarded words such as *hypnopedia* 'learning during sleep' from the analysis.

All 178 types are proper names of companies, products, websites etc., 88 being loanwords, and only 56 most likely coined in Romanian, according to the criteria mentioned above; there are 34 items for which we cannot establish or refute a connection between Romanian and identical international words.

Only items created in Romanian will be discussed, since we are concerned with Romanian libfixing.

In Romanian, the international libfix *-pedia* combines with a variety of base types as shown in Table 2. Most frequently, it is bound to whole words (36 items), a behaviour that resembles that of an affix. Out of these, in 8 instances, the libfix is preceded by an interfix, e.g. *-o-* in *tristopedia* 'ironic name for a website' < *trist* 'sad' + *-o-* + *-pedia*, or *-i-* in *blondipedia* < *blonde* 'blondes' + *-i-* + *-pedia*. However, *-o-* may also be interpreted as part of the *-opedia* clipping of *encyclopedia*, an interpretation supported by the form *comentopedia* 'the title of the comments section of a newspa-

per' < *comentarii* 'comments' + *-opedia,* where it combines with a clipped base in a blend-like manner. In this context, we have interpreted *-opedia* as an allomorph of *-pedia* (in case not-yet-morphemes can be said to have allomorphs).

Table 2: Base types for Romanian libfixed *-pedia* words.

Base types	n.	Example
whole word	36	*citatepedia* 'name of a famous quotation website' < *citate* 'quotations' + *-pedia*
acronym	4	*RATT-pedia* 'website on the public transportation system of Timișoara' < *RATT* 'the public transportation system of Timișoara city' + *-pedia*
multi-word unit	1	*mysportpedia* 'name of a website' < *my sport* + *-pedia*
combining form	2	*cronopedia* 'name of a literary circle and its website' < *crono-* 'chrono' + *-pedia*
clipped base	13	*redupedia* 'name of a discount website' < *reduceri* 'discounts' + *-pedia*

Combinations with clipped bases, e.g. *redupedia,* and with combining forms, e.g. *chronopedia,* are evidence that, although productive, *-pedia* is still on the blending – morpheme continuum, thus underlining its libfix status. To assess its position on this continuum, we need to consider also the metrical properties of the outputs.

Being a stressed formant, *-pedia* should create items stressed on the same syllable, i.e. *-pe-.* Although the libfix has been extracted from a 6-syllable word, *encyclopedia,* most items display a 5-syllable structure (Table 3). This is not random. Out of the 13 items formed from clipped bases, 9 words have 5 syllables, which means that they were purposely clipped to fit the metrical pattern. Our hypothesis is that, in this particular case, the metric structure was not modelled by the source-word *encyclopedia,* but by the more recent, but very popular *Wikipedia,* which may also explain the occurrence of *-i-* as an interfix. The hypothesis is also supported by the fact that some of the *-pedia* formations are in fact names of collaborative websites, such as *citapedia, turismpedia.*

With a type frequency of 178 items, *-pedia* also comprises the items with the highest token frequency. For instance, *softpedia* has the highest token frequency, occurring 1315 times (0.505 pmw) in the RoTenTen16 corpus. Its productivity is definitely triggered by the advance of the Internet and social media and the existence of large collections. The metrical properties of *-pedia* words indicate the influence of the model blend Wikipedia, placing it towards the morpheme end of the continuum.

Table 3: The metrical structure of the Romanian *-pedia* libfixed formations; all items are stressed on the antepenultimate syllable.

Number of syllables	n.	Example
4 syllables	8	*mobpedia* 'name of website dedicated to furniture' < *mobilă* 'furniture' + *-pedia*
5 syllables	33	*imopedia* 'name of a real estate website' < *imobiliare* 'real estate' + *-pedia*
6 syllables	11	*pensiopedia* 'name of a website on private pensions' < *pensie* 'pension' + *-pedia*
7 syllables	4	*cosmeticopedia* 'name of an online cosmetics store' < *cosmetice* 'cosmetics' + *-opedia*

4.2 *-gate*

The second most productive libfix is *-gate*, 'public scandal (usually followed by cover-up attempt)', with 125 types, out of which 95 are loanwords, 30 Romanian, and 1 ambiguous. It is a relatively young formant, since it was extracted from *Watergate*, the name of the office building in Washington DC where people connected with the Nixon administration broke into a headquarters of the Democratic Party in 1972. In the years after Nixon's resignation, *-gate* was a popular formative. Its productivity heavily depends on extralinguistic factors, i.e. public scandals (Mattiello 2023). The OED currently labels *-gate* as a combining form.

Table 4: Base types for Romanian libfixed *-gate* words.

Base types	n.	Example
whole word	25	*Băsescugate* < *Băsescu* 'name of a former Romanian president' + *-gate*
acronym	3	*SRIgate* < *SRI* 'Romanian Intelligence Service' + *-gate*
multi-word unit	2	*Băsescuberceamondialgate* < *Băsescu* 'name of a former Romanian president' + *Bercea Mondial* 'name of a shady businessman' + *-gate*

Unlike *-pedia*, *-gate* does not combine with clipped bases or with combining forms, but mainly with whole words (25 items out of 30, see Table 4 above), which is typical for affixes (see Bauer 1983: 213).

The metrical structure of *-gate* formations also indicates an affix-like behaviour: the newly formed words retain the stress of their bases, the libfix being unstressed, just like in English. Moreover, Table 5 shows that the formations display a variety of metrical patterns: *Watergate* is a trisyllabic word with stress on the first, whereas less than half of the Romanian formations with *-gate,* i.e. 13 items, are 3-syllable words, and only 9 have first syllable stress, preserving the metrical

structure of the source-word. Therefore, prosodic patterns indicate a loose connection with the source-word *Watergate*.

Table 5: The metrical structure of the Romanian *-gate* libfixed formations; all formations preserve the stress of the base.

Number of syllables	n.	Example
2 syllables	3	*romgate* < *rom* 'Roma' + *-gate*
3 syllables	13	*Udreagate* < *Udrea* 'name of a former Romanian minister' + *-gate*
4 syllables	9	*Turcescugate* < *Turcescu* 'name of a former Romanian journalist' + *-gate*
5 syllables	2	*Teleormangate* < *Teleorman* 'name of a Romanian county' + *-gate*
6 syllables	1	*Elisabetagate* < *Elisabeta* 'name of a Romanian prostitute' + *-gate*
7 syllables	1	*Băsescu-Mondial-gate* < *Băsescu* 'name of a former Romanian president' + *Mondial* 'name of a shady businessman' + *-gate*
9 syllables	1	*Băsescuberceamondialgate* < *Băsescu* 'name of a former Romanian president' + *Bercea Mondial* 'name of a shady businessman' + *-gate*

Due to its productivity, morphological and metrical properties, in Romanian *-gate* seems to be situated more towards the affix part of the continuum.

4.3 *-aholic*

Another productive libfix is *-aholic*, with 120 types: 77 international, 13 ambiguous, 30 Romanian. The OED has recently changed its status from combining form in the 2nd edition to suffix in the 3rd edition (Mattiello 2018: 14). This change suggests that recent formations with *-aholic* take only whole words as bases. However, this is not the situation in Romanian, as shown in Table 6. Although in the majority of original Romanian types *-aholic* is bound to whole words, in a few formations it is attached to clipped bases, e.g. *călătoholic* 'addicted to travelling' < *călători* 'to travel', and combining forms, e.g. *ihtioholic* 'addicted to fish' < *ihtio-* 'ichtyo-', which do not usually combine with affixes (see Bauer 1983: 214–215; Mattiello 2023: 13).

With respect to the metrical structure, the most frequent pattern (14 items) is a 4-syllable word stressed on the penultimate syllable (see Table 7), which is exactly the metrical pattern of *alcoholic*. We believe that this prosodic structure was reinforced by *workaholic*, which appears to be the model blend that triggered further analogical creations and the liberation of *-aholic*, according to the diachronic study of Mattiello (2018: 10). In some cases, the choice of a specific allomorph of

Table 6: Base types for Romanian libfixed *-aholic* words.

Base types	n.	Example
whole word	23	*curcubeoholic* 'addicted to rainbows' < *curcubeu* 'rainbow' + *-oholic*
multi-word unit	1	*"view-who-s-online"-oholic* 'a person who compulsively checks who is online' < *view who's online* + *-oholic*
combining form	1	*ihtioholic* 'addicted to fish' < *ihtio-* 'ichthyo-' + *-holic*
clipped base	5	*călătoholic* 'addicted to travelling' < *călători* 'to travel' + *-holic*

-aholic was made so as to fit the model metrical structure. For instance, *biciholic* 'addicted to biking' is formed on the base *bici,* a disyllabic established clipping of *bicicletă* 'bicycle', which combines with *-holic* to form a 4-syllable word. Nevertheless, deviations from the metrical model suggest that the connection to the source-word is loosening even in Romanian.

Table 7: Metrical structure of Romanian *-aholic* formations; all formations are penultimate-syllable stressed.

Number of syllables	n.	Example
3 syllables	1	*Betholic* 'addicted to Beth' < *Beth* 'Beth Hart' + *-holic*
4 syllables	14	*brânzoholic* 'addicted to cheese' < *brânză* 'cheese' + *-oholic*
5 syllables	11	*impozitholic* 'addicted to tax' < *impozit* 'tax' + *-holic*
6 syllables	2	*înghețatoholic* 'addicted to icecream' < *înghețată* 'icecream' + *-oholic*
7 syllables	2	*telenoveloholic* 'addicted to soap opera' < *telenovele* 'soap operas' + *-oholic*

-Aholic has been discussed in numerous papers as an example of meaning secretion (Warren 1990, Mattiello 2018, 2023 a.o.): the base clipping comes with a "semantic clipping": *-aholic* 'addicted to' preserves part of the meaning of *alcoholic* 'addicted to alcohol'. Moreover, it seems that *-aholic* is undergoing melioration. If *alcoholic* 'addicted to alcohol' has a negative meaning, *curcubeoholic*, literally 'rainbow-aholic', should be understood as 'passionate about rainbows', with a positive connotation. This semantic shift is yet another indicator that *-aholic* is losing connection to its source-word.

The morphological, metrical and semantic features mentioned indicate *-aholic* as more advanced on the path to morphemization.

4.4 *-zilla*

En. *-zilla* is registered in OED as a combining form. Extracted from *Godzilla*, it has become more and more productive, and it is mostly bound to monosyllabic whole word bases (Mattiello 2018: 12). In Romanian, we identified 33 types in the three corpora: 20 international, 12 Romanian, and 1 ambiguous. The Romanian formations are coined from whole words or clipped bases, as shown in Table 8.

Table 8: Base types for Romanian libfixed *-zilla* formations.

Base types	n.	Example
whole word	8	*veveritzilla* 'huge squirrel' < *veveriță* 'squirrel' + *-zilla*
clipped base	4	*șnitzilla* 'huge schnitzel' < *șnițel* 'schnitzel' + *-zilla*

Just like in English, by adding the libfix to monosyllabic bases, Romanian formations tend to preserve the metrical pattern of *Godzilla*, a 3-syllable word with a stressed penultimate syllable (see Table 9). However, other metrical patterns deviate from the original, suggesting *-zilla* is also turning independent.

Table 9: Metrical patterns of Romanian libfixed *-zilla* formations; all items are penultimate-syllable stressed.

Number of syllables	n.	Example
3 syllables	7	*mortzilla* 'nickname of a forum user' < *mort* 'dead' + *-zilla*
4 syllables	3	*mamazilla* 'momzilla' < *mamă* 'mother' + *-zilla*
5 syllables	2	*cotcodaczilla* 'huge hen' < *cotcodac* 'cluck (interjection)' + *-zilla*

Semantically, *-zilla* has also undergone a process of meaning secretion, as it preserves only certain meaning features of the source-word, such as 'huge' or 'frightening'. In English, corpus data show that *-zilla* is added usually to nouns, mostly names of animals or family members. Mattiello (2018: 12) found only one example where it combines with an inanimate base, *brandzilla*, which she interprets as a metaphorical extension. Romanian examples suggest that this extension is continued cross-linguistically, as shown by examples such as *Contzilla* 'name of an accounting website' < *contabilitate* 'accounting', *gropzilla* 'huge pothole' < *groapă* 'pothole', or even an on-the-spot creation *blackzilla* 'overcooked meat' (although formed from an English base, we could not identify an English *blackzilla* with the same meaning, making this word another pseudo-Anglicism).

Given all of the above, we can argue that, although not very productive, *-zilla* is halfway through the morphemization process, slightly leaning towards the bound morpheme end of the continuum; in time, it may turn into an augmentative suffix.

4.5 *-aton* (= En. *-athon*)

Extracted from *marathon*, *-athon* has been productive in English for quite some time. Marchand (1960: 161) called words such as *danceathon* and *swimathon* "barbarisms". Warren (1990) includes *-athon* on her list of secreted combining forms, as it does not preserve the entire meaning of the source-word, but has gone through a process of semantic broadening, creating names of prolonged activities resembling a marathon.

In Romanian, *-aton* is not at all productive. Out of the 20 items containing this libfix, 18 are English loanwords and only 2 Romanian formations (see Table 10). The loanwords may be graphically adapted. For instance, 2 items occur with both English and Romanian spelling: *hackathon* and *hackaton*, *swimathon* and *swimaton*. Another 3 items have been attested in our corpora only with a Romanian spelling of the libfix: *ciclaton, sexaton, techaton*. This suggests that, although the libfix is not productive in Romanian, people understand it and connect it to its source-word, both En. *marathon* and the Ro. *maraton*.

Table 10: Romanian formations with the libifix *-aton* = *-athon.*

Base types	n.	Example
whole word	2	*varaton* 'long vaccination session during the summer' < *vară* 'summer' + *-aton*
		Mayaton 'long concert of Maya' < *Maya* 'name of a singer' + *-aton*

As can be seen in Table 10, the two Romanian items are formed by adding the graphically adapted libfix to whole words. Both items preserve the metrical structure of *maraton* 'marathon', i.e. they are three-syllable words with final-syllable stress. Moreover, *varaton* may also be interpreted as a blend of *vară* 'summer' and *maraton* 'marathon', with a larger overlapping segment, a characteristic of initial-phase libfixes. As Kemmer (2003: 89) points out in her case study of En. *-erati*, when a splinter is detached from a blend and is further attached to other bases, speakers tend to choose bases that share phonological segments with the splinter.

The low productivity of *-aton* and the metrical characteristics of the two Romanian creations indicate that this libfix stands more towards the blending end of the continuum.

4.6 *-ghedon* (= En. *-geddon*)

The analysis of Ro. *-ghedon* reveals similar patterns to Ro. *-athon*. We identified only 12 items in the three Romanian corpora: 8 international words, 3 Romanian formations, and 1 ambiguous. In English, *-geddon* is secreted from *Armaggedon*, originally a proper name designating the Biblical battlefield where the final fight between the Good and the Evil would take place (Revelation 16:16). The meaning of *Armageddon war* further broadened to name any decisive confrontation. Lalić-Krstin (2014: 260) found 102 English items that she considers blends with *armaggedon* as a second source-word. This suggests high productivity for the libfix in English.

Just like in the case of *-athon*, En. *-geddon* is recognized by Romanian speakers who connect it with En. *Armageddon* and/or Ro. *armaghedon*, which is stressed on the final syllable. Since we only relied on written corpora, it is impossible to determine whether Romanian speakers connect a particular libfix with its English or Romanian counterpart. Nevertheless, both items would be highly activated in the speakers' minds. This may explain why 2 out of the 8 international libfixed formations occur both with English and Romanian spelling: *eurogeddon* and *euroghedon*, *Romageddon* and *Romaghedon*. Conversely, the variation can lead to ambiguous situations such as *tramvagedon*, in which the libfix is spelled neither as in English (*-geddon*, with a double *-d*-), nor as in Romanian (*-ghedon*).

The examples in Table 11 suggest that *-ghedon* is also an initial-phase libfix in Romanian. In 2 cases, the libfix is bound to clipped bases; in the third example, a double interpretation is possible, both as a libfix attached to a whole word and as an overlap blend of *sarma* 'meat roll' + *armaghedon*. All 3 Romanian items preserve the metrical structure of *armaghedon*.

Table 11: Romanian formations with the libfix *-ghedon* = *-geddon*.

Base types	n.	Example
whole word	1	*sarmaghedon* 'a meal where one eats too many meat rolls' < *sarma* 'meat rolls' + *-ghedon*
clipped base	2	*tramvagedon* 'imaginary tram crash' < *tramvai* 'tram' + *-gedon* *pornaghedon* 'a sudden end to a politician's career due to video recordings of him having sex' < *porno* 'porn' + *-aghedon*

The Romanian dataset is too small to draw strong conclusions; nevertheless, the existing forms suggest that *-ghedon* lies more towards the blending end of the continuum. It may develop into a more productive libfix and further into a com-

bining form, due to the productivity of En. *-geddon*. For the moment, it is infrequent and still preserves blending features.

4.7 *-izdă*

The only original Romanian libfix is extracted from a taboo word, *pizdă* '(slang) vulva', therefore all items containing it are mainly offensive, but there is also an idea of playfulness and sarcasm. This is why real type and token frequency may actually be higher in this case: taboo language is even more taboo in writing.

In 64 out of 70 types, the libfix is bound to clipped bases (see Table 12), and only in a few it is attached to whole words or acronyms, a situation which is completely different from the international libfixes discussed above. Interestingly, all 64 items formed from clipped bases preserve the metrical structure of the base, e.g. *jurnalizdă* is a 4-syllable word, with its stress on the penultimate syllable, just like the base *jurnalistă* 'female journalist'. This embedding of *-izdă* in the metrical contour of the base is favoured by the phonological similarity of the base. 58 of the 64 clipped bases end in *-istă*, with the only difference in the voicing of the consonants, e.g. *feminizdă* '(offensive) female feminist' < *feministă* 'female feminist', *artizdă* '(offensive) female artist' < *artistă* 'female artist'. The other 6 words are also professional names formed with a different suffix, e.g. *bucătărizdă* '(offensive) female cook' < *bucătăreasă* 'female cook' < *bucătar* 'male cook'.

Table 12: Base types for libfixed formations with *-izdă*.

Base types	n.	Example
whole word	5	*Merkelizdă* '(offensive) Merkel' < *Merkel* + *-izdă*
acronym	1	*gesepizdă* '(offensive) female journalist at the Sports Gazette' < *GSP* 'Sports Gazette' + *-izdă*
clipped base	64	*jurnalizdă* '(offensive) female journalist' < *jurnalistă* 'female journalist' + *-izdă*

Due to the taboo nature of the libfix, it is difficult to embark on a diachronic study of the *-izdă* formations. Yet, there are reasons to assume that the libfix was liberated after the creation of a blend from the source-word *pizdă* with a professional name ending in *-istă*, catalysed by the phonological similarity of the final word parts. Then, the splinter took as bases other feminine professional names that ended differently, e.g. *bucătărizdă* '(offensive) female cook' < *bucătăreasă* 'female

cook' < *bucătar* 'male cook'. In that phase, the process resembled affix substitution, and the next step was the combination with whole bases in an affix-like manner.

Interestingly, for 18 out of the 70 feminine nouns formed with -*izdă*, there are masculine back-formations in our corpora, e.g. *ziarizd* '(offensive) male journalist' < *ziarizdă* '(offensive) female journalist < *ziaristă* 'female journalist'. Moreover, we found 9 masculine words derived with -*izd* for which there was no feminine counterpart in the corpora, e.g. we found *minimalizd* '(offensive) male adept of minimalism', but not **minimalizdă* '(offensive) female adept of minimalism'. This suggests an emerging masculine form of the libfix. Semantically, such masculine formations are offensive both for males and for females, as they imply that the males they are referring to have a more female-like behaviour that is seen as weak or, in any case, derogatory. As a result, these words are also disrespectful towards women.

As it is no longer confined to a feminine noun base, -*izdă* is undergoing a process of meaning generalization. Yet, it is still subjected to semantic restrictions, as it combines with animates, with the aim of pejorating the base word.

Assessing the exact position of -*izdă* on the blending – bound morpheme continuum is difficult, if not impossible. The combination with clipped bases could suggest the blending end of the continuum; yet, as shown above, the clipping pattern, i.e. the fact that -*izdă* usually replaces an affix, indicates an affix-like behaviour. The meaning generalization and the creation of a masculine counterpart also underline a higher degree of morphemization. As both the masculine and the feminine coinages retain the emotional force of a taboo word, the connection to the source-word is maintained rather semantically than metrically, i.e. the outputs do not preserve the metrical structure of the source-word.

4.8 Libfixing as Structural Borrowing

Libfixing, understood here as the liberation of a word part from a source-word as a consequence of blending, has never been observed, studied or discussed in relation to the Romanian language before. The seven case studies presented above illustrate a novel word-formation process in Romanian.

Studying Romanian corpora has allowed us to embark on a fine-grained analysis of its origin, and to establish whether it is an internal evolution of the Romanian word-formation system or a contact-induced language change.

As shown in Section 2.2, structural borrowing implies the borrowing of abstract morphological schemata carried out via loanwords. For libfixing, we argue that the process implies first borrowing lexical items, then the combination of international libfixes with native bases and, then, in this context, new native libfixes may emerge.

Our case studies support the hypothesis that all international libfixes have entered Romanian via a series of loanwords. For each international libfix, there are only a few autochthonous formations (see Table 13). Nevertheless, the fact that such libfixes combine with native bases is evidence that Romanian speakers have indeed reanalysed the loanwords, extracted and (re)applied the morphologic rule. The emergence of a native libfix, -*izdă*, a quite productive one, is further evidence that Romanian libfixing is a case of structural borrowing.

Table 13: Total types and autochthonous types.

Libfix	Total type frequency	Autochthonous types	
		n.	%
-*pedia*	178	56	31.46%
-*gate*	125	30	24%
-*aholic*	120	30	25%
-*zilla*	33	12	36.36%
-*aton*	20	2	10%
-*ghedon*	12	3	25%
-*izdă*	70	70	100%

Due to the existing English-Romanian language contact, and with the omnipresence of English on the Internet, structurally borrowing libfixing and its incorporation into Romanian are multidimensional rather than linear. When borrowed, each libfix brings into Romanian not only a form-meaning pair, but also traces of its history, i.e. the connection to its source-word.

For instance, OED already acknowledges -*athon* as a combining form in English, but its connection with En. *marathon* is still visible in the Romanian outputs. Romanian speakers did borrow *hackathon, swimathon, danceathon* and other English formations, and recognized -*athon* as a word formant, but also recognized *marathon* as its source and thus its underlying blending formation. In Romanian, the only two formations (*varaton, Mayaton*) may also be interpreted as blends of Ro. *maraton*, since they differ from this word by only one letter and since similarity with the source-words is one of the modelling factors of blends (see, for instance, Gries 2012: 159–162). This is why the two autochthonous formations resemble Ro. *maraton* so much. Autochthonised spellings such as *swimaton, techaton* for English loanwords suggest that the words are also processed in relation with Ro. *maraton*. In this context, we argue that the borrowing of libfixing as a process somehow overlaps with or even is triggered by the increasing productivity of Romanian lexical blending under the influence of English, which may also be considered a case of structural borrowing (see Section 2.2). As we have

shown, lexical blending and libfixing have a strong connection that is passed into the recipient language.

Lexical blending was attested in Romanian prior to the English influence, but only marginally, whereas recently it has become more widespread (see Section 2.2). The increased productivity of lexical blending and the ease with which Romanian speakers now understand and coin lexical blends have opened the path to borrow libfixing. On the one hand, since Romanian native speakers are currently blending more than in the past, it is more likely that the same word will be blended several times and offshoot a libfix; on the other hand, speakers recognize international libfixes as productive word formatives still connected to their source-words.

In their network analysis of English libfixes, Norde and Sippach (2019: 380) show that libfixes form clusters with strong interparadigmatic links, which make English speakers produce and understand libfixed formations easily. Our results further support the idea that these links extend even over language boundaries, and are passed into the recipient languages as non-native speakers are exposed to English libfixed formations almost on a daily basis.

The Romanian libfix *-izdă* is another argument that libfixing has become part of the Romanian word-formation system. The libfix itself is highly productive compared to the number of native creations based on international libfixes. However, its morpho-phonological properties differ from the international libfixes: *-izdă* is bound mostly to clipped bases, and its source-word never imposes its metrical contour over the resulting words. This may be due to the fact that the base is actually the semantic head of the new word, the libfix only adding some extra semantic features. We do not have diachronic corpora to establish the exact moment when this libfix emerged; our own Internet searches indicated *ziarizdă* '(offensive) female journalist' as the oldest formation, from 2008. However, we believe it followed the same process as other libfixes, i.e. it was created as a result of a series of blends. Masculine back-formations, such as *ziarizd* '(offensive) male journalist' indicate semantic generalization, and, along with type frequency, suggest it is advancing on the path to morphemehood. More importantly, the case of *-izdă* suggests that Romanians did not borrow a schema understood as a form-meaning pair, but some sort of procedural knowledge: how to liberate libfixes from blends.

5 Conclusions

Romanian libfixing is an emerging word-formation process whereby speakers form creative, expressive and attention-grabbing words in a partly predictable and routinized way, since the same phonological string with more or less the same meaning is repeatedly attached to various bases. Yet, the combinations are also partly restricted prosodically and semantically: in some cases, bases are clipped to fit certain metrical patterns or to deviate from them as little as possible; in other cases, bases are restricted to a certain semantic category, such as animates.

In line with Norde and Sippach (2019: 380), we considered libfixes to be part of a continuum from blending to combining forms and affixes. The seven case studies present seven word-formants in different stages of morphemization: -*aton* and -*ghedon* are closer to the blending end, -*zilla* is half way through, -*pedia* and -*aholic* are more productive and lean towards the combining form end of the continuum, whereas -*gate* is the closest to becoming a fully-established morpheme. The autochthonous -*izdă* is productive, but still linked to its taboo source-word; its partial semantic generalization and its combination with whole words indicate it is also half-way through the morphemization process.

Our corpus-based analysis shows that Romanian libfixing is indeed a particular case of both material and structural borrowing: speakers borrowed libfixed formations and combined the international libfixes with native bases, but, more importantly, Romanian speakers borrowed the know-how of serial blending and of liberating libfixes, thus enlarging the Romanian word-formation system with a creative and productive process.

References

Arnaud, Pierre J. 2013. Word-formation and word-creation: A data-driven exploration of inventiveness in neologisms. *Quaderns de Filologia-Estudis Lingüístics*, 18. 97–113.

Arndt-Lappe, Sabine & Ingo Plag. 2013. The role of prosodic structure in the formation of English blends. *English Language and Linguistics* 17(3). 537–563. https://doi.org/10.1017/S1360674313000154.

Arndt-Lappe, Sabine. 2018. Expanding the lexicon by truncation: Variability, recoverability, and productivity. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin & Esme Winter-Froemel (eds.), *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and Ludicity*, 141–170. Berlin & New York: De Gruyter Mouton.

Barbu Mititelu, Verginica, Dan Tufiş & Elena Irimia. 2018. The Reference Corpus of the Contemporary Romanian Language (CoRoLa). In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, Takenobu Tokunaga (eds.), *Proceedings of*

the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 1178–1185. Miyazaki, Japan.

Barrena Jurado, Alejandro. 2019. A study on the 'wordgasm': the nature of blends' splinters. *Lexis, Journal in English Lexicology* 14, doi: 10.4000/lexis.3916.

Bauer, Laurie. 1983. *English Word Formation*. Cambridge University Press.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge University Press.

Beliaeva, Natalia. 2019. Blending creativity and productivity: on the issue of delimiting the boundaries of blends as a type of word formation. *Lexis. Journal in English Lexicology* 14, doi: 10.4000/lexis.4004.

Dressler, Wolfgang. 2000. Extragrammatical vs. marginal morphology. In Ursula Doleschal & Anna Thornton (eds.), *Extragrammatical and marginal morphology*, 1–10. München: LINCOM.

Fischer, Roswitha. 1998. *Lexical Change in Present-day English: A Corpus-based Study of the Motivation, Institutionalization, and Productivity of Creative Neologisms*. Tübingen: Gunter Narr Verlag.

Fradin, Bernard. 2000. Combining forms, blends and related phenomena. In Ursula Doleschal & Anna M. Thornton (eds.). *Extragrammatical and Marginal Morphology*, 11–59. München: Lincoln Europa.

Furiassi, Cristiano & Henrik Gottlieb. 2015. *Pseudo-English*. *Studies on False Anglicisms in Europe*. Berlin/Boston/Munich: De Gruyter Mouton.

Hamans, Camiel. 2017. Language change and morphological processes. *Yearbook of the Poznań Linguistic Meeting*, 3. 1–23, doi: 10.1515/yplm-2017-0001.

Hamans, Camiel. 2021. A lesson for covidiots: About some contact induced borrowing of American English morphological processes into Dutch. *Studia Anglica Posnaniensia* 56. 659–691.

Haspelmath, Martin. 2002. *Understanding morphology*. London: Arnold.

Kemmer, Suzanne. 2003. Schemas and lexical blends. In Herbert Cuyckens, Thomas Berg, René Dirven & Klaus-Uwe Panther (eds.), *Motivation in Language: Studies in Honor of Günther Radden*, 69–97. Amsterdam: John Benjamins.

Konieczna, Ewa. 2012. Lexical blending in Polish: A result of the internationalisation of Slavic languages. In Vincent Renner, François Maniez & Pierre J. L. Arnaud (eds.), *Cross-Disciplinary Perspectives on Lexical Blending*, 51–73. Berlin: Walter de Gruyter.

Lalić-Krstin, Gordana. 2014. Forms and meanings of the source word Armageddon in English lexical blends. In Tvrtko Prćić (ed.), *Engleski jezik i anglofone književnosti u teoriji i praksi: Zbornik u čast Draginji Pervaz [English Language and Anglophone Literatures in Theory and Practice: Festschrift in Honour of Draginja Pervaz]*, 257–273. Novi Sad, Filozofski fakultet].

Lehrer, Adrienne. 2007. Blendalicious. In Judith Munat (ed.), *Lexical Creativity, Texts and Contexts*, 115–136. Amsterdam & Philadelphia: John Benjamins.

Marchand, Hans. 1960. *The Categories and Types of Present-day English Word-formation: A Synchronic-diachronic Approach*. Wiesbaden: Otto Harrassowitz.

Mattiello, Elisa. 2017. *Analogy in Word-formation: A study of English Neologisms and Occasionalisms*. Berlin: Walter de Gruyter.

Mattiello, Elisa. 2018. Paradigmatic morphology: Splinters, combining forms, and secreted affixes. *SKASE Journal of Theoretical Linguistics* 15(1). 2–23.

Mattiello, Elisa. 2023. *Transitional Morphology: Combining Forms in Modern English*. Cambridge: Cambridge University Press.

Miloš Jakubícek, Adam Kilgarriff, Vojtech Kovář, Pavel Rychlý & Vit Suchomel. 2013. The TenTen corpus family. In Andrew Hardie & Robbie Lowe (eds.), *7th International Corpus Linguistics Conference. Abstract Book*, 125–127. Lancaster, UCREL.

Niculescu-Gorpin, Anabella-Gloria & Monica Vasileanu. 2025. Romanian speakers' processing and understanding of lexical blends and blend-related formations. *Italian Journal of Linguistics* (in press).

Norde, Muriel & Sarah Sippach. 2019. Nerdalicious scientainment: A network analysis of English libfixes. *Word Structure* 12.3. 353–384.

Norde, Muriel & Graeme Trousdale. 2023. Issues in diachronic construction morphology. *Constructions and Frames* 15(2). 145–159.

Oxford English Dictionary Online. 2nd/3rd ed. Oxford: Oxford University Press. www.oed.com.

Plag, Ingo. 2018. *Word-formation in English*. 2nd edition. Cambridge University Press.

Renner, Vincent. 2018. Structural borrowing in word-formation: An exploratory overview. *SKASE Journal of Theoretical Linguistics*, 15(2). 2–12.

Renner, Vincent. 2019. French and English blends in contrast. *Languages in Contrast* 19(1). 27–47.

Ronneberger-Sibold, Elke. 2015. Word creation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe*, 485–500. Berlin & Boston: De Gruyter

Stamenov, Christo. 2015. Borrowing word-formation: *-ing* suffixation and blending in Bulgarian. *Съпоставително езикознание/Сопоставительное языкознание* [*Comparative Linguistics*], 3. 163–197.

Stoichițoiu-Ichim, Adriana. 2006. *Creativitate lexicală în româna actuală* [Lexical Creativity in Present-day Romanian]. București, Editura Universității din București.

Ten Hacken, Pius, & Renáta Panocová. 2020. Word formation, borrowing and their interaction. In Pius ten Hacken & Renáta Panocová (eds.), *The interaction of borrowing and word formation*, 3–14. Edinburgh: Edinburgh University Press, 2020.

Vaccarelli, Francesca. 2019. Miscele inglesi (e italiane) per formare nuove parole [English (and Italian) mixes for coining new words]. *Lingua Italiana Magazine Online di Istituto della Enciclopedia italiana fondata da Giovanni Treccani* [*Italian Language Online Magazine of the Institute of the Italian Encyclopedia founded by Giovanni Treccani*], doi: 10.7393/DLIOL-01.

Vasileanu, Monica & Anabella-Gloria Niculescu-Gorpin. 2022. Lexical blending in present-day Romanian: A corpus-based study. *Revue Roumaine de Linguistique* 57(4). 363–376.

Vasileanu, Monica, Anabella-Gloria Niculescu-Gorpin & Cristina-Andreea Radu-Bejenaru. 2024. Keep calm and carry on blending: Experimental insights into Romanian lexical blending. *Word Structure* 17(1–2). 55–89.

Warren, Beatrice. 1990. The importance of combining forms. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Contemporary Morphology*, 111–132. Berlin: Mouton de Gruyter.

Zwicky, Arnold & Geoffrey Pullum. 1987. Plain morphology and expressive morphology. *Annual Meeting of the Berkeley Linguistics Society*, 13, 330–340.

Zwicky, Arnold. 2010. Libfixes. *Arnold Zwicky's blog*, https://arnoldzwicky.org/2010/01/23/libfixes (last accessed 2 September 2023).

C Dynamics of Creativity and Routine in Synchrony

Maximilian Frankowsky & Barbara Schlücker

More than Just Formal Deviation

Creativity and Routine in Name-Based Lexical Patterns

Abstract: Being creative with language means doing something in a new way that has some advantage for conveying socio-pragmatic, contextual, or lexical information. The paper discusses two name-based lexical patterns in German that exemplify the interplay between creative language use and routine in word formation. Both patterns are instances of nominal compounding. The first consists of a proper name modifier and a common noun head (PN compounds, e.g. *Merkel-Besuch* 'Merkel visit'), the second of two identical proper names or common nouns forming proper names (Name-ICCs, e.g. *foto-foto* 'photo photo' [company name], *sabinesabine* [online forum user name]). We analyze data from historical and present-day German corpora to investigate the diachronic and synchronic development of the patterns. We argue that creative language does not necessarily have to be particularly artistic or eye-catching, but that the mere copying of form or even changes that only affect semantics or function can also be considered creative.

1 Introduction

One area in which human creativity is readily apparent is language. Especially when speakers are faced with the task of naming new things (or renaming old things), we can see how human creativity uses the language system to make the naming process successful. We show that linguistic creativity need not involve elaborate changes in form–it may also consist in purely semantic changes or even simple repetition.

This paper deals with name-based nominal compound patterns in German. These compounds have at least one proper name as a constituent. For example, the compound *Shetlandpony* ('Shetland pony') consists of the proper name *Shetland* as the first constituent and the common noun *Pony* as the head, referring to horses of a particular breed. Another example is *Kennedybrücke* ('Kennedy

bridge'), which is the name of a particular bridge named after President Kennedy.[1] There are also compound patterns where both constituents are proper names, e.g. *Anna-Lena* (first name) or *Nordrhein-Westfalen* (name of a federal state of Germany).

As can be seen from these examples, nominal compounds containing proper names might either be proper names themselves (e.g., *Anna-Lena*, *Kennedy-Brücke*) or common nouns (e.g., *Shetlandpony*). Compounds that are proper names may also consist of two identical proper names, e.g. *Baden-Baden* (name of a German city) or *sabinesabine* (online forum user name). Finally, there are also compounds that are proper names although the constituents are common nouns, both with different and identical constituents, e.g. *Feldberg* ('field mountain', a mountain name) and *foto-foto* ('photo photo', a company name). The latter type is thus rather name-related than name-based in the strict sense. Nevertheless, we will include formations with identical common noun constituents in our study, as they are particularly interesting with regard to the topic of the study.

The formation of name-based lexical patterns can provide valuable insights into understanding linguistic creativity. Being creative with language is a balancing act: speakers may want to try something new, but at the same time they want to be understood and achieve a communicative goal. The two case studies show how this can work, and that creative use of these patterns does not necessarily lead to new routines.

We first set out our understanding of creativity and routine (Section 2). We then present two case studies (on proper name compounds and compounds with identical constituents) that illustrate the interplay of creativity and routine in word formation (Section 3) and discuss them against the background of the theoretical assumptions on linguistic creativity (Section 4), followed by a summary of the findings (Section 5).

2 Creativity and Routine

Defining creativity is at least as difficult as defining language. Previous definitions of various linguistic areas have quite diverging conceptions of creativity. Some see creativity as an individual ability, while others regard it as emerging through interaction with interlocutors or as the result of social processes. Creativity is also treated alternatively either as a process or an outcome.

---

1 In addition, the proper name constituent may also appear as the compound head, e.g., *Helicopter-Cem* 'helicopter + Cem (first name of a German politician)', cf. Belosevic (2022).

When creativity is related to word formation, two conditions are usually assumed: first, something must be new, and second, it must be useful in some way (e.g., Stein 1953: 311; Sternberg and Lubart 1999: 3; Runco and Jeager 2012). This implies the act of foregrounding or defamiliarization: the modification of a pattern causes the newly created to stand out in some way, because something known has been made unknown through language. The defamiliarization realizes Jakobson's poetic function of language, since it leads to "a focus on the message for its own sake" (Jakobson 1960: 356).

In the following, we focus on the approach by Jones (2015). His definition of creativity includes four aspects: (1) The material from which people create things. (2) The cognitive processes that make linguistic creativity possible in the first place. (3) The social processes that are essential to linguistic creativity. (4) The result of the three aspects mentioned above, namely what has been created, such as a text, an action, or words. We will take a brief look at these four parts of this "creativity map" (Jones 2015: 1) and relate each to the domain of word formation.

First of all, there is the material to which creative processes are applied, the elements with which speakers can be creative, and the rules they can deviate from, that is, the routine. The term routine is understood as the set of rules that determine a pattern and is used to refer to the opposite of creativity. In terms of word formation, using a word that has come to life through routine means not being creative.

The second part of Jones' analysis concerns the cognitive processes that determine that speakers do not always follow the routine, but deviate from the rules from time to time. Speakers know the word formation routines and can modify them creatively. Schmid (2014) and Smith and Mackie (2000) suggest that this linguistic knowledge is ultimately based on association and that all linguistic structure is based on this general cognitive process. Linguistic knowledge is a network of more or less routinized associations. The strength of the associations is facilitated by routine, which in turn is facilitated by repeated processing. Speakers can choose whether to use a routinized word formation pattern (a choice for semantic transparency and compositionality) or to modify it (a choice for economy of expression since a new meaning has been added that would otherwise have to be expressed in a different way, see Körtvélyessy, Štekauer and Kačmár 2021: 1018). When a speaker forms new words based on a modified pattern, they are more difficult to interpret. Both the words and the underlying pattern are new and require a transfer effort on the part of the addressee. However, patterns are linked by associations, and deviations from routine are often understandable despite the modifications.

The third part is the social processes that, in addition to individual personality traits such as extroversion and intelligence (Hoffmann 2018; Kandler et al. 2016) give rise to linguistic creativity and the space in which creativity is applied. Social processes are a major reason why speakers deviate from the rules at all since this can

be advantageous. Extravagance and expressivity play an important role here. The term extravagance is linked to originality and creativity (Ishiyama 2014; Ivorra Ordines 2023) since it emphasizes the speaker's conscious attitude to be creative. Extravagance corresponds to Keller's (1994: 139) maxims to "talk in such a way that you are noticed" and to "talk in an amusing and funny way". Forming words in an extravagant way leads to expressivity (e.g. Hopper & Traugott 1993), a term that focuses on the linguistic forms themselves. Expressivity in turn serves sociopragmatic demands (Schmid 2020: 78). Sometimes speakers are successful when they play with the linguistic material (or imitate peers who do so), e.g. for social reasons such as negotiating identities and social relationships within a peer group (Maybin and Swann 2007: 512). Creativity is therefore always interactional, jointly constructed, and subject to historical and cultural references. In addition, according to Maybin and Swann (2007: 513), there is a critical-analytical dimension: creative language is necessarily evaluative and the creative episodes have the potential for deeper critique (developing shared critical standpoints and values, playing with authorities, etc.).

The fourth and final element of linguistic creativity is the outcome of the creative process, that is, what has been created. Here, however, we encounter problems that are not trivial to solve. The crux of Jones' descriptive approach is to define what must be present in order for the outcome of the whole process (a word formation, an interaction) to be considered creative.

Admittedly, regarding word formation, this way of defining linguistic creativity only shifts the problem. On the one hand, one can ask at what point something can be considered new, and on the other hand, one can ask what exactly is new.

Regarding the first question, it can simply be assumed that something is new if its frequency in the respective database is zero up to a certain point and is greater than zero afterward. In other words, something was not there before and is there now. Hohenhaus' (2005) approach goes in this direction. He considers a word formation to be new if it is actively formed by a speaker and not retrieved from the "storage of already existing listemes in the lexicon" (Hohenhaus 2005: 364).

Regarding the second question, either the words themselves or the patterns from which they emerge can be new. Hohenhaus assumes that if the words themselves are new, one can already speak of creativity. According to this view, novel formations are per se creative, even if the new word is based on a very frequent pattern. This is in line with an onomasiological point of view according to which speakers choose one expression out of many to respond to the need to name something and in this way achieve their communicative goals (Körtvélyessy, Štekauer and Kačmár 2021: 1022). Furthermore, Pennycook (2007) sees creativity not so much as a change of linguistic material, i.e. that a new form has been created, but that this form is placed in a different context. Since language is always multimodal, it can be assumed that, for example, a lexeme also contains information

about the contexts in which it is usually used. Here, "new" refers more generally to all aspects belonging to a linguistic unit. In the case of a creatively used word, this also includes aspects such as the situations in which it can be used or the addressees who have to interpret the word. Creativity, then, is simply "a new way of doing something" (Jones 2015: 5), and in terms of word formation, it concerns either the form of a word, its meaning, or its use.

However, we take a different approach in this regard. To us, word formation creativity does not mean that a new word has been created, but that a new pattern, i.e. a conventional pairing of form and meaning, has been created that has an advantage over existing patterns. Thus, we define word formation creativity as referring to the patterns themselves, not to the words. Creativity does not occur when a new word is coined, but when speakers begin to bend the rules or apply them in alternative ways.

So for instance, given the frequency and regularity of noun-noun compounding in German, a newly coined compound such as *Ampelmaut* ('traffic light toll', i.e. toll to be paid when stopping at a traffic light) would not be unusual or surprising, and may not even be perceived as new by the recipient. The underlying pattern is productive but not creative and rather satisfies the recipient's need for comprehensibility. Speakers have linguistic knowledge about the patterns they have encountered before and how often they are used. Therefore, it is not the routine formation of a word that is creative, but the deviation from the routine and the modification of existing patterns. In this way, the pattern itself must be new/unexpected and useful. In other words: speakers do not simply choose one expression among many to respond to the need to name something but they create an entirely new way of responding. If they are successful, i.e. if the new way they have formed a word is useful, other speakers may come up with novel words in a similar way and the innovations may spread.

An example of creative word formation that has led to a new pattern is the *i*-derivation in German. Although *i*-suffixation has existed since Old High German (cf. Henzen 1965: 143), it was long restricted to the function of coining hypocoristic personal names and kinship names, which also formed the bases for the suffixes (e.g., *Robi* as a short form of the first name *Robert*; *Mutti* as a short form for *Mutter* 'mother'). Fleischer (1982: 201) assumed that common noun bases (with a few exceptions such as *Ami* as a short form for *Amerikaner* 'American citizen') do not exist in the written language. Köpcke (2002), on the other hand, points to a more recent development, according to which *i*-suffixes also combine with common nouns as well as adjectival and verbal bases. These formations also have a hypocoristic or joking meaning and the pattern is now very productive (cf. also Fleischer and Barz 2012: 214), e.g. *Ersti* ('freshman', from *Erstsemester* 'first semester'), *Blödi* (friendly scolding for a stupid person; from the adjective *blöd* 'stupid').

Thus, in the case of the *i*-derivation, the original pattern was modified and became much more productive since.

Our definition of word formation creativity follows the approach of Munat (2015: 93), Lieber (2010), and others, who clearly contrast word formation creativity with productivity. Word formation creativity is the ability to intentionally coin a new word on a new/modified pattern. Whereas productivity is a rule-governed routine, creativity is a deviation from the routine. Sometimes creativity is regarded as an extragrammatical process and creative morphology accordingly encompasses phenomena such as blends, acronyms, reduplicatives (Mattiello 2013). For Ronneberger-Sibold (2012: 16), language creativity is "not only the coining of entirely new words not based on any previously existing linguistic elements but includes all operations for the production of new lexemes which are not covered by regular word formation". Similarly, according to Bauer (1983: 63), productivity is the formation of new words by the application of the rules of grammar, while creativity is "the native speaker's ability to extend the language system in a motivated but unpredictable (non-rule-governed or irregular) way".

According to Bauer (2001: 64) this distinction between creativity and productivity in language is gradual because creativity is also rule-related. When speakers use a lexicalized word, they act more routinely and less creatively than when they form new words on the basis of a highly productive word formation pattern, such as noun-noun compounding in German, which in turn is less creative than when they form words on the basis of a less productive pattern such as blending (Ronneberger-Sibold 2012). Language creativity is thus "a graded phenomenon ranging from the more conventional and predictable to the less conventional and unpredictable" (Kecskes 2016: 3).

To sum up, the creativity of a word formation pattern is determined by the properties of that pattern and the degree of awareness with which the corresponding expressions are formed. As in the definitions above, creative use of language is intentional and fulfills an expressive function so that the speaker stages the expression for the audience to recognize it.

Creativity in word formation is worth examining empirically. According to Langlotz (2015), people's creative potential is directly reflected in their language structure and use. There is empirical evidence that the general tendency towards routine correlates negatively with the creative strengths of language participants and that this is particularly true for word formation tasks (Körtvélyessy, Štekauer

and Kačmár 2021).[2] It is therefore interesting to see how speakers overcome the routine, modify familiar patterns and use them in unexpected, creative ways as well as to see which innovations are here to stay and which are transitory. Schmid (2014: 246) compares this spread of linguistic innovation to the spread of a disease or a virus. The question of whether modifications lead to the development of new routines is therefore an inherently diachronic one.

In the following, we discuss the diachronic development of name-based and name-related compound patterns in German. The first study deals with modifications in the period from 1600 to 1900 while the second study examines modifications that are currently underway. We show that modifications are not limited to the formal properties but that they also include semantic-functional properties. In addition to the corpus data that form the basis of our study, we also use metalinguistic information of various kinds, where accessible, to assess the social and expressive impact of the modifications.

3 Case Studies

3.1 Nominal Proper Name Compounds

The first pattern to be discussed is nominal compounding with a proper name as the left constituent and a common noun as the head, here referred to as proper name (henceforth PN) compounding, e.g. *Nilfahrt* ('Nile ride'), *Kosovo-Konflikt* ('Kosovo conflict'). The pattern also includes formations with phrasal names such as *Louis-Armstrong-Lied* ('Louis Armstrong song'). In this case study, we focus on the period from 1600 to 1900. This period shows a clear increase in frequency of PN compounds which can be associated with semantic and functional modifications of the pattern, through which the range of possible uses becomes larger. At the same time, extravagant effects also seem to play a role in this. As mentioned in Section 1, PN compounds are either proper names themselves (onymic type; e.g. *Kennedybrücke* 'Kennedy bridge') or common nouns (appellative type; e.g. *Shetlandpony* 'Shetland pony'). This functional dichotomy is crucial for the question of creativity and the emergence of new routines.

---

2 The authors use a standardized method to determine a person's creative strength (*Torrance Test of Creative Thinking*) and investigate the extent to which this is related to the word formation behavior of the test subjects. As a result, creative potential, i.e. the ability to develop novel and useful ideas (divergent thinking ability), appears to be related to the way in which individual language users coin new words.

Onymic PN compounds have already been used since Old High German, in particular place names, e.g. *Neckarburg* ('Neckar castle', name of an early medieval castle near the river Neckar), cf. Bach (1953), for instance. Thus, the formal pattern of PN compounding is without doubt a very old and native lexical pattern of German. At the same time, appellative PN compounds are quite rare. A few examples of appellative PN compounds are given in the Middle High German grammar by Klein, Solms, and Wegera (2009), e.g. *kipperwīn* ('Cyprus wine'), *rōmvart* ('Rome trip'). (Note that even though these compounds are not place names, they also have a local meaning.) In contrast, in present-day language, appellative PN compounding is a very productive pattern (cf. Kosmata in prep.) and is not restricted to place reference and location, e.g. *Merkel-Besuch* ('Merkel visit'), *Bunsenbrenner* ('Bunsen burner').

The study by Schlücker (2020) indicates that the period 1600–1900 is crucial both for the increase in frequency of the pattern as well as for its functional expansion. More specifically, it can be argued that these two changes are interrelated. The study is based on data from the DTA (*Deutsches Textarchiv*) corpus. A total of 15,765 tokens (3,783 types) were collected.[3] Of these, 6,162 tokens were annotated and serve as the basis for the study.[4] The period investigated is divided into six time periods of 50 years each. Table 1 shows a continuous increase in type and token frequency, especially a rapid increase in the second half of the 19th century. Even if some reservations are in order because the basic size of the corpus is not totally balanced and because of the sampling procedure (so that it does not make sense, for example, to calculate a type-token ratio; see Schlücker (2020: 244) for more details), these numbers nevertheless show that the pattern of PN compounding is continually being expanded by new formations.

A closer look at the data reveals that both the onymic type and the appellative type occur in each period and that both are increasing in frequency. However, there is a clear shift in the relative frequency of occurrence: while the appellative type accounts for only 26.3% of the types and 37.7% of the tokens in the

---

[3] As in other corpora, compounds cannot be searched for automatically in the DTA corpus. Hundreds of individual queries were carried out to search for strings with a set of proper names at the beginning or appellatives at the end (e.g. *Rhein* 'Rhine' + X, X + *Brücke* 'bridge'). The results were cleaned up manually. The initial list of proper names and appellatives to be searched consisted of items that had been used in previous studies with other corpora (cf. Schlücker 2017). During the search, new proper name modifiers and appellative heads were constantly added to the queries. In total, the data comprises 1,746 proper name modifier and 705 appellative head types.

[4] For reasons of time and capacity, the annotation of all tokens was impossible. Therefore, three tokens (if available) were annotated for each type per time period, i.e. a maximum of 18 tokens, amounting to a total of 6,162 tokens.

Table 1: Total development of PN compounds (from Schlücker 2020: 246); corrected numbers of types. The last period covers one more year.

		Types	Tokens
17th c.	1600–1649	76	130
	1650–1699	219	440
18th c.	1700–1749	316	553
	1750–1799	505	763
19th c.	1800–1849	627	912
	1850–1900	2,229	3,364
total		3,972	6,162

first period (1600–1649), it is already 51.5% and 48.0% of the types (51.1% and 48.9% of the tokens) in the last two periods (1800–1849; 1850–1900).

In the following, we discuss the idea that the observed increase in frequency can be explained by modifications of the pattern, through which the range of application expands and the forms are used more frequently. However, this is less about formal than about semantic-functional changes. Specifically, three modification processes can be observed:

– Modifications of the name class of the compound (onymic type)
– Modifications of the name class of the PN constituent (appellative type)
– Modifications of the semantic relations between the compound constituents (appellative type)

The first group of modifications concerns the onymic type and relates to the name class of the entire compound. As mentioned above, the oldest attestations of PN compounds are essentially place names. This is also true for the initial period of this study: in the first period, 93% of the onymic PN compounds are place names, whereas personal names cover 5% and object names 2%. In contrast, in the final period, the proportion of place names decreases to 78% while object names rise to 19% (plus 1% of personal names), cf. Figure 1.

This is a functional extension of the pattern of onymic PN compounding, as it is increasingly used not only for naming places but also for naming artefacts of all kinds, such as *Rheinarmee* 'Rhine army', *Juliusspital* 'Julius hospital', *Friedrichsuniversität* 'Friedrich University', *Ludwigslied* 'song of Ludwig'. In addition, there is also a functional extension within the group of place names. During the entire period under review, the most frequent subgroup of place names are water names, in
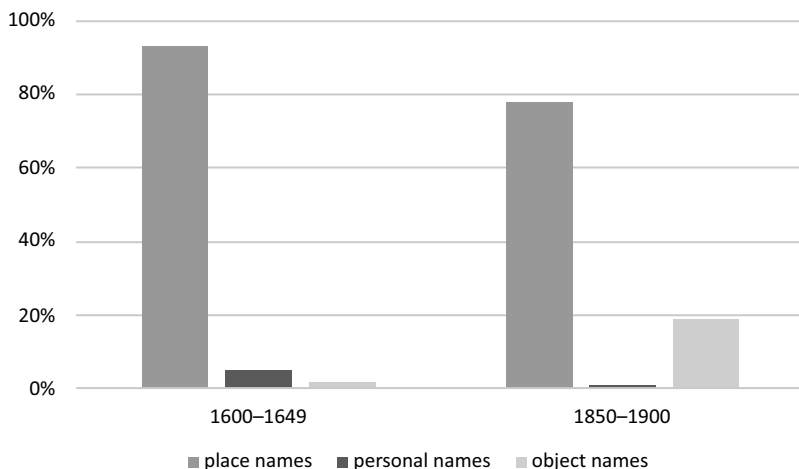
Figure 1: Proportion of name classes of onymic PN compounds.

particular river names such as *Rheinstrom* 'Rhine stream', *Oderfluss* 'river Oder'. However, the proportion of water names decreases from over 50% in the first period to below 30% in the last two periods, while at the same time the proportions of other types of place names such as names of states, countries and landscapes (e.g. *Ruhrgebiet* 'Ruhr area'), mountain names (*Humboldtgletscher* 'Humboldt glacier') and street names (*Beethovenstraße* 'Beethoven street') clearly increase. This demonstrates the functional expansion that the pattern of PN compounding undergoes between 1600 and 1900 and which forms an obvious (partial) explanation for the general increase in frequency (cf. Schlücker 2020: 247–248).

The second group of modifications concerns the name class of the onymic modifier in the appellative type. At first glance, there are no major changes regarding these modifiers. They are essentially personal, object, and place names, with the last group having the largest proportion overall (with some variation) and its proportion increasing diachronically. There is, however, a small but crucial modification within the group of personal names. Overall, personal names in the modifier position in PN compounds between 1600 and 1900 are almost exclusively first names, mostly very old and often biblical names, names of historical personalities or saints' names, e.g. *Sophien-Tempel* 'temple of Sophia', *Jakobsleiter* 'Jacob's ladder', *St. Thomastag* 'St. Thomas' day'. Family names, in contrast, do not occur at all or only in very small numbers (< 1%). However, in the last period (1850–1900)

their proportion rises to 7%; examples are *Goethegedicht* 'Goethe poem' or *Morseapparat* 'Morse telegraph'. This is particularly interesting because in present-day German, family names form the largest part of personal names as modifiers in PN compounds. Schlücker (2017: 71) compares the proportion of first names in different present-day language corpora and finds that it accounts for 77% to 89% of all personal name modifiers, depending on the corpus, whereas the proportion of first names is only 2% to 9%. In view of the high productivity of the pattern in the 20th century and the large proportion of family name modifiers (18% of all appellative PN compounds, cf. Kosmata in prep.), it can be concluded that the novel inclusion of family names as modifiers from 1850 onwards results in a considerable functional expansion of the pattern and thus leads to an increase in productivity. Furthermore, this development is presumably closely related to the third group of modifications and probably made them possible in the first place.

The third group of modifications concerns the semantic relations in appellative PN compounds. These changes are particularly interesting since they modify the pattern in such a way that it becomes a potential competitor construction to syntactic constructions. It is generally assumed that the constituents of noun-noun compounds are related by implicit semantic relations. They are part of the compound meaning and have either to be learned or inferred from the use of the compound in context. Even if the number of relations is not limited in principle, there seem to be about ten to twelve core relations, such as LOCAL (*garden bench*), HAVE (*associations assets*), PART-OF (*string instrument*), MAKE (*honey bee*), ABOUT (*animal book*) (cf. e.g., Levi 1978, Fanselow 1981, among many others).

Semantic relations are also found in PN compounds. However, the semantic relations that are most frequent in PN compounds are not identical to the most frequent relations in noun-noun compounds (i.e. those with an appellative modifier) and there are relations that occur exclusively in PN compounds (cf. Schlücker 2017), such as the commemorative relation (cf. Warren 1978; Koptjevskaja-Tamm 2009), where something is named after a founder, inventor, discoverer or in honor of a person (*Kennedybrücke* 'Kennedy bridge'). The commemorative relation is still rather rare in the DTA data, though (examples include *Röntgenstrahlen* 'X-rays', lit. Röntgen rays, *Morseapparat* 'Morse telegraph', *Bunsenbrenner* 'Bunsen burner'), and apparently is more characteristic of the pattern in the 20th century. However, this relation co-occurs almost exclusively with family name modifiers. Thus, the two are interrelated and the modifications of name class and semantic relation lead to a functional expansion and – especially in the 20th century – to a clear increase in frequency.

In addition, there is also a clear increase in synthetic PN compounds during the period under investigation (cf. Schlücker 2020). Synthetic compounds have deverbal heads and the first constituent is an internal or external argument of the underlying verb.[5] The proportion of synthetic PN compounds increases clearly in the data, from about 5% in the first period to over 10% in the last. Also frequent are PN compounds with relational heads, such as *Schwedenkönig* 'king of Sweden'. This is important because these argumental relations belong to the semantic relations that can also be realized by genitive constructions. For this reason, these compounds can function as substitutes for genitive constructions (provided that the name is used referentially), e.g. *Homerkritiker – Kritiker Homers* 'Homer's critic', *Schillerverehrer – Schillers Verehrer* 'Schiller's admirer', *Nilüberschwemmung – Überschwemmung des Nils* 'flooding of the Nile', *Schwedenkönig – Schwedens König* 'Sweden's king', *Nettelbeck-Bild – Bild Nettelbecks* 'Nettelbeck portrait', *Rösel-Freund – Rösels Freund* 'Rösel friend' (cf. Schlücker 2018). These (as well as other kinds of) PN compounds have therefore been regarded as competitor constructions to the genitive. They often have family name modifiers, so it can be assumed that the increase of family name modifiers is also related to the increase of compounds with argumental relations, in addition to the commemorative relation.

All of the modifications and innovations of the 1600–1900 period discussed here are regularly found in present-day language PN compounds, so creativity has become routine. For instance, recent examples such as *Gorman-Gedicht* 'Gorman poem', *Merkel-Besuch* 'Merkel visit' (where Merkel can be interpreted as either the internal or external argument) or *Selenskyj-Rede* 'Selenskyj speech', whether lexicalized or not, not only point to the continuing productivity of the pattern but also illustrate its potential use as a substitute construction.

Finally, the novelty of the changes can be demonstrated not only on the basis of the corpus data but also on the basis of reactions to this language use which reflect their salience and extravagant nature (cf. Neels, Hartmann & Ungerer 2023). Two grammarians of that time, Wustmann (1891) and Briegleb (1928, 1932), have commented extensively on these constructions. First of all, both note the increase of PN compounds in language use. Whereas Briegleb refers to PN compounds in general, Wustmann explicitly mentions PN compounds with personal names. PN compounds and their increasing use are clearly evaluated negatively: "die jetzt immer massenhafter auftretenden Zusammensetzungen mit Personennamen" ['the compounds with personal names that are now appearing more and more en masse'] (Wustmann

---

5 So strictly speaking, they do not have semantic relations in the narrow sense, since the relation between the constituents is not freely inferred, but rather determined by the argument structure of the deverbal head.

1891: 204) and "(. . .) die heutige Wucherung von Zusammensetzungen mit Eigenna-men. Ein Kleist-Denkmal, das Kant-Grab, Shaw-Worte, der Matteotti-Mörder [. . .]. Alles verfallene Sprache" ['(. . .) today's proliferation of compounds with proper names. A Kleist monument, the Kant tomb, Shaw words, the Matteotti murderer [. . .]. All decayed language.'] (Briegleb 1928: 20).

In all cases, the criticism refers to the fact that compounds are used instead of genitives, or, in some cases, instead of constructions with prepositional phrases (e.g. *Italienreise – Reise nach Italien* 'journey to Italy'). In the case of place names, also the use of PN compounds instead of constructions with name-based adjectives is criticized (e.g. *Italienreise – italienische Reise* 'Italian journey') (cf. Wustmann 1891: 202; Briegleb 1932: 27–29). PN compounds are labeled as "falsche Zusammen-setzung" ['wrong compounding'] (cf. Briegleb 1932: 29), among other things, and their use is explained by laziness and slackness of the speakers, since, according to Briegleb (1932: 28), the genitive requires effort of linguistic formation, but com-pounding does almost not. Finally, Wustmann (1891: 206) also explicitly discusses synthetic PN compounds. While he assumes that other kinds of PN compounds might be possible in some cases, he categorically rules this out for synthetic com-pounds and calls them the "Gipfel der Sinnlosigkeit" ['pinnacle of futility'].

Thus, the second and third modification processes described above can be re-garded as creative, since they modify the original pattern and have extravagant effects: the comments on the increase of PN compounding in general and on the 'competitor construction' in particular point to the novelty and salience of the ap-pellative pattern at that time. With regard to the various characteristics of extra-vagance that have been proposed in the literature (cf. Ungerer & Hartmann 2020), it can therefore be stated that the pattern discussed here deviates from linguistic norms and expectations and that this deviation creates attention. Using this pat-tern thus obeys the maxim of extravagance as formulated by Haspelmath (1999: 1055): "Talk in such a way that you are noticed". However, nothing can be said about other possible characteristics of extravagance, such as the emotional in-volvement of the speaker (cf. Petré 2016, 2017).

Obviously, this language use has been socially successful, since in the 20th century until now we find a consolidation of the changes and the development of a routine (cf. Schmid 2020 on feedback-loop processes in conventionalization and entrenchment). This transition from creativity to routine in the course of dia-chronic development is also reflected in a further, albeit weak, increase in pro-ductivity in the 20th century (cf. Kosmata in prep).

The first group, the modifications regarding the name class of the onymic PN compounds, is a bit different. So far, there is no explicit evidence that these changes were perceived as particularly extravagant and/or expressive. Neverthe-

less, these changes can also be classified as creative, because they offer a new way of referring to extra-linguistic entities, and we also find here the development of a routine.

3.2 Name-ICCs

The second case study concerns compounds with identical constituents (identical constituent compounds, henceforth: ICCs, Finkbeiner 2014; Hohenhaus 2004). For a long time, ICCs were only mentioned in linguistic texts to proclaim their impossibility, i.e., to define constraints on compounding in German. For example, Kürschner (1974) states: "The following conditions apply [to nominal compounds]: (i.) A≠B; i.e., different lexicon entries must be chosen as A and B in order to exclude forms such as *garden-garden, *pencil-pencil" (Kürschner 1974: 148, our translation). Erben also states that "for compounds in German, the rule of combining d i s s i m i l a r morphemes or lexemes normally applies" (Erben 1981: 39, our translation, emphasis in original). The assumption that there are no compounds with identical constituents in German may be the reason why German linguistics has paid so little attention to ICCs for so long.

Contrary to this widespread assumption, however, there are ICCs in German. First, there are determinative ICCs, which are ultimately just 'normal' noun-noun compounds with identical constituents, but which otherwise have no special properties (Donalies 2011: 72; Fleischer & Barz 2012: 96). Examples of this ICC type are *Kindeskind* 'grandchild' or *Helfershelfer* 'accomplice'. Both compounds have the linking element *-(e)s-*, which is common in German noun-noun compounds, and the semantic relation OF between the constituents ('a child **of** a child', 'the helper **of** a helper'). Second, there are ICCs that are not determinative and have no modifier-head relation such as *Film-Film* 'blockbuster, no documentation' or *Mannmann* 'prototypical man'. Like determinative ICCs, they denote sub-concepts, but they do so by restricting the given concept to its prototypical core. A *Film-Film* is a 'real movie', i.e. the first thing that comes to mind when thinking about movies; a *Mannmann* is a 'real man' who has all the stereotypical characteristics of a man, i.e. is physically strong, dominant etc. (Bross and Fraser 2020; Finkbeiner 2014; Freywald 2015; Frankowsky 2022, 2024; Günther 1979; Kentner 2017).

However, these two ICC types are not the subject of this paper. Rather, the focus here is on a third ICC type for which there is hardly any literature to date, namely ICCs that are proper names (henceforth: Name-ICCs), e.g. *Dior-Dior* (referring to a perfume of the Parisian luxury goods manufacturer Dior). Similar to compounds like *Anna-Lena*, they are name-based compounds, except that the constituents are identical.

So far, Name-ICCs have only been mentioned in completely different contexts: in business linguistics,[6] Name-ICCs have been discussed in connection with word formation patterns of brand naming. Platen (2013), a compilation of brand naming techniques, refers to noun-noun compounds with identical constituents as "Echowörter" ['echo words'] with a monoreferential meaning (Platen 2013: 53). He considers the function of Name-ICCs like *Dior-Dior* to lie primarily in the expressivity that arises in such words. He classifies them under the category "proper name formations" and sees such formations as a variant of emphatic repetition and affective reduplication that is used specifically in product naming (Platen 2013: 54). In addition, Kauffman (2015: 4), in a linguistic typological review paper, discusses Name-ICCs that consist of spatial proper names and are, in turn, place names, e.g. *Baden-Baden* (German city name). Kauffman also mentions "name doubling" and recognizes a social function in such formations, which is certainly also due to the fact that nicknames, especially in English, are often doubled, e.g. *Jon-Jon*, *Lou-Lou* (Kauffman 2015: 3).

Kentner (2017) investigates a similar phenomenon in German: speakers create pseudonyms in the form of ICCs with a personal name as a base constituent, e.g., *Sabinesabine* (Kentner 2017: 244). The base is doubled for formal extension in the course of name formation (cf. also Dürscheid 2005: 48). With the help of acceptability tests, Kentner's study provides evidence that rhyme reduplication (*Sabinepabine*) is preferred in the formation of proper names that are themselves based on proper names. Total reduplication, i.e., the use of Name-ICCs (*Sabinesabine*) is rather disfavored (Kentner 2017: 252–253).

Name-ICCs such as *Sabinesabine* belong to a certain subgroup of Name-ICCs, which we refer to here as Name-ICC type 1, namely Name-ICCs with proper name constituents. However, there are other patterns that can also be regarded as Name-ICCs that have been completely overlooked in previous research: Name-ICCs with common noun instead of proper name constituents, e.g. *AutoAuto* 'car car' or *Katze Katze* 'cat cat'. Neither the formal nor the semantic-functional properties of this type have been described so far. In particular, it is unclear whether these formations still have descriptive meaning parts or are purely referential, as is generally assumed for proper names.

In order to answer these questions and to provide a comprehensive description of the name-related lexical pattern "Name-ICC", a large-scale corpus study was conducted. We used the DECOW16 corpus (Schäfer and Bildhauer 2012),

---

6 The discipline of business linguistics, which developed at the beginning of the 20th century (partly in the context of the Prague School), analyzes business language as a means of communication. However, business linguistics did not survive as a separate discipline.

which consists of about 20 billion tokens and contains written texts from the Internet. This corpus contains unedited texts from computer-mediated communication and interaction-oriented writing, among other things, such as blog posts, chats, and forum discussions, which straddle the line between orality and literalness. For this reason, this corpus, despite the written nature of the media, also contains uncontrolled communication, which is less inhibiting to innovative linguistic techniques than, for example, the corpora of printed newspapers (Albert 2013: 163). A lexeme-based approach was used to search for Name-ICCs. 1,034 lexemes were selected based on the World Loanword Database (WOLD), a database that provides basic concepts and vocabularies (Haspelmath and Tadmor 2009). Out of WOLD's 1,814 basic concepts, 1,034 are expressed by nouns in German. These lexemes were chosen for the query.[7]

The data analysis has three main results. First, Name-ICCs are not uncommon. The corpus query yielded 2,566 tokens of Name-ICCs. Considering that the search was based on a limited set of base nouns, this is a remarkable result. Name-ICCs were found for 277 of the 1,034 base nouns, i.e. for more than a quarter of the nominal concepts from the WOLD. By comparison, appellative ICCs are found for only 96 of the 1,034 base lexemes and only 838 tokens in total (Frankowsky 2024). Name ICCs are therefore the most common ICC type. Second, there are three types of Name-ICCs. Third, Name-ICCs generally differ from noun-noun compounds in terms of word structure, formal features, and semantics. These last two results need to be looked at in detail.

The three Name-ICC types can be distinguished with recourse to the word class as well as the semantics of the constituents and the ICC. Besides Name-ICC type 1 which has proper names as constituents (e.g. *DiorDior* or *Sabinesabine*), there is Name-ICC type 2 which has common noun constituents. Examples of this would be *AutoAuto* 'car car', referring to a German entertainment show, and *fotofoto* 'photo photo', referring to a company. These Name-ICCs can be regarded as a special type of onymization. In contrast to Name-ICCs like *Sabinesabine* and just like the name *Feldberg* 'field mountain' (name of a mountain in southern Germany) in these Name-ICCs neither the first nor the second constituent is a proper name, but the formation still results in a proper name. Unlike proper names such as *Feldberg*, however, the constituents are identical. Moreover, only one constituent of a Name-ICC provides descriptive meaning. This descriptive meaning is used to indicate what the referents are dealing with. The names are thus partially motivated: *AutoAuto* 'car car', refers to an entertainment show that has to do with a

---

7 The different word forms for each lexeme were searched for by using wildcards ([x*x*] / [x*_x*]) so that separately written ICCs are included in the data.

car (which is destroyed during the show), *foto-foto* 'photo photo' refers to a company that produces photo products. Another difference to compounds such as *Feldberg* is that there is no semantic relation between the two constituents.

The Name-ICC type 3 works slightly differently. Examples are *Löwelöwe* 'lion lion', the name of an Austrian rock band, and *Katze Katze* 'cat cat', the name of a cocktail bar in Graz. During word formation, the semantic features of the common noun constituents get lost; Name-ICCs retain only referential meaning. The semantic features of the base noun are completely irrelevant for the reference of the formation: *Löwelöwe* does not refer to a lion but to a rock band, *Katze Katze* does not refer to a cat but to a bar. The referents of these Name-ICCs have nothing to do with lions or cats. Thus, type 3 Name-ICCs are arbitrarily related to the objects to which they refer. The partial motivation of type 2 Name-ICCs and the fact that this functional aspect is not present in type 3 Name-ICCs such as *Katze Katze* justifies the assumption of two different Name-ICC types. Figure 2 shows the basic properties and examples of all three types of Name-ICCs.

Figure 2: Properties and examples of the three Name-ICC types.

All three Name-ICC types deviate from noun-noun compounds in general in that they violate the alleged dissimilarity constraint of German compounding (Erben 1981; Kürschner 1974). As immediate repetitions of form, they are extremely unusual formations for German, a language that is particularly averse to reduplication from a linguistic typological perspective ("reduplication avoider", Freywald 2015: 905; Lohde 2006: 43; Stolz 2006: 115; Stolz, Stroh and Urdze 2011: 565). In lay linguistic publications on German, repetition is generally viewed very negatively since it is equated with redundancy (Sick 2009: 26). It can therefore be assumed that reduplicative structures generally attract attention, if not rejection. Stolz, Stroh, and Urdze (2011: 5) call this an "anti-T[otal] R[eduplication] attitude". The fact that Name-ICCs nevertheless exist suggests that it is precisely this extravagant, eye-catching structure through which speakers achieve their communicative goal. The use of a Name-ICC leads to expressivity and can be a very effective way of naming. With Name-ICCs, the speakers achieve what Bauer (2000) assumes for all playful and innovative word formation products: "[P]layful formations [. . .] and some literary creations [. . .] may go beyond the bounds of normal rules specifically to gain effect" (Bauer 2000: 838). If speakers use this conspicuous, extravagant expression, it may be because of the extravagant effects it conveys. In other words: speakers of German cannot repeat the first constituent of a compound without attracting attention. The message itself, i.e. the name of the product or user, is the focus.

Name-ICCs deviate from canonical compounds not only regarding their overall structure but also regarding other formal features. They almost invariably show zero inflection – a property characteristic of names in general (Nowak and Nübling 2017; Nübling 2012). Furthermore, linking elements that are otherwise obligatory in noun-noun compounds, such as *-n-* after schwa, are missing in most Name-ICCs (*Löwenfell* 'lion skin' – *Löwe∅löwe, Katzenhaar* 'car hair' – *Katze∅Katze*). Although they are not completely excluded in Name-ICCs, they are much less frequent. So while according to Nübling and Szczepaniak (2009) about 35% of canonical N+N compounds have linking elements, they are only found in 17% of Name-ICCs, e.g. *Mausemaus* [online forum user name]. In addition, Name-ICCs differ in spelling. Name-ICCs have a higher proportion of internal capitalization, hyphenation, and space. They are spelled regularly in less than a third of the cases. Another feature of Name-ICCs is the frequent use of quotation marks. In DECOW16, only 0.7% of all nouns are enclosed in double quotation marks, but almost 11% of the Name-ICCs. In many cases, this is due to the fact that they are proper names or actual quotations. However, the use of double quotes may also reflect that the writers are aware of the status of Name-ICCs as deviant formations ("distancing quotation marks", e.g. Klockow 1980, Meibauer 2007). Another indication of this is that speakers

sometimes use metalinguistic comments to distance themselves from the use of a Name-ICC. These comments cannot be easily quantified, but some of them are given here:

(1) *ich bezweifel es stark.* **Biber Biber** *hört sich komisch an*
     'I strongly doubt it. **Beaver Beaver** sounds strange'

(2) *Sie hatten halt so tolle Namen wie: Mipsmops,* **Melone Melone** *. . . naja Kinderkacke halt*
     'They had just so great names like: **Melon Melon** . . . well, it's just child poop [=nonsense]'

(3) *oder über den charmanten Humor von* "**Mädchen Mädchen**". *Blöder Titel –*
     aber kein blöder Film. 'or about the charming humour of "**girl girl**". Stupid title – but not a stupid film'

The fact that *Biber Biber* is called strange here, the use of *Melone Melone* discredited as nonsense and *Mädchen Mädchen* as stupid shows that Name-ICCs deviate greatly from the routine and can therefore not go uncommented. In summary, although Name-ICCs are structurally and formally divergent, they are regularly used by speakers.

4 Discussion

We have presented two case studies of word formation patterns associated with nominal compounds, namely Name-ICCs and PN compounds. Both word formation patterns are used creatively in the sense that speakers modify existing patterns to achieve their communicative goals. Furthermore, the two case studies show that the criteria for defining word formation creativity must be extended.

Previous definitions from the literature apply well to the phenomenon discussed in the second case study. The modifications of the noun-noun compound pattern that speakers make to form Name-ICCs are formal in nature. Name-ICCs have an unusual reduplicative structure and are characterized by an underuse of linking elements which are mandatory otherwise, by zero inflection, and a spelling that deviates from the norm. Name-ICCs thus exemplify extensive modifications of a pattern, and one might ask whether such an extreme word formation pattern can be successful at all. In fact, the formation of Name-ICCs is generally not a very common word formation process in German. However, it is quite successful in some

domains. On the one hand, Name-ICCs often occur in marketing contexts (Platen 2013). By using the Name-ICC pattern, names for shows, perfumes, music bands, and other products are easy to create, and the repetition of a proper name or a common noun which has descriptive semantics helps to remember them easily. This mnemonic effect may be the reason why Name-ICCs are used as brand naming devices. The second domain where Name-ICCs can be found is self-naming in social media (Kauffmann 2015; Kersten and Lotze 2022). Here, speakers follow the motto: 'Take your name or a word that describes the topic you are interested in, double it, and you have a good nickname.'

This domain specificity shows once again that information from the situational context must be taken into account when processing ICCs (Frankowsky 2022, 2024). Whether a Name-ICC is used or not depends to some extent on the contextual contingency of an utterance and on "pragmatic associations" (Schmid 2020). Pragmatic associations contain contextual information and transfer information from the current perceptual input to higher levels of processing that mediate between perception and the processing of contextual and functional aspects of meaning. Conversely, the interpretation and evaluation of Name-ICCs can also be viewed as being accessed through pragmatic associations.

As mentioned before, Kentner (2017) provides experimental evidence that Name-ICCs are used reluctantly when it comes to naming persons. He collected acceptability data on pseudonym/nickname formation, showing that Name-ICCs are strongly disfavored over rhyming and ablaut reduplication. Thus, speakers are much more likely to accept names like *Sabinepabine* (rhyme reduplication) or *Mipsmops* (ablaut reduplication) as nicknames/pseudonyms than total reduplication as in Name-ICCs (*Sabinesabine, Mopsmops* 'pug pug'). Importantly, however, Kentner's study did not ask about grammaticality or correctness, but rather "participants were asked to judge the items in a playful context" (Kentner 2017: 252). According to Kentner, it can be assumed that Name-ICCs are rejected by speakers. However, the study investigates the acceptability of reduplicative names in an experimental setting. Speakers may reject the formations when explicitly asked, but they still use them actively, as shown by our data. The extensive formal deviations observed in Name-ICCs attest the balancing act between violation and comprehensibility: Name-ICCs are clearly a tool to attract attention. When the deviation from the norm is extensive, speakers are aware of the deviation. The pattern can still be successful, but the number of domains in which it can be applied may be reduced and the need for legitimation strategies increases. Speakers recognize the deviation of the pattern from the routine, which can be seen from the metalinguistic comments in the data. This highlights the extravagance of the construction, signalling that the speaker is aware of the deviation, which in turn protects the speaker's face and allows them to use the construction. This illustrates that

linguistic creativity is not limited to the speaker alone, but depends on social processes and interaction with interlocutors.

The question of whether Name-ICCs will be used for proper name formation in the long run cannot be answered at this time. On a formal level, Name-ICCs are related to determinative ICCs such as *Kindeskind* 'grandchild' as well as ICCs with a prototypical meaning such as *Mannmann* 'prototypical man', which might increase the risk of misunderstanding. However, the formal properties of Name-ICCs may distinguish them at least from the determinative ICC pattern: Name-ICCs do not inflect, often lack obligatory linking elements and are sometimes written with a space.

Moreover, due to the extravagant effect created by the identity of the constituents, Name-ICCs help speakers to succeed in the field of product marketing and self-naming on the Internet. Thus, Name-ICCs can be considered useful in situations where it is important to stand out. All this makes Name-ICCs a prime example of creative word formation. However, the aforementioned extravagant effect may wear off over time, partly because this effect is used by another pattern currently on the rise, namely ICCs with a prototypical meaning (Frankowsky 2022: 171), but also because using Name-ICCs in naming processes might become boring and outdated in the long run if more things are named this way.

In contrast to Name-ICCs, the modifications to the PN compound pattern have long been consolidated and become routine. The pattern modifications discussed for PN compounds are not form-related, but rather semantic-functional. We have discussed modifications regarding two semantic-functional properties, name class, and semantic relation. PN compounding was originally a pattern used for proper names only, in particular place names, and its use as a common noun developed only later. Within the onymic type, the proportion of object names has increased diachronically, at the expense of place names. This is an important modification, since this shift means that a wider range of entities can be named by using this pattern. This is an obvious explanation why the pattern is increasingly used. However, while the onymic pattern undergoes a functional expansion, there is no evidence that the modification of the pattern realizes an expressive function or that the speakers are particularly aware of the change. Within the appellative type, a completely new name class has been observed, i.e. family names, which had not been used in the modifier position of this pattern before. Again, this means extending the scope of possible uses, especially since they were presumably closely related to the emergence of (among others) two types of semantic relations, i.e. commemorative and argumental relations.

As the comments of the grammarians of the time show, the modifications of the appellative type have been noticed and evaluated negatively, which hints at a deliberate use of the pattern. At the same time, the extravagant effect obviously outweighs the norm violations caused by the modifications. Gradually, the modi-

fications become permanent, the extravagance diminishes and finally disappears, and a new routine emerges which exists undoubtedly in present-day German ("habitualization"/ "routinization", Schmid 2020).

Regarding the answer to the question of what determines the creativity of a pattern, we argue that not only formal modifications can lead to the emergence of new routines, but also semantic-functional ones, such as changing the status of compounds from common nouns to proper nouns and using novel name classes and semantic relations. Thus, the criteria discussed in the literature on creativity in word formation need to be extended. In addition to (a) the formal properties of a pattern, (b) the degree of consciousness with which an expression is formed, and (c) the expressive function of a pattern, (d) the functional-semantic modifications of a pattern must be considered if one wants to describe creative word formation processes comprehensively. Speakers thus can be creative not only in the way they formally realize a naming process but also in what they refer to.

5 Conclusion

The phenomena discussed here do not seem to be particularly creative at first glance: in Name-ICCs, linguistic material is simply copied. In PN compounds, speakers do not even change the form of the expression at all. In fact, however, these changes can be seen as sophisticated examples of linguistic creativity. With PN compounds, the pattern is extended by various semantic-functional properties, so that its area of use becomes larger, which in turn is presumably the reason for the observed increase in frequency. These changes are creative because they modify the original pattern, which is also noticed by the interlocutors. Name-ICCs, on the other hand, violate original constraints on compounding in German and have distinctive formal properties. They are particularly favored in product marketing and Internet pseudonyms.

We propose that creativity in word formation should consider not only formal aspects, but also functional-semantic modifications of patterns, such as changes in name classes and semantic relations. Both types of phenomena, those that change a pattern in terms of its form and those that change it in terms of semantics and function, must be considered creative. What matters is that speakers do something in a new way, that doing something in a new way is advantageous, and that others recognize this. Creativity in language is not limited to formal aspects but extends to how meanings are conveyed and interpreted in different contexts. This perspective broadens our understanding of linguistic creativity.

References

Albert, Georg. 2013. *Innovative Schriftlichkeit in digitalen Texten: syntaktische Variation und stilistische Differenzierung in Chat und Forum*. Berlin: Akademie-Verlag.

Bach, Adolf. 1953. *Deutsche Namenkunde II. Die deutschen Ortsnamen. Vol. 2.* Heidelberg: Carl Winter Universitätsverlag.

Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2000. System vs. norm: coinage and institutionalization. *Morphologie/Morphology* 17(1). 832–840.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Belosevic, Milena. 2022. *Veggie-Renate* und *Merci-Jens*: Semantik und Pragmatik onymischer Personennamenkomposita. *Zeitschrift für germanistische Linguistik* 50(2). 289–319. https://doi.org/10.1515/zgl-2022-2056.

Briegleb, Otto. 1928. *Vom Wesen der Zusammensetzung. (Zusammensetzung mit der Mehrzahl?)*. Meinigen: Renßner'sche Hofbuchdruckerei.

Briegleb, Otto. 1932. *Wider die Entartung der Sprache durch Erstarrung – Falsche Zusammensetzung – Wortverstümmelung*. Leipzig: Verlag von Friedrich Brandstetter.

Bross, Fabian & Katherine Fraser. 2020. Contrastive focus reduplication and the modification puzzle. *Glossa: a journal of general linguistics* 5(1). 1–18.

Donalies, Elke. 2011. *Basiswissen Deutsche Wortbildung*. Tübingen u.a.: Francke.

DTA-Korpus. 2018. Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Herausgegeben von der Berlin-Brandenburgischen Akademie der Wissenschaften, Berlin. http://www.deutschestextarchiv.de/

Dürscheid, Christa. 2005. Normabweichendes Schreiben als Mittel zum Zweck. *Muttersprache* 115(1). 40–53.

Erben, Johannes. 1981. Neologismen im Spannungsfeld von System und Norm. In Brigitte Schlieben-Lange (ed.), *Geschichte und Architektur der Sprachen*. 35–43. Berlin & New York: De Gruyter.

Fanselow, Gisbert. 1981. *Zur Syntax und Semantik der Nominalkomposition: ein Versuch praktischer Anwendung der Montague-Grammatik auf die Wortbildung im Deutschen*. Tübingen: Niemeyer.

Finkbeiner, Rita. 2014. Identical constituent compounds in German. *Word Structure* 7(2). 182–213.

Fleischer, Wolfgang. 1982. *Wortbildung der deutschen Gegenwartssprache*. 5., unveränd. Aufl. Tübingen: Niemeyer.

Fleischer, Wolfgang & Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache*. Berlin/Boston: De Gruyter.

Frankowsky, Maximilian. 2022. Extravagant expressions denoting quite normal entities: Identical constituent compounds in German. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant Morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology*, 156–179. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.223.07fra.

Frankowsky, Maximilian. 2024. *N+N-Komposita mit identischen Konstituenten im Deutschen. Theorie und Empirie zu reduplikativer Komposition*. Berlin & Boston: De Gruyter.

Freywald, Ulrike. 2015. Total Reduplication as a productive process in German. *Studies in Language* 39(4). 905–945. https://dol.org/10.1075/s1.39.4.06fre

Günther, Hartmut. 1979. N+N: Untersuchungen zur Produktivität eines deutschen Wortbildungstyps. In: Leonhard Lipka & Hartmut Günther (eds.), *Wortbildung*, 258–280. Darmstadt: Wissenschaftliche Buchgesellschaft.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068. https://doi.org/10.1515/ling.37.6.1043.

Haspelmath, Martin & Uri Tadmor. 2009. *Loanwords in the world's languages: a comparative handbook*. New York: Mouton de Gruyter.

Henzen, Walter. 1965. *Deutsche Wortbildung*. 3rd edn. Tübingen: Niemeyer.

Hoffmann, Thomas. 2018. Creativity and construction grammar: Cognitive and psychological issues. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 259–276.

Hohenhaus, Peter. 2004. Identical constituent compounding – a corpus-based study. *Folia Linguistica* 38. 297–331. https://doi.org/10.1515/flin.2004.38.3–4.297.

Hohenhaus, Peter. 2005. Lexicalization and Institutionalization. In Pavol Stekauer & Rochelle Lieber (eds.), *Handbook of Word Formation*, 353–373. Dordrecht: Springer.

Hopper, Paul J. & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge University Press.

Ishiyama, Osamu. 2014. The Nature of Speaker Creativity in Linguistic Innovation. In Evie Coussé & Ferdinand von Mengden (eds.), *Usage-Based Approaches to Language Change*, 147–65. Amsterdam and Philadelphia, PA: Benjamins.

Ivorra Ordines, Pedro Ivorra. 2023. *Por mí como si te operas.* Constructional idioms of rejection from a constructionist approach. *Yearbook of Phraseology* 14. 89–120.

Jakobson, Roman. 1960. Linguistics and poetics. In Thomas Sebeok (ed.), *Style in Language*, 350–377. Cambridge (Mass.): MIT Press.

Jones, Rodney H. 2015. Introduction. In Rodney H. Jones (ed.), *The Routledge Handbook of Language and Creativity*, 1–21. Abingdon: Routledge.

Kandler, Christian, Rainer Riemann, Alois Angleitner, Frank M. Spinath, Peter Borkenau & Lars Penke. 2016. The Nature of Creativity: The Roles of Genetic Factors, Personality Traits, Cognitive Abilities, and Environmental Sources. Journal of Personality and Social. *Psychology* 111, 230–249.

Kauffmann, Charles, A. 2015. Reduplication reflects uniqueness and innovation in language, thought and culture. *Omniglot: the online encyclopedia of writing systems and languages*. http://www.omni glot.com/language/articles/reduplication.htm/ (accessed 9 November 2023).

Kecskes, Istvan. 2016. Deliberate Creativity and Formulaic Language Use. In Keith Allan (ed.), Pragmemes and Theories of Language Use. Perspectives in Pragmatics, [Special issue]. *Philosophy & Psychology* 9. 3–20.

Keller, Rudi. 1994. *On language change: The invisible hand in language*. London & New York: Routledge.

Kentner, Gerrit. 2017. On the emergence of reduplication in German morphophonology. *Zeitschrift für Sprachwissenschaft* 36(2). 234–277.

Kersten, Saskia & Netaya Lotze 2022. Anonymity and authenticity on the web. Towards a new framework in internet onomastics. *Internet pragmatics* 5(1). 38–65. https://doi.org/10.1075/ ip.00074.ker.

Klein, Thomas, Hans-Joachim Solms & Klaus-Peter Wegera. 2009. *Mittelhochdeutsche Grammatik. Teil 3. Wortbildung*. Tübingen: Niemeyer.

Klockow, Reinhard. 1980. *Linguistik der Gänsefüßchen. Untersuchungen zum Gebrauch der Anführungszeichen im gegenwärtigen Deutsch*. Frankfurt am Main: Haag und Herchen.

Köpcke, Klaus-Michael. 2002. Die sogenannte i-Derivation in der deutschen Gegenwartssprache. Ein Fall für output-orientierte Wortbildung. *Zeitschrift für germanistische Linguistik* 30(3). 293–309. https://doi.org/10.1515/zfgl.2003.001.

Koptjevskaja-Tamm, Maria. 2009. Proper-name compounds in Swedish between syntax and lexicon. *Rivista di Linguistica / Italian Journal of Linguistics* 21. 119–148.

Körtvélyessy, Lívia; Pavol Štekauer & Pavol Kačmár. 2021. On the role of creativity in the formation of new complex words. *Linguistics* 59(4). 1017–1055. https://doi.org/10.1515/ling-2020-0003

Kosmata, Eva. in prep. *Zur Produktivität lexikalischer Muster: Eine Fallstudie über die deutsche Eigennamenkomposition im 20. Jahrhundert*. Dissertation, Freie Universität Berlin.

Kürschner, Wilfried. 1974. *Zur syntaktischen Beschreibung deutscher Nominalkomposita auf der Grundlage generativer Transformationsgrammatiken*. Tübingen: Niemeyer.

Langlotz, Andreas. 2015. Language, creativity, and cognition. In Rodney H. Jones (ed.), *The Routledge Handbook of Language and Creativity*, 40–60. Abingdon: Routledge.

Levi, Judith. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Lieber, Rochelle. 2010. *Introducing morphology*. Cambridge: Cambridge University Press.

Lohde, Michael. 2006. *Wortbildung des modernen Deutschen: ein Lehr- und Übungsbuch*. Tübingen: Narr.

Mattiello, Elisa. 2013. *Extra-grammatical morphology in English: Abbreviations, blends, reduplicatives, and related phenomena*. Berlin & Boston: Walter de Gruyter.

Maybin, Janet & Joan Swann. 2007. Everyday Creativity in Language: Textuality, Contextuality, and Critique. *Applied Linguistics* 28(4). 497–517. https://doi.org/10.1093/applin/amm036.

Meibauer, Jörg. 2007. Syngrapheme als pragmatische Indikatoren: Anführung und Auslassung. In Sandra Döring & Jochen Geilfuß-Wolfgang (eds.), *Von der Pragmatik zur Grammatik*, 21–37. Leipzig: Universitätsverlage.

Munat, Judith. 2015. Lexical creativity. In Rodney H. Jones (ed.), *The Routledge Handbook of Language and Creativity*, 92–106. Abingdon: Routledge.

Neels, Jakob, Stefan Hartmann & Tobias Ungerer. 2023. 3 A quantum of salience. In Hendrik De Smet, Peter Petré & Benedikt Szmrecsanyi (eds.), *Context, Intent and Variation in Grammaticalization*, 47–78. De Gruyter Mouton. https://doi.org/10.1515/9783110753059-003.

Nowak, Jessica & Damaris Nübling. 2017. Schwierige Lexeme und ihre Flexive im Konflikt: Hör- und sichtbare Wortschonungsstrategien. In Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (eds.), *Sichtbare und hörbare Morphologie*, 113–144. Berlin & Boston: De Gruyter.

Nübling, Damaris. 2012. Auf dem Wege zu Nicht-Flektierbaren: Die Deflexion der deutschen Eigennamen diachron und synchron. In Rothstein, Björn (ed.), *Nicht-flektierbare und nicht-flektierte Wortarten*, 224–246. Berlin & New York: De Gruyter.

Nübling, Damaris, Fabian Fahlbusch, Rita Heuser. 2015. *Namen: Eine Einführung in die Onomastik*. Tübingen: Narr Francke Attempto.

Nübling, Damaris & Renata Szczepaniak. 2009. Religion+s+freiheit, Stabilität+s+pakt und Subjekt(+s+) pronomen: Fugenelemente als Marker phonologischer Wortgrenzen. In Peter O. Müller (ed.), *Studien zur Fremdwortbildung*, 195–222. Hildesheim: Olms.

Pennycook, Alastair. 2007. The rotation gets thick. The constraints get thin: Creativity, recontextualization, and difference. *Applied linguistics* 28(4). 579–596.

Petré, Peter. 2016. Unidirectionality as a cycle of convention and innovation: Micro-changes in the grammaticalization of [be *going* to INF]. *Belgian Journal of Linguistics* 30(1). 115–146.

Petré, Peter. 2017. The extravagant progressive: an experimental corpus study on the history of emphatic [BE V*ing*]. *English Language & Linguistics* 21(2). 227–250.

Platen, Christoph. 2013. *Ökonymie: Zur Produktnamen-Linguistik im Europäischen Binnenmarkt*. Berlin & Boston: Max Niemeyer. https://doi.org/10.1515/9783110941203

Ronneberger-Sibold, Elke. 2012. Blending between grammar and universal cognitive principles: Evidence from German, Farsi, and Chinese. In Vincent Renner, François Maniez & Pierre Arnaud (eds.), *Cross-disciplinary perspectives on lexical blending*, 115–144. Berlin & Boston: Walter de Gruyter. https://doi.org/10.1515/9783110289572.115.

Runco, Mark A. & Garrett J Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24(1). 92–96.

Schäfer, Roland & Felix Bildhauer. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. *LREC*. 486–493.

Schlücker, Barbara. 2017. Eigennamenkomposita im Deutschen. In Johannes Helmbrecht, Damaris Nübling & Barbara Schlücker (eds.), *Namengrammatik. [Special issue]. Linguistische Berichte* 23. 59–93. Hamburg: Buske.

Schlücker, Barbara. 2018. Genitives and proper name compounds in German. In Tanja Ackermann, Horst J. Simon & Christian Zimmer (eds.), *Germanic Genitives* (Studies in Language Companion Series 193), 275–299. Amsterdam & Philadelphia: Benjamins. https://doi.org/10.1075/slcs.193.10sch.

Schlücker, Barbara. 2020. Von *Donaustrom* zu *Donauwelle*. Die Entwicklung der Eigennamenkomposition von 1600–1900. *Zeitschrift für Germanistische Linguistik* 48(2). 238–268. https://doi.org/doi.org/10.1515/zgl-2020-2002.

Schmid, Hans-Jörg. 2014. Lexico-grammatical patterns, pragmatic associations and discourse frequency. In Thomas Herbst, Hans-Jörg Schmid & Susen Faulhaber (eds.), *Constructions, collocations, patterns*, 239–293. Berlin: Walter de Gruyter.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford University Press.

Smith, Eliot. R. & Diane. M. Mackie. 2000. *Social psychology*. Philadelphia: Psychology Press. https://doi.org/10.1093/oso/9780198814771.003.0005

Sick, Bastian. 2009. *Der Dativ ist dem Genitiv sein Tod – Folge 3: Noch mehr Neues aus dem Irrgarten der deutschen Sprache*. Köln: Kiepenheuer & Witsch.

Stein, Morris I. 1953. Creativity and culture. *The journal of psychology* 36(2). 311–322.

Sternberg, Robert J. & Todd I. Lubart. 1999. The concept of creativity: Prospects and paradigms. In Robert J. Sternberg (ed.), *Handbook of Creativity*, 3–15. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511807916.003

Stolz, Thomas. 2006. (Wort-)Iteration: (k)eine universelle Konstruktion. In Kerstin Fischer & Anatol Stefanowitsch (eds.), *Konstruktionsgrammatik. Von der Anwendung zur Theorie*, 105–132. Tübingen: Stauffenburg.

Stolz, Thomas, Cornelia Stroh & Ania Urdze. 2011. *Total reduplication the areal linguistics of a potential universal*. Berlin: Akademie-Verlag. https://doi.org/10.1524/9783050050973

Ungerer, Tobias & Stefan Hartmann. 2020. Delineating extravagance: Assessing speakers' perceptions of imaginative constructional patterns. *Belgian Journal of Linguistics* 34. 345–356.

Warren, Beatrice. 1978. *Semantic patterns of noun-noun compounds* (Gothenburg Studies in English 41). Göteborg: Acta Universitatis Gothoburgensis.

Wustmann, Gustav. 1891. *Allerlei Sprachdummheiten. Kleine deutsche Grammatik des Zweifelhaften, des Falschen und des Häßlichen*. Leipzig: Grunow. https://download.digitale-sammlungen.de/pdf/1413989909bsb11023653.pdf.

Stefan Hartmann & Tobias Ungerer

Chaos Theory, Shmaos Theory

Creativity and Routine in English *shm*-Reduplication

Abstract: This paper investigates an "extravagant" construction at the interface of morphology and syntax: English *shm*-reduplication, a pattern in which a word is immediately repeated, but the initial consonant or consonant cluster is either replaced by /ʃm/, or /ʃm/ is added to the beginning of a word if it begins with a vowel. So far, research on *shm*-reduplication has been limited to small samples of the construction and has mainly focused on its phonological and morphosyntactic properties rather than its semantics. The present study aims at filling this gap with a corpus-based analysis of a larger dataset from the web corpus ENCOW-16AX. Our findings suggest that *shm*-reduplication combines elements of routine and creativity: While the pattern is largely unconstrained in its semantics, its unusual syntactic profile and tight connection with specific communicative contexts mark it as an innovative and extravagant construction.

1 Introduction

Recent years have seen increased interest in so-called expressive morphology, i.e., morphology that "is associated with an expressive, playful, poetic, or simply ostentatious effect of some kind" (Zwicky and Pullum 1987: 335). Examples include pejorative compound patterns (e.g., German *rattenscharf* 'great/attractive', lit. 'rat sharp'; Meibauer 2013), *doubler-upper* compounds (e.g., *stayer-onner-for-nower*; Lensch 2018), libfixes (e.g., *safety-o-cracy*; Norde and Sippach 2019), and pseudo-participles (e.g., German *bebrillt* 'be-glassed', Kempf and Hartmann 2022). Alternatively, such phenomena have been characterized as instances of linguistic "extravagance," a term popularized by Haspelmath (1999) to describe speakers' use of noticeable language to stand out (see also Ungerer and Hartmann 2020). Another particularly interesting pattern in this regard is English *shm*-reduplication, as exemplified in (1).

(1)   a. *And I did, and didn't actually say anything, just sort of nn uh yuh un uh.* **Language, schmanguage**.
        (http://www.guardian.co.uk/theobserver/2000/sep/03/features.magazine37, ENCOW)

b. ***Peer review, schmeer review.*** *Good article!*[1]
(http://www.labspaces.net/blog/profile/624/Odyssey, ENCOW)

In this reduplicative pattern, a word is immediately repeated, but the initial consonant or consonant cluster (if any) is replaced by /ʃm/ (McCarthy and Prince 1986; Nevins and Vaux 2003; Mattiello 2013: 153). In writing, the cluster appears as <shm> or <schm>. So far, most research on English *shm*-reduplication has focused on its phonological and syntactic properties (e.g., Nevins and Vaux 2003; Grohmann and Nevins 2004; Saba Kirchner 2010; Kołłątaj 2016; but see Mattiello 2013 for a discussion of semantic and pragmatic aspects). Previous studies also had to rely on a fairly limited amount of authentic data as the construction does not occur very often in standard language data that make up for the bulk of widely-used corpora. For instance, a quick case-insensitive search for words starting with *shm*- and *schm*- in the 100-million-word British National Corpus (BNC) reveals that the construction is not attested there at all (although there are some *shm*-attestations without reduplication, e.g., *schmendrick* or *he schmiled*); and while Mattiello's (2013) study is based on data from multiple dictionaries and neologism databases, she could only work with 22 attestations of *shm*-reduplication (Mattiello 2013: 8). However, web corpora like ENCOW (Schäfer and Bildhauer 2012; Schäfer 2015) make available large amounts of non-standard language data that allow for investigating low-frequency phenomena like *shm*-reduplication in more detail.

Against this background, the aim of the present paper is twofold. On the one hand, we will present a corpus-based account of the morphosyntactic and semantic properties of *shm*-reduplication. On the other hand, and on a more theoretical level, we will discuss the "extravagant" potential of the pattern against the backdrop of the trade-off between "creativity" and "routine". The remainder of this paper is structured as follows. Section 2 gives an overview of previous research on *shm*-reduplication and discusses how it can be approached from the theoretical perspective of Construction Grammar. Section 3 introduces the data sources and presents the descriptive results of our corpus study. Section 4 links the corpus results to a wider theoretical discussion of how creativity and routine interact in giving rise to the extravagant nature of *shm*-reduplication. Section 5 provides a summary and conclusion.

---

1 We would like to take this opportunity to thank the anonymous reviewers as well as the editors of the present volume for their helpful feedback. The usual disclaimers apply (in other words: as for any remaining shortcomings – errors, shmerrors).

2 The English *shm*-Reduplication Construction

As Nevins and Vaux (2003) and Finkbeiner, Meibauer, and Wiese (2016: 4) point out, *shm*-reduplication seems to have its origin in Yiddish, although it is a matter of debate whether it originated in Yiddish or whether it is an English-internal development based on "the numerous Yiddish words beginning with this cluster" (OED 2023, "schm-"; also see Southern 2005 for an in-depth study of the pattern's origins). Nevins and Vaux (2003) follow Weinreich (1980: 623–624) in assuming that the construction dates back several centuries in Yiddish; according to the OED, formations with *shm-* are found from the 1920s onwards.

The sound combination /ʃm/ seems to be closely associated with words of Yiddish origin with a slightly negative or even strongly derogatory meaning component. Compare Table 1, which lists the lemmas of all word forms beginning with the grapheme combination <schm> or <shm> (with lowercase <s> to exclude proper names) in the BNC. According to their OED definitions, most of these words have pejorative connotations: *schmaltz*, for instance, is used to describe "[e]xtreme or excessive sentimentality" (OED 2023, "schmaltz"), and *schmuck* refers to a "a stupid or foolish person" (OED 2023, "schmuck"). Given these emotive-evaluative meaning components, the sound combination /ʃm/ qualifies as a phonæstheme (Firth 1930), i.e., a frequently recurring sound-meaning pattern on the submorphemic level (see, e.g., Bergen 2004; Stroebel 2017).

Table 1: Words beginning with <shm-> or <schm-> in the BNC.

Frequency	Lemma
13	*schmaltz*
9	*schmuck*
8	*schmooze*
5	*schmaltzy*
2	*schmall, schmeichal, schmutter, shmuck*
1	*schmaltze, schmecker, schmendrick, schmidt, schmile, schmoe, schmoozer, shmaltz, shmatte, shmeckle, shmoozing*

In line with the semantics of its Yiddish-derived base words, *shm*-reduplication is typically connected with "mockery and ridiculing" (Kołłątaj 2016: 243) and "derogatoriness" (Mattiello 2013: 46), and it has been described as "ironic" (Inkelas and Zoll 2005: 42). According to Mattiello (2013: 153), "this type of construction generally shows a dismissive usage, but can also be employed to downplay a situation or problem that is potentially overwhelming or threatening, or to lighten a situation with humour".

shm-reduplication is a special case of echo reduplication (Grohmann and Nevins 2004), which in turn is a special case of reduplication, a phenomenon that is widespread in the languages of the world (Inkelas 2014: 169) but relatively rare in Indo-European languages (Schwaiger 2015: 478). This makes exceptions such as identical constituent compounding (*salad-salad,* Hohenhaus 2004; Finkbeiner 2012; Frankowsky 2022) or ablaut reduplication (*riff-raff, tip-top*, Minkova 2002) all the more salient. Reduplication shows a wide range of functions that have been characterized in terms of a radial network by Regier (1994), e.g., signaling smallness, affection, contempt, or intensity. As such, many reduplicative patterns can be seen as key examples of evaluative morphology in the sense of Grandi and Körtvelyéssy (2015), who argue that the function of patterns that have been described as "expressive" or "evaluative" morphology can be characterized along two parameters: descriptive or quantitative evaluation on the one hand (diminution and augmentation), and qualitative evaluation on the other (e.g., intensification, endearment, contempt; also see Grandi 2017). Many evaluative patterns can fulfill several of these functions in different contexts – for example, diminutives, in many languages, can express not only smallness but also endearment and contempt, alongside a broad variety of other meanings (Jurafsky 1996). Although echo reduplication tends to be associated with a smaller range of meanings (Inkelas 2014: 184), even the semantically fairly homogeneous pattern of *shm*-reduplication can have a range of different functions, as we will discuss in more detail below.

The functions of *shm*-reduplication, as of other reduplication patterns, can partly be explained in terms of iconicity (see, e.g., Fischer 2011; Kentner 2017, 2022). Bybee et al. (1994: 167) assume that reduplicatives emerged as maximally iconic patterns in which the repetition of a verb signals the repetition of the action described by the verb. Fischer (2011: 64) sees the principle of "more of the same" as the key iconic source that can explain many apparently non-transparent instances of reduplication. This increase in quantity can occur in the vertical dimension (augmentation, intensification) or horizontal dimension (plurality, iteration, distribution; also see Kentner 2023: 105). This can entail, among other possibilities, a meaning of 'smallness': "Observing a multitude of similar items spread out on a horizontal plane makes each individual item appear relatively small and blurry." (Kentner 2023: 105) Moreover, Fischer (2011: 65) assumes that reduplication in child-directed speech "reflects the onomatopoeic imitation of the actual CV-CV syllabic babbling sounds made by children". Kentner (2023: 105) points out that this invites both negative and positive connotations. This general account can also be applied to *shm*-reduplication, which can be interpreted as a diminutivizing strategy ascribing a lack of importance to the reduplicant.

Turning from the pattern's function to its form, the replacement of the onset (or, sometimes, word-internal material) that is characteristic of echo reduplication

(Inkelas 2014: 170) has been described as "melodic overwriting" (McCarthy and Prince 1986: 68; Inkelas and Zoll 2015: 42). A prime example of melodic overwriting is *m*-reduplication, which has been characterized as "virtually pan-Asian" (Inkelas and Zoll 2015: 42) or "pan-Anatolian" (Donabedian and Sitaridou 2021: 409), e.g., Turkish *beyaz-meyaz* 'white and such things/allegedly white' (Donabedian and Sitaridou 2021: 414). From Turkish, *m*-reduplication has been borrowed into many other languages (including Kiezdeutsch, a variety of German used by young speakers in multilingual urban settings, see Wiese and Polat 2016). According to Donabedian and Sitaridou (2021: 414), *m*-reduplication is "mostly used to refer to a whole conceptual domain or to recall any word from the context with scepticism or irony". *m*-reduplication shares the latter function with *shm*-reduplication, while the former is fulfilled by other morphological patterns in English (e.g., *-ish* suffixation).

The status of reduplication in general, and of echo reduplication in particular, has been a matter of debate (Downing and Inkelas 2015: 520–526; Schwaiger 2018). Inkelas and Zoll (2005: 2–4) distinguish two major approaches to reduplication, one focusing on phonology, the other on semantic as well as syntactic properties. From the perspective of usage-based Construction Grammar (see, e.g., Ungerer and Hartmann 2023), we can conceive of *shm*-reduplication as a construction, i.e., a pairing of form and meaning at various levels of generality. By combining the concatenation of words with a quasi-paradigmatic stem alternation, *shm*-reduplication blends features traditionally attributed to syntax with features that are commonly seen as morphological. A constructionist approach, however, does not assume a strict division between morphology and syntax and can therefore be seen as particularly well-suited for investigating phenomena such as *shm*-reduplication, which can be "located at the border of word-formation and syntax" (Finkbeiner, Meibauer and Wiese 2016: 4).[2] Following Nagaya (2020: 267), who analyzes reduplication phenomena in terms of Construction Morphology (Booij 2010), a simplified formalization of the *shm*-reduplication construction could look as follows:

(2) $< [X_i \, shm\text{-}X_i]_j \leftrightarrow [\text{DISMISSIVE } [\text{SEM}]_i]_j >$

This simple schema shows the form side of the construction on the left-hand side of the double arrow, and the meaning (or function) side on the right. A construc-

---

2 In fact, a reviewer points to another aspect in which *shm*-reduplication falls in between morphology and syntax: While identical constituent compounding (ICC) like *salad-salad* shows compound stress, *salad shmalad* shows phrasal stress. Note that this corresponds with the function of *shm*-reduplication: While ICCs refer to a specific entity and can therefore be readily interpreted as compound nouns, *shm*-reduplicatives express a certain evaluation of the reduplicant that is added to the unaltered original word.

tionist perspective would assume that speakers of English have a schema like (2) available as part of their linguistic knowledge, i.e., they know that echo reduplication with *shm-* modifies the semantics (SEM_i) of a given element (X_i) in such a way that it is (at least prototypically) interpreted as dismissive.

As this brief overview has shown, previous studies have highlighted several key characteristics of *shm-*reduplication, but they are based on small samples of the construction and do not provide in-depth quantitative analyses of its formal and semantic properties. In the following sections, we will explore what quantitative methods, when applied to a larger corpus sample, can tell us about the morphosyntactic and semantic profile of English *shm-*reduplication.

3 Corpus Study

3.1 Data and Methods

As *shm-*reduplication is expected to be largely a phenomenon of informal or conceptually oral language in the sense of Koch and Oesterreicher (1985), we opted for a corpus that documents computer-mediated communication, namely the webcorpus ENCOW16AX (Schäfer and Bildhauer 2012; Schäfer 2015).[3] ENCOW is part of the COW (= "Corpora from the Web") family of large web corpora in different languages. It contains almost 17 billion tokens from more than 9 billion documents.

We searched the corpus for words starting with <shm-> or <schm->. In the next step, the normalized Levenshtein distance (Levenshtein 1966) between the target word and the immediately preceding word was calculated, as well as between the target word and the second word preceding it. Levenshtein distance is a widely-used measure of edit distance between two strings: For instance, it takes one edit to get from *house* to *mouse*; taking the number of characters into account, this yields a normalized Levenshtein distance of 0.2. Similarly, it takes five edits to get from *house* to *castle*; divided by the length of the longer item, this yields a normalized Levenshtein distance of 0.83. As such, this measure allows for detecting graphemically similar words. In this way, we were able to detect *shm-*reduplicants based on

---

[3] It should be noted, however, that ENCOW does not exclusively consist of computer-mediated communication – it contains a variety of text types, including some texts that are older than the Internet but happen to be available online. Still, the chances of finding instances of *shm-*reduplication are much higher in such a corpus than in other corpora. Compared to other widely-used web corpora, ENCOW has the advantage that it is freely accessible, which entails significant advantages for reproducibility and replicability (see Hartmann 2024).

individual lexemes, such as *Brexit, shmexit*, and based on compounds, such as *Chaos theory, shmaos theory*, or phrases, such as *ecological problems, shmecological problems*. A few cases involving more complex compounds or phrases may have been overlooked by taking this approach, and cases where the *shm-* is inserted word-internally, as in *obscene obshmene* (Nevins and Vaux 2003), are also neglected, but overall, this operationalization seems to offer a good balance between precision and recall. The items with a normalized Levenshtein distance of 0.4 and below were then checked manually. To analyze the morphosyntactic characteristics of the dataset, we manually annotated each instance for part of speech, syntactic position (e.g., separate sentence, sentence-initial but syntactically non-integrated, etc.), and morphological complexity (simplex or complex).

Overall, we ended up with 1,642 tokens of *shm*-reduplication, 1,557 of which had a simplex base (879 different types), while 85 instances had a compound or phrasal base (78 types).

In terms of semantic characteristics, we relied on several complementary approaches to assess the semantic spectrum of the pattern and its expressive potential. On the one hand, we used semantic vector-space modeling (see, e.g., Perek 2016) to map out the overall semantic spectrum of the construction and examine whether it clusters around particular areas of the possible semantic space. The basic assumption of this approach is that similar words tend to occur in similar contexts (Boleda 2020). More specifically, we employed a word2vec model originally created for Hartmann and Ungerer (2024), which was trained with Schmidt and Li's (2022) *wordVectors* package for R (R Core Team 2023) on the basis of the first of the 17 downloadable sets of ENCOW sentence shuffles (comprising around 600 million sentences). The training used a five-word window (i.e., five words before and after the target word were taken into account) and a skip-gram approach with negative sampling (i.e., the algorithm randomly samples other words in the text that are not neighbors; see Mikolov et al. 2013 for details). Note that the vectors resulting from this process are based on the entire dataset, i.e., not just on the occurrences of the target words in the *shm*-reduplication construction. For the visual representation of the results, we also follow Hartmann and Ungerer (2024) in using t-distributed Stochastic Neighborhood Embedding (t-SNE, van der Maaten and Hinton 2008), which allows for representing *n*-dimensional data in two-dimensional space.

On the other hand, we used collostructional analysis (Stefanowitsch and Gries 2003) to determine the lexemes that are strongly attracted to, or repelled by, the construction. This allowed us to gauge, from a more qualitative perspective, whether there are any semantic regularities among the items that most typically combine with *shm*-reduplication.

Finally, we drew on Warriner, Kuperman, and Brysbaert's (2013) affective meaning norms to check whether the construction tends to attract lexemes that

can be considered particularly expressive. We focused on two dimensions of affectedness: valence (defined as the pleasantness of the emotions invoked by a word) and arousal (capturing the intensity of emotion provoked by a word). Using crowdsourcing methods, Warriner, Kuperman, and Brysbaert had participants rate a total of 13,915 lexemes on 9-point scales (1 = lowest, 9 = highest) ranging from *unhappy* to *happy* (for valence) and from *calm* to *excited* (for arousal). Our rationale was that, if the valence and arousal values of *shm*-reduplicants differed from the rest of the norming set, this would suggest that the construction tends to be used to convey speakers' emotional involvement in the subject matter. Naturally, Warriner, Kuperman, and Brysbaert's norms have potential limitations: For instance, as one of our reviewers suggests, participants may have found it difficult to judge the affective properties of abstract words such as *title* or *easy*. Nevertheless, given that norming ratings (especially for valence) were quite consistent across participants (Warriner, Kuperman, and Brysbaert 2013: 1194) and that affective norms have informed a large body of research on emotions and word processing (Kuperman et al. 2014), we used them here as a (tentative) way of probing the expressive potential of our construction.

3.2 Results

We first present the results of our morphosyntactic analysis (Section 3.2.1), before examining the semantic characteristics of the construction (Section 3.2.2).

3.2.1 Morphosyntactic Analysis

Figure 1 illustrates how different parts of speech are distributed among the instances in our dataset. Overall, *shm*-reduplication is clearly biased towards nominal uses, which account for almost 70% of all instances. Adjectives form the second most common group (29.1%), while verbs and adverbs are very infrequent (together less than 2% of the data). An example of each category is provided in (3).

(3) a. **Maps, schmaps..** *They work just fine* (http://www.zdnet.com/sinofskys-de parture-from-microsoft-politics-or-products-to-blame-7000007297)
   b. **Trustworthy Shmustworthy.** *Who's the judge anyway.* (http://www.matt cutts.com/blog/snarky-or-not)
   c. *In the end I built – oh,* **built schmilt***.*
      (http://www.gold.ac.uk/glits-e/glits-earchive/2010-2011/contentspage/inter textualtransformationandtextualerasure)

d. ***Maybe, schmaybe****, but I hold UK standards in education in high regard* (http://www.chinasmack.com/2013/stories/dual-track-pension-system-civil-servants-6000-farmers-55-rmb.html)
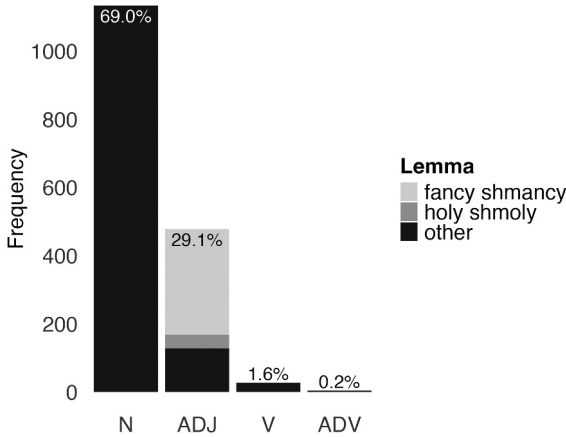
Figure 1: Distribution of parts of speech in the data.

Moreover, Figure 1 also shows that most adjectival uses of the construction are, in fact, restricted to two highly frequent types. First, *fancy shmancy* (310 attestations) accounts for 64.9% of adjectival uses, and for 18.9% of all instances of the construction, thus constituting by far the most frequently attested example of *shm*-reduplication. Second, *holy shmoly* (40 attestations) accounts for 8.4% of adjectival uses, while making up 2.4% of the entire dataset. This means that the rest of the construction is even more heavily biased towards nominal uses: While a range of other adjectives are attested in the construction (e.g., *indie schmindie*, *legal schmegal*), they together make up less than 10% of our data.

As a second step, we examined the syntactic positions in which *shm*-reduplication occurs. Our findings are summarized in Figure 2. The results indicate that most instances of *shm*-reduplication fill syntactically isolated positions.[4] Around half the time (48.9%), they occur in a separate sentence (or occasionally, split into two sentences), as illustrated in (4a). Moreover, when they form part of a sen-

---

4 As a reviewer points out, this tendency can be observed for other reduplicative patterns as well, e.g., reduplicative phrases such as German *Argumente hin, Argumente her* (lit. 'arguments to, arguments fro'; Finkbeiner 2017), which may be connected to the pragmatic function of these patterns: The base words have previously been mentioned or are otherwise salient in the discourse and are taken up as a quasi-quotation.

tence, they frequently occur in syntactically non-integrated positions, i.e., not as an argument or adjunct of the verb. Most of these non-integrated positions are sentence-initial (15.4%), as in (4b); while sentence-medial (1.4%) and sentence-final positions (3.4%) are also attested but less frequent, see (4c) and (4d). Syntactic non-integration is often signaled by a punctuation mark (e.g., comma, colon, dash), although this is not always the case given the inconsistency in the use of punctuation in informal (especially internet) language.

(4) a. *Everest schmeverest. Try climbing an active volcano!* (http://ngadven ture.typepad.com/blog/the-adventure-top-10)
   b. *Beatles, schmeatles – who else is missing from iTunes?* (http://blog.wash ingtonpost.com/clicktrack/2010/11/us_royalty_bridging_indie_rock.html)
   c. *Who cares about all that self analysis as artists, **ego schmeego**, it's just painting!!!* (http://clicks.robertgenn.com/nom-de-brush.php)
   d. *The issue is scarcely ever raised in public by US officials – **commitment, shmemitment**.* (http://badattitudes.com/MT/archives/2007/09/mr_olmert_ tear.html)
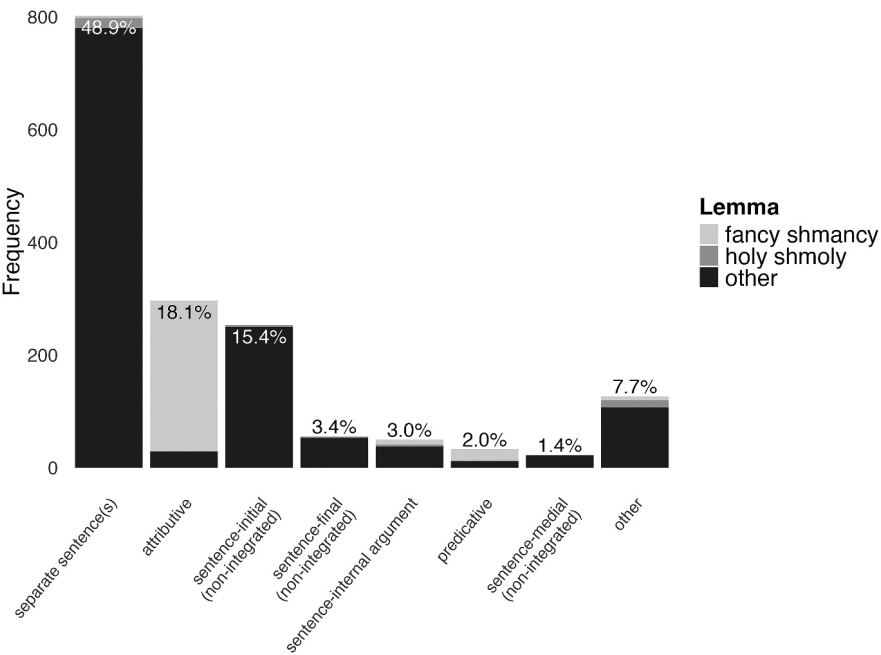
Figure 2: Distribution of syntactic positions.

In contrast, syntactically integrated positions within the sentence are less common. In particular, while attributive and predicative uses make up a considerable proportion (together 20.1%) of the dataset, most of these instances (around 90%) consist of *fancy shmancy*, as illustrated in (5a). Interestingly, this means that many other adjectives, when they occur in the construction, are not integrated into sentence-internal syntactic structures but rather appear in separate sentences, as in (5b). Beyond this, only relatively few instances of *shm*-reduplication (3%) occur as sentence-internal arguments, for example as a subject in (5c). Finally, there were a number of cases that we classified as "other," for example because they appeared sentence-internally but as quotations, or because their syntactic status was unclear.

(5) a. *i'd buy myself a **fancy schmancy** dress for Christmas* (http://www.shef
     fieldforum.co.uk/archive/index.php/t-166255.html)
   b. ***Fair, shmair.** That's life in Big City, USA, Planet Real World.* (http://www.
     j-bradford-delong.net/movable_type/2003_archives/001699.html)
   c. ***Details schmetails** will be forthcoming.* (http://derspatchel.livejournal.
     com/810339.html?page=3%26cut_expand=1)

Turning to the morphosyntactic complexity of the reduplicants, 64 tokens in our dataset (58 types) are based on compounds, usually N+N compounds as in *chaos theory, shmaos theory*, while 21 tokens (20 types) are based on phrases. The distinction between compounds and phrases is, of course, notoriously difficult (see, e.g., Schlechtweg 2018: 39–135), and there are a few doubtful cases in our data as well (e.g., *civil liberties*, annotated as a phrase here, although it could also be argued to be lexicalized as a holistic unit). Overall, however, compounds seem to outnumber phrases, which indicates that *shm*-reduplication tends to target single units. Phrasal reduplicants, as in *early morning, shmearly shmorning*, are less typical and hence arguably more salient or "extravagant" (see Section 4 for a more detailed discussion of this term). This is also supported by the observation that the proportion of phrases is higher among the 11 attestations in which both constituents are subject to melodic overwriting than among the remaining attestations in which *shm-* is only inserted into one (always the first) constituent: Five out of 11 instances with double melodic overwriting are phrases (45.5%), e.g., *optical density, schmoptical schmensity* or *enigmatic smile, schmnegmatic schmile*. Among the 74 complex instances in which only the first constituent is subject to melodic overwriting, only 16 are phrases (21.6%). Assuming that double melodic overwriting entails a higher degree of salience than single melodic overwriting, it seems plausible that speakers exploit the full extravagant potential of the pattern

by combining an unusual type of reduplicant (a phrase) with a highly salient doubling of melodic overwriting.

3.2.2 Semantic Analysis

For the semantic analyses, only *shm*-reduplications with a simplex base were taken into account; see below for a separate analysis of the (few) compound and phrasal attestations.

Figure 3 displays the results of the semantic vector-space analysis. The size of the words corresponds to their (logged) frequency in the construction. To read the many small-print (i.e., low-frequency) words, readers may refer to the online version of the diagram (see Section 6 "Data availability"). The ellipses in Figure 3 show clusters that were identified using partitioning around medoids (*pam*), a non-hierarchical clustering technique in which observations are clustered around the exemplar from which the average semantic distance to all other members of the cluster is minimal (see, e.g., Levshina 2015: 317).[5] Although some of the resulting clusters are quite heterogeneous, some interesting tendencies can be detected. Going clockwise through the graph, the cluster in the upper-left corner mostly contains lexemes from the domains of media and technology, e.g., *internet, cloud, graphics, usability*. Next to it, the top-center cluster revolves around financial and administrative terms, e.g., *capital, price, agency, logistics*. The cluster on the upper right comprises terms related to science and legal matters, e.g., *statistics, ethics, rule, privacy*. The cluster below that contains concepts from politics and society (*democracy, tradition*), abstract values (*honor, dignity*), and discourse (*contradiction, fact*). At the bottom center, we see a variety of terms related to technology (*helicopter, gear*), food and drinks (*pizza, booze*), and animals (*goose, turtle*). The bottom-left cluster is also rather heterogeneous, comprising concepts from the domains of religion (*holy, martyr*), people and celebrities (*princess, beckham*), sports (*basketball*), and time (*saturday, morning*). Finally, the center of the diagram contains terms from medicine and science (*cancer, photon, multicomponent*) as well as some more generic adjectives (*easy, moderate*).

Interestingly, there are some tendencies that apply across the different clusters. For example, the construction seems to show a preference for more or less technical terms from domains like science and technology, as well as from social and political domains. At the same time, the semantics of the lexemes that un-

---

5 The plot was created using the packages *ggplot2* (Wickham 2016) and *ggforce* (Pedersen 2024). For detecting the *pam* clusters, the package *cluster* (Maechler et al. 2023) was used.

Figure 3: Results of a semantic vector-space analysis using a word2vec model trained on ENCOW data.

dergo *shm*-reduplication is extremely broad, which indicates that constraints on the construction's productivity may pertain more to formal and phonological features than to the semantics of the base words.

This also becomes clear when taking a closer look at the results of the collostructional analysis. There have been some debates regarding the adequate statistical methods to use when performing collexeme analyses, and the general consensus that is currently emerging is that different measures can be seen as complementary as they cover different relevant dimensions (see, e.g., Schneider 2020; Hartmann and Ungerer 2024). In Table 2, we draw on three measures, the first being the log likelihood G^2, a bidirectional measure that quantifies the *mutual* attraction between lexemes and constructions. G^2 is strongly correlated with other popular collostructional measures such as the *p*-value of the Fisher-Yates Exact Test or chi-squared residuals (Gries 2023), which is why only one of these three measures is reported here. These measures have in common that they are

sensitive to frequency. The odds ratio, by contrast, is independent of sample size and can therefore be used as a measure of "pure" association that is not influenced by sample size. Like G^2, it is a bidirectional measure. Delta P (ΔP), in turn, is unidirectional and assesses the degree to which a construction "attracts" a specific lexeme, or vice versa (Schneider 2020). For example, *fancy* has a very high construction-to-word ΔP, as *fancy shmancy* makes up a considerable proportion of all attestations of *shm*-reduplication in our data. By contrast, its word-to-construction ΔP is fairly low as *fancy* is attested much more often outside of than within the construction.

Table 2 shows the top 30 attracted collexemes, sorted by G^2. Perhaps most importantly, the results clearly show a skew in the distribution of the collexemes, especially when looking at the G^2 value, which takes the frequency of the lexemes into account. *fancy shmancy* is by far the most frequent exemplar, and consequently, *fancy* is the most strongly attracted lexical item, followed by *holy*, as in *holy shmoly*. This confirms the observation made in Section 3.2.1 that these two types are particularly prominent examples of *shm*-reduplication, which make up the lion's share of its adjectival instances. Apart from that, the semantic groups identified in the vector-space analysis above can also be seen in the results of the collostructional analysis, with terms from (pop) culture and media as well as technical terms and socio-political vocabulary ranking high.

Table 2: Top 30 attracted collexemes.

Collexeme	Corpus frequency	Frequency in construction	Odds ratio	ΔP word-to-construction	ΔP construction-to-word	G^2
fancy	194089	310	3.68	0.0016	0.2	4564.42
holy	202491	40	2.69	0.0002	0.026	414.13
oedipus	89	10	5.5	0.11	0.0066	231.33
indie	49388	11	2.75	0.00022	0.0072	116.31
multicomponent	509	5	4.42	0.0098	0.0033	90.74
politics	548713	12	1.74	0.00002	0.0077	71.6
twitter	81869	8	2.4	0.0001	0.0052	71.43
monkeon	35	3	5.41	0.086	0.002	67.67
ethics	937	4	4.06	0.0043	0.0026	65.9
lawsuit	126192	8	2.21	0.00006	0.0052	64.54
photon	32776	6	2.68	0.00018	0.0039	61.07
internet	10245	5	3.11	0.00049	0.0033	60.68
libros	134	3	4.81	0.022	0.002	59.42
rights	2	2	7.08	1	0.0013	58.77
statistics	375	3	4.35	0.008	0.002	53.2
millennium	64147	6	2.39	0.00009	0.0039	53.04

Table 2 (continued)

Collexeme	Corpus frequency	Frequency in construction	Odds ratio	ΔP word-to-construction	ΔP construction-to-word	G²
privacy	267201	8	1.88	0.00003	0.0052	52.66
vermeer	8	2	5.97	0.25	0.0013	49.77
dilepton	57	2	5.03	0.035	0.0013	41.44
jointery	73	2	4.92	0.027	0.0013	40.43
rule	1820515	11	1.18	0.00001	0.0067	38.42
oom	228	2	4.42	0.0088	0.0013	35.84
beckham	296	2	4.31	0.0068	0.0013	34.79
accuracy	179452	5	1.87	0.00003	0.0032	32.19
graphics	202436	5	1.82	0.00002	0.0032	31
build	1	1	6.86	1	0.00066	29.38
hillphones	1	1	6.86	1	0.00066	29.38
mexican	1	1	6.86	1	0.00066	29.38
windows	1	1	6.86	1	0.00066	29.38
iphone	22409	3	2.57	0.00013	0.002	28.66

Given that the overall number of attestations is comparatively low and, more importantly, that very few instances of the pattern are attested more than a handful of times, the results of the collexeme analysis have to be taken with caution, regardless of which measure we use. But even some of the facts that potentially contribute to reducing the validity of the results reported in Table 2 can be quite informative about the pattern itself: For example, some of the collexemes have a high rank because they are attested very infrequently in the corpus as a whole, which entails that small differences in either corpus or construction frequency can make a large difference regarding the rank of the lexical item in question. For example, both corpus and construction frequencies are low in the case of some proper names like *Beckham, Monkeon* (apparently the name of a forum user) and *Vermeer*, which tend to occur in very specific contexts only. Others are highly technical, such as *dilepton* or *jointery*. The case of *oom* illustrates a further problem of collostructional analysis, namely that it is blind to polysemy (Dekalo and Hampe 2017). In ENCOW, *oom* typically occurs as a discourse marker (as in *uhm*) or in Dutch material (as a preposition meaning 'around'). In *oom-shmoom*, however, *oom* represents the Hebrew pronunciation of *U.N.*, as (6) shows.

(6)   *United Nations is 'oom', and as David Ben Gurion said about the philanderings of the U.N., "**oom shmoom**!"* (http://www.michaeltotten.com/archives/2009/05/did-hezbollah-k.php, ENCOW)

These limitations, however, are informative about the pattern in at least three ways: First, the high type-token ratio indicates that the pattern is very productive and that it allows for creating many ad-hoc coinages tailored to specific situations. Second, the fact that some very infrequent lexemes occur in the construction also supports the impression that its uses are often tailored to highly specific contexts. Third, the fact that the data contain material from other languages such as Hebrew or Yiddish testifies to the inherently multilingual (and multicultural) character of the construction. *shm*-reduplication seems to be connected to the Yiddish language, and to Jewish culture, at least to some extent. The third-ranked collexeme, *Oedipus*, also bears witness to this as it refers to a classic Jewish joke (Paley 2019), the punchline of which is quoted in (7).[6]

(7)   *Remember the wise words of Jewish mothers everywhere. "**Oedipus, shmoedipus** – who cares, so long as he loves his mother."*
        (http://newhumanist.org.uk/595/hail-mary-, ENCOW)

The next step of our analysis was to assess the emotional valence of the words that occur in *shm*-reduplication, by drawing on Warriner, Kuperman, and Brysbaert's (2013) widely used affective meaning norms. As described in Section 3.1, these norms provide valence and arousal ratings for almost 14,000 English lexemes, measured on a scale from 1 (lowest) to 9 (highest). Figure 4 shows the distribution of valence and arousal ratings for all lemmas from our dataset that also occur in the norming data. The font size of the lexemes corresponds to their (logged) frequency in the *shm*-reduplication construction. For better readability of the high-frequency lexemes, lexemes attested at least three times are shown in black while the less frequent ones are displayed in gray. The graph is also available online (see Section 6 "Data availability"), allowing readers to zoom into its details.

If Kołłątaj's (2016: 243) assumption that "[t]he first component of a reduplicative compound in itself is normally endowed with a positive meaning" was correct, then we would expect the data to be skewed towards high valence ratings. Indeed, the median of the valence value is slightly higher for lexemes that occur in the *shm*-reduplication construction than for those lexemes in Warriner, Kuperman, and Brysbaert's (2013) norms that are not attested in our dataset. According to a two-sample t-test, the difference is highly significant (t=5.49, df=636, p<0.001). This suggests that the construction shows a certain affinity to lexical items with a positive semantic prosody, although there is also a considerable number of items on the left-hand (i.e., low-valency) side of the scale, some of which occur quite fre-

---

6 Thanks to Barbara Wehr for pointing this out to us.

Figure 4: Distribution of the lexemes occurring in the *shm*-reduplication construction across the continuum of valence and arousal according to Warriner, Kuperman, and Brysbaert's (2013) affective words list.

quently in the construction. As for the arousal values, we did not find any significant difference between the lexemes occurring in the construction and the lexemes not attested in the construction (two-sample *t*-test, *t*=1.13, *df*=636, *p*=0.26). Overall, then, the construction seems to show a slight preference towards items with a positive semantic prosody, but at the same time, it combines with a broad variety of different items across the entire spectrum of valence and arousal values.

Finally, let us take a look at the instances of *shm*-reduplication with a compound or phrasal base that we have neglected in the semantic analysis so far. Only five compounds are attested more than once in our data (*black mail, climate change, earth day, tea party, peer review*). Still, they are quite representative of the semantic spectrum of *shm*-reduplication as discussed above, as they contain environmental and socio-political terms as well as one term from the academic domain. The 11 compounds or phrases in which both constituents are subject to melodic overwriting also show the full range of the semantic spectrum discussed above, from socio-political and legal terms (*individual mandate, civil liberties*) to pop culture (*kindergarten cop*) and technical terms from science and technology (*optical density*).

3.3 Discussion

Overall, our corpus results illustrate that *shm*-reduplication has a relatively "specialized" syntactic and pragmatic profile, while at the same time being flexibly used across a variety of semantic domains. On the one hand, the syntactic analysis suggests that the construction is prototypically applied to nouns, and that most of its instances are syntactically non-integrated, occurring either as separate sentences or at the sentence periphery. This also provides indirect evidence for the pragmatic function of *shm*-reduplication, which is typically used to express dismissive comments, often in the form of asides or afterthoughts. A related finding comes from our semantic valency analysis, which provides at least some tentative evidence that *shm*-reduplication tends to combine with lexemes that are slightly above average in terms of their emotional valency. This hints again at the fact that the construction is pragmatically biased towards addressing concepts that are emotionally relevant to the speaker and/or hearer, and which the speaker wants to express a potentially emotion-laden stance towards. Note, however, that the skew towards positive semantic prosody only concerns the base words that enter into the construction, not the valency of the entire *shm*-instances. Considering the constructional semantics outlined in Section 2, this means that speakers tend to use *shm*-reduplication to express a dismissive stance towards terms that may otherwise invoke positive associations.[7]

On the other hand, our semantic-vector space analysis indicates that the lexemes that undergo *shm*-reduplication belong to a diverse range of semantic fields, including for example food, music, science, technology, and society. Nevertheless, as was also reflected in the collostructional analysis, the construction seems especially productive within the domains of science and technology, the latter also including media- and internet-related terms. This may be partly an artifact of the data we used, since these topics are naturally quite well-represented in web corpora. Still, the high type frequency of the construction in these domains can also be seen as evidence that *shm*-reduplication is a particularly prevalent construction within online communities, where it is applied to concepts that are typically of concern to internet users.

Our corpus analysis also yielded several other interesting findings. One is that *shm*-reduplication can be applied to compound words and phrases, and that there

---

7 Importantly, as a reviewer correctly points out, the pragmatic function of the construction also hinges on its phonological properties. As such, phonological properties constrain the pattern's domain of application – for instance, monosyllabic words emerge as strongly repelled items in the collostructional analysis. One reason for this may be that the rhyming pattern works better with polysyllabic words.

are even some salient cases in which *shm-* is doubly attached to the constituents (e.g., *civil liberties, schmivil schmiberties*). We will discuss this together with other particularly creative examples of the construction in Section 4. Moreover, we identified two collexemes that are particularly strongly associated with *shm-*reduplication: the (highly frequent) *fancy shmancy* and the (somewhat less frequent) *holy shmoly*. Interestingly, both of these display characteristics that are atypical of the rest of the construction. Not only is *fancy shmancy* an adjective, while other adjectival uses of *shm-*reduplication are rare; but it also almost exclusively occurs attributively and predicatively, i.e., in prototypical sentence-internal positions, rather than in separate sentences or at the sentence periphery. Meanwhile, *holy shmoly* lacks the typical dismissive pragmatics of the construction, rather being used as an intensified exclamation that signals the speaker's surprise. It is actually unclear if *holy shmoly* is an instance of the *shm-*reduplication construction as we have defined it here, also given that it is merely one among several playful phonological variations, such as *holy moly*, *holy camoly*, and *holy guacamole*. *fancy shmancy*, on the other hand, typically expresses a depreciative element, characterizing something as "extremely fancy [. . .] in a pretentious or ostentatious way" (OED 2023, "fancy-schmancy"). While it therefore qualifies as an instance of *shm-*reduplication, it may be exactly in virtue of its high type frequency that *fancy shmancy* has developed a distinct syntactic profile from the rest of the construction. As a result, *fancy shmancy* can potentially be regarded as a subconstruction in its own right, which might serve as a salient analogical model for other adjectival *shm-*reduplicants – a point that we will return to in Section 4.

4 The Interplay of Creativity and Routine

The results of our corpus study illustrate the close interplay between aspects of linguistic *creativity* and *routine* in sanctioning speakers' use of *shm-*reduplication. We will explore these implications now in the broader context of concepts and ideas that have been developed in the recently growing constructionist literature on linguistic creativity (e.g., Bergs 2018, 2019; Hoffmann 2018, 2020; Trousdale 2020; Uhrig 2020).

A useful starting point is Sampson's (2016) distinction between *F-creativity* (for "fixed" creativity) and *E-creativity* (for "enlarging" or "extending" creativity), which has also been put to fruitful use in other constructionist work (e.g., Bergs 2018; Hoffmann 2018). According to Sampson (2016: 19), F-creative activities "characteristically produce examples drawn from a fixed and known (even if infinitely large) range," while E-creative activities "characteristically produce exam-

ples that enlarge our understanding of the range of possible products of the activity". Applied to language, F-creativity is usually understood in terms of the productivity of a conventionalized rule or pattern, thus resembling the traditional Chomskyan use of the term "creativity" (Chomsky 1965; see Bergs 2018: 277–279). E-creativity, on the other hand, involves some notable deviation from the established linguistic norms, thus extending existing patterns in previously non-licensed ways.

The phenomenon of *shm*-reduplication displays several features that can be regarded as E-creative, i.e., deviating from typical linguistic norms. First, word-initial /ʃm/ is an atypical onset in English, occurring only in relatively few, usually loaned words (see Section 2). Second, reduplication as a morphological process is very rare in English, and in Indo-European languages in general (Schwaiger 2015: 478).[8] Third, as our corpus results show, most instances of *shm*-reduplication are not syntactically integrated with the sentence, but they rather form separate (syntactically incomplete) sentences or fill peripheral positions that are structurally marked off from the sentence core. These non-prototypical syntactic positions lend themselves to the specific discourse function of *shm*-reduplication, which is typically used to express dismissive comments, often in the form of asides or afterthoughts.

Together, these features differentiate *shm*-reduplication from the rest of the linguistic system, thus arguably contributing to its E-creative nature. As a result, speakers can use the construction to demonstrate their linguistic skills and thereby attract attention. In this sense, *shm*-reduplication can be seen as an "extravagant" construction, thus illustrating another concept that has gained prominence in (diachronic) constructionist research (e.g., Petré 2016; Neels, Hartmann and Ungerer 2023; Hartmann and Ungerer 2024). Popularized by Haspelmath (1999), extravagance can be defined as speakers' desire to be noticed and stand out from their peers (see also Ungerer and Hartmann 2020). This also fits with the finding from our corpus analysis that *shm*-reduplication shows a slight tendency to involve lexemes with relatively high emotional valence, thus increasing the chances that the expressions are relevant for hearers and catch their attention. As illustrated by *shm*-reduplication, extravagance and E-creativity are often cyclically related: Speakers' creative breach of linguistic norms constitutes a mechanism (though perhaps not the only one) through which expressions achieve extravagant effects; while the social motivation of attracting others' attention provides an explanation for why speakers choose to be creative in the first place.

---

8 As a reviewer points out, however, this mainly pertains to standard registers, while reduplication may be more frequent in non-standard registers.

While *shm*-reduplication appears to be E-creative when compared with other, more canonical constructions, the way in which the pattern is extended to new instances also provides evidence of F-creativity. In particular, the high type frequency together with the results of our semantic analysis indicate that *shm*-reduplication constitutes a productive pattern that is applied across a variety of semantic domains. This suggests that speakers have formed a well-entrenched *shm*-schema that can be used to license new instances in flexible but ultimately predictable (and thus F-creative) ways. The relatively homogeneous nature of these instances is highlighted by the fact that most of them adhere to the specific syntactic and pragmatic characteristics of the construction (i.e., occurring in non-integrated positions and expressing dismissive comments). Interestingly, therefore, the narrow syntactic and pragmatic profile that underlies the pattern's E-creativity relative to other constructions (see above) also seems to scaffold its F-creativity by increasing the internal consistency among its instances and thus arguably facilitating the pattern's productive extensions. F-creativity thus illustrates the role that routines – perhaps more appropriately termed *creative routines* in this context – play in *shm*-reduplication, where the repeated use of an originally E-creative pattern leads to the formation of an increasingly well-entrenched schema in speakers' minds. The fact that an increase in the routine-like nature of *shm*-reduplication may, over time, lead to a decrease in its perceived E-creativity is in line with previous observations that language change often involves a cyclical interaction between speakers' desire for extravagance and conformity (Haspelmath 1999; Neels, Hartmann and Ungerer 2023).

Of course, there are still some pockets of E-creativity left *within* the *shm*-reduplication construction: Certain exemplars definitely appear more creative than others. For example, our analysis indicates that *shm-* is sometimes applied to both constituents of compound words (e.g., *impulse buying, schmimpulse schmuying*), which is likely to attract attention. There are also other examples in which *shm*-reduplication is used in particularly playful ways: In (8a), for example, the speaker adds *schmillion* to an enumeration of number words, thus inducing an ad-hoc interpretation of the word as an extremely large numeral while at the same time expressing sarcasm towards the content of the previous sentence. In (8b), the speaker first transforms *dentists* into a made-up form *dontists* before applying *shm*-reduplication, thus increasing the phonological variation and strengthening the dismissive connotation.

(8)   a. *. . . that amount of money out of thin air every week this year.* ***Billion, trillion, quadrillion, schmillion****.* (http://www.housepricecrash.co.uk/ newsblog/2009/11/blog-money-illusion-26316.php)

   b. ***Dentists, dontists, schmontists****, they are all bloody expensive these days* (http://www.sheffieldforum.co.uk/archive/index.php/t-179797.html)

On the other hand, there are also further signs that elements of routine are at work in *shm*-reduplication. First, while the pattern is semantically relatively unconstrained, our vector-space analysis nevertheless points to some domains in which the construction is particularly frequent, for example internet- and media-related concepts. Together with the informal impression we gained during our analysis that many of our examples stem from internet forums, i.e., semi-private subcommunities, this suggests that computer-mediated communication serves not only as a social context in which *shm*-reduplication is used particularly prolifically but also as a topic to which the construction is applied. The creation of such sociofunctional "niches" may contribute to the formation of conversational routines that boost the construction's productivity. Second, our analysis provided evidence of two particularly frequent types, *fancy shmancy* and *holy shmoly*. These may arguably have the status of mini-constructions in their own right, especially given that they diverge from some of the construction's prototypical characteristics (with *fancy shmancy* displaying the syntactic behavior of a typical adjective, and *holy shmoly* lacking the dismissive connotation). Nevertheless, they form salient routines that may potentially scaffold the creation of new *shm*-instances via analogical extension (see, e.g., De Smet 2012: 8). For example, *fancy shmancy* may serve as a model for other attributive and predicative uses of *shm*-reduplicated adjectives.

5 Conclusion

In this paper, we have presented a corpus-based analysis of *shm*-reduplication based on a comparatively large sample drawn from the ENCOW corpus. *shm*-reduplication can be seen as a paradigm example of an "extravagant" pattern at the interface of word-formation and syntax. We have focused on the morphosyntactic usage of the construction as well as on its semantic characteristics. Regarding the former, we have shown that *shm*-reduplication is usually syntactically non-integrated, most prototypically occurring in separate sentences. The most frequent instance *fancy shmancy* makes up the bulk of attributive and predicate uses, which suggests that it can be considered a construction in its own right (similarly to the second most frequent instance, *holy shmoly*). Regarding the semantics of *shm*-reduplication, we have shown that the construction is not constrained to any particular semantic domain, even though it combines particularly frequently with terms from the socio-political realm, science, and technology. In terms of its semantic prosody, the pattern combines with lexemes across the emotional valence scale, even though it shows a slight preference for words with positive emotional valence.

Based on these characteristics, the pattern behaves quite similar to other extravagant constructions such as "snowclones" like *X is the new Y* or *the mother of all X* (Hartmann and Ungerer 2024), which also display preferences for some semantic domains but are still open for a semantically virtually unconstrained inventory of slot fillers. In addition, the productivity pattern of *shm*-reduplication shows some similarities to these snowclones as well as to extravagant word-formation patterns like pseudo-participles (e.g., *bebrillt* 'be-glass-ed', see Kempf and Hartmann 2022): Apart from the outliers *holy shmoly* and especially *fancy shmancy* with many attestations, most instantiations of the pattern are attested only a few times, or even just once. This indicates that the instances of the construction are usually ad-hoc coinages that are strongly tied to the specific contexts in which they are used. The latter usage profile may be a common characteristic of linguistic patterns that can be characterized as extravagant, i.e., that are somehow unusual and salient and thus, at least partly, serve the purpose of gaining the hearers' attention.

Whether or not this is the case would be an interesting question for a potential follow-up study that systematically compares a set of extravagant constructions across different domains (e.g., morphology, syntax, phraseology) and ideally across different languages. A contrastive perspective could be particularly fruitful for *shm*-reduplication, as the pattern is attested in other languages (e.g., German) as well. A comparison with the above-mentioned *m*-reduplication could also prove insightful, especially given that the semantic domains of both constructions overlap but are not fully congruent. A further potentially promising way of extending the analysis of *shm*-reduplication could be to add a multimodal perspective: Given its dismissive semantics, we could expect that in spoken language, it is often combined with dismissive gestures such as members of the family of AWAY gestures (Bressem and Müller 2014).[9] In addition, given our assumption that *shm*-reduplication is a conceptually more oral phenomenon, taking multimodal data into account can potentially provide more insights into the use of *shm*-reduplication in communicative interaction.

These examples show that there is still a lot of uncharted territory to explore when it comes to the interplay of creativity and routine in the use of *shm*-reduplication, and of expressive patterns in word formation and syntax in general. With this paper, we hope to have shed some new light on the use of *shm*-reduplication in authentic (written) data, thus laying the groundwork for follow-up studies of this pattern and related phenomena.

---

9 A pilot study based on data from the TV News Archive (https://archive.org/details/tv, accessed 27 February 2024) shows that this is the case in 13 out of 23 attestations.

6 Data Availability

The data, R scripts, and additional plots referred to in the text are available via the Open Science Framework (OSF): https://osf.io/f2cu3/

References

Bergen, Benjamin K. 2004. The psychological reality of phonaesthemes. *Language* 80(2). 290–311. https://doi.org/10.1353/lan.2004.0056.

Bergs, Alexander. 2018. Learn the rules like a pro, so you can break them like an artist (Picasso): Linguistic aberrancy from a constructional perspective. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 277–293. https://doi.org/10.1515/zaa-2018-0025.

Bergs, Alexander. 2019. What, if anything, is linguistic creativity? *Gestalt Theory* 41(2). 173–183. https://doi.org/10.2478/gth-2019-0017.

Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6(1). 213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303.

Booij, Geert E. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Bressem, Jana & Cornelia Müller. 2014. The family of Away gestures: Negation, refusal, and negative assessment. In Cornelia Müller, Alan J. Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Sedinha Tessendorf (eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, 1592–1604. Berlin & Boston: De Gruyter.

Bybee, Joan L., Revere Perkins & William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Cienki, Alan. 2017. Utterance Construction Grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard* 3(s1). https://doi.org/10.1515/lingvan-2016-0048.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

De Smet, Hendrik. 2012. *Spreading Patterns: Diffusional Change in the English System of Complementation*. Oxford: Oxford University Press.

Dekalo, Volodymyr & Beate Hampe. 2017. Networks of meanings: Complementing collostructional analysis by cluster and network analyses. *Yearbook of the German Cognitive Linguistics Association* 5(1). 143–176. https://doi.org/10.1515/gcla-2017-0011.

Donabedian, Anaïd & Ioanna Sitaridou. 2021. Anatolia. In Evangelia Adamou & Yaron Matras (eds.), *The Routledge Handbook of Language Contact*, 404–433. London & New York: Routledge.

Downing, Laura J. & Sharon Inkelas. 2015. What is reduplication? Typology and analysis part 1/2: The typology of reduplication. *Language and Linguistics Compass* 9(12). 516–528. https://doi.org/10.1111/lnc3.12152.

Finkbeiner, Rita. 2012. Naja, normal und normal. Zur Syntax, Semantik und Pragmatik der x-und-x-Konstruktion im Deutschen. *Zeitschrift für Sprachwissenschaft* 31(1). 1–42. https://doi.org/10.1515/zfs-2012-0001.

Finkbeiner, Rita. 2017. "Argumente hin, Argumente her." Regularity and Idiomaticity in German N Hin, N Her. *Journal of Germanic Linguistics* 29(3), 205–258. https://doi.org/10.1017/S1470542716000234.

Finkbeiner, Rita, Jörg Meibauer & Heike Wiese. 2016. What is pejoration, and how can it be expressed in language? In Rita Finkbeiner, Jörg Meibauer & Heike Wiese (eds.), *Pejoration*, 1–18. Amsterdam & Philadelphia: John Benjamins.

Firth, John. 1930. *Speech*. Oxford: Oxford University Press.

Fischer, Olga. 2011. Cognitive iconic grounding of reduplication in language. In Pascal Michelucci, Olga Fischer & Christina Ljungberg (eds.), *Semblance and Signification*, 55–81. Amsterdam/ Philadelphia: John Benjamins.

Frankowsky, Maximilian. 2022. Extravagant expressions denoting quite normal entities: Identical constituent compounds in German. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant Morphology. Studies in Rule-Bending, Pattern-Extending and Theory-Challenging Morphology*, 156–179. Amsterdam & Philadelphia: John Benjamins.

Grandi, Nicola. 2017. Evaluatives in morphology. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.250 (last modified 29 March 2017).

Grandi, Nicola & Livia Körtvélyessy. 2015. Introduction: Why evaluative morphology? In Nicola Grandi & Livia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: Edinburgh University Press.

Gries, Stefan Th. 2023. Overhauling collostructional analysis: Towards more descriptive simplicity and more explanatory adequacy. *Cognitive Semantics* 9(3). 351–386. https://doi.org/10.1163/ 23526416-bja10056.

Grohmann, Kleanthes K. & Andrew Ira Nevins. 2004. Echo reduplication: When too-local movement requires PF-distinctness. *University of Maryland Working Papers in Linguistics* 13. 84–108.

Hartmann, Stefan. 2024. Open Corpus Linguistics – or How to overcome common problems in dealing with corpus data by adopting open research practices. In Mark Kaunisto & Marco Schilk (eds.), *Challenges in Corpus Linguistics: Rethinking Corpus Compilation and Analysis*, 89–105. Amsterdam & Philadelphia: John Benjamins.

Hartmann, Stefan & Tobias Ungerer. 2024. Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns. *Journal of Linguistics* 60(3). 599–634. https://doi.org/10.1017/ S0022226723000117.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068. https://doi.org/10.1515/ling.37.6.1043.

Hoffmann, Thomas. 2018. Creativity and Construction Grammar: Cognitive and psychological issues. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 259–276. https://doi.org/10.1515/zaa-2018-0024.

Hoffmann, Thomas. 2020. Construction grammar and creativity: Evolution, psychology, and cognitive science. *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2018.

Hohenhaus, Peter. 2004. Identical constituent compounding – a corpus-based study. *Folia Linguistica* 38(3–4). https://doi.org/10.1515/flin.2004.38.3-4.297.

Inkelas, Sharon. 2014. Non-concatenative derivation: Reduplication. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 169–189. Oxford: Oxford University Press.

Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication: Doubling in Morphology*. Cambridge: Cambridge University Press.

Jurafsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72(3). 533–578. https://doi.org/10.2307/416278.

Kempf, Luise & Stefan Hartmann. 2022. What's extravagant about *be-sandal-ed feet*? Morphology, semantics and pragmatics of German pseudo-participles. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant Morphology: Studies in Rule-Bending, Pattern-Extending and Theory-Challenging Morphology*, 19–50. Amsterdam & Philadelphia: John Benjamins.

Kentner, Gerrit. 2017. On the emergence of reduplication in German morphophonology. *Zeitschrift für Sprachwissenschaft* 36(2). 233–277. https://doi.org/10.1515/zfs-2017-0010.

Kentner, Gerrit. 2022. do not repeat: Repetition and reduplication in German revisited. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant Morphology: Studies in Rule-Bending, Pattern-Extending and Theory-Challenging Morphology*, 182–205. Amsterdam: John Benjamins.

Kentner, Gerrit. 2023. Reduplication as expressive morphology in German. In J. P. Williams (ed.), *Expressivity in European Languages*, 1st edn., 103–120. Cambridge: Cambridge University Press.

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36. 15–43. https://doi.org/10.1515/9783110244922.15.

Kołłątaj, Andrzej. 2016. Reduplication in English – typology, correlation with slang and metaphorisation. *Philolog. Studia Neofilologiczne* 6. 237–248. https://doi.org/10.34858/polilog.6.2016.021.

Kuperman, Victor, Zachary Estes, Marc Brysbaert & Amy Beth Warriner. 2014. Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General* 143(3). 1065–1081. https://doi.org/10.1037/a0035669.

Lensch, Anke. 2018. Fixer-uppers. Reduplication in the derivation of phrasal verbs. In Rita Finkbeiner & Ulrike Freywald (eds.), *Exact Repetition in Grammar and Discourse*, 158–181. Berlin & Boston: De Gruyter.

Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10(8). 707–710.

Levshina, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam/Philadelphia: John Benjamins.

Maaten, Laurens van der & Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9(86). 2579–2605.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert & Kurt Hornik. 2023. cluster: Cluster analysis basics and extensions. Manual. https://CRAN.R-project.org/package=cluster.

Mattiello, Elisa. 2013. *Extra-Grammatical Morphology in English: Abbreviations, Blends, Reduplicatives, and Related Phenomena*. Berlin & Boston: De Gruyter.

McCarthy, John & Alan Prince. 1986. Prosodic morphology. https://hdl.handle.net/20.500.14394/32398.

Meibauer, Jörg. 2013. Expressive compounds in German. *Word Structure* 6(1). 21–42. https://doi.org/10.3366/word.2013.0034.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado & Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In Christopher J. Burges, Leon Bottou, Max Welling, Zoubin Ghahramani & Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 3111–3119. Red Hook, NY: Curran Associates.

Minkova, Donka. 2002. Ablaut reduplication in English: The criss-crossing of prosody and verbal art. *English Language & Linguistics* 6(1). 133–169. https://doi.org/10.1017/S1360674302001077.

Nagaya, Naonori. 2020. Reduplication and repetition from a constructionist perspective. *Belgian Journal of Linguistics* 34. 259–272. https://doi.org/10.1075/bjl.00051.nag.

Neels, Jakob, Stefan Hartmann & Tobias Ungerer. 2023. A quantum of salience: Reconsidering the role of extravagance in grammaticalization. In Hendrik De Smet, Peter Petré & Benedikt Szmrecsanyi (eds.), *Context, Intent and Variation in Grammaticalization*, 47–78. Berlin & Boston: De Gruyter.

Nevins, Andrew & Bert Vaux. 2003. Metalinguistic, shmetalinguistic: The phonology of shm-reduplication. *Chicago Linguistic Society* 39(1). 702–721.

Norde, Muriel & Sarah Sippach. 2019. *Nerdalicious scientainment*: A network analysis of English libfixes. *Word Structure* 12(3). 353–384. https://doi.org/10.3366/word.2019.0153.

OED. 2023. *Oxford English Dictionary*. https://oed.com

Paley, Andrew. 2019. "Seeing the world as it is: The importance of honest sight." In *Texas Jewish Post*. https://tjpnews.com/seeing-the-world-as-it-is-the-importance-of-honest-sight (last modified 22 July 2019).

Pedersen, Thomas Lin. 2024. ggforce: Accelerating "ggplot2". Manual. https://CRAN.R-project.org/package=ggforce.

Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1). 149–188. https://doi.org/10.1515/ling-2015-0043.

Petré, Peter. 2016. Unidirectionality as a cycle of convention and innovation: Micro-changes in the grammaticalization of [be *going to* INF]. *Belgian Journal of Linguistics* 30. 115–146. https://doi.org/10.1075/bjl.30.06pet

R Core Team. 2023. R: A language and environment for statistical computing. Manual. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

Regier, Terry. 1994. A preliminary study of the semantics of reduplication. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=183743bf7da4da277d9ddf2f4aa1d34a58e110f8 (accessed 27 February 2024).

Rubino, Carl. 2013. Reduplication. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/27.

Saba Kirchner, Jesse. 2010. *Minimal reduplication*. Santa Cruz, CA: University of California, Santa Cruz dissertation. https://doi.org/doi:10.7282/T3VQ31K8.

Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, Experiment and Argument in Linguistics and Philosophy of Language*, 15–26. Bern: Peter Lang.

Schäfer, Roland. 2015. Processing and querying large corpora with the COW14 architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lüngen & Andreas Witt (eds.), *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3)*, 28–34.

Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Nicoletta Calzolari, Khalid Choukri, Terry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 486–493. Istanbul: European Language Resources Association.

Schlechtweg, Marcel. 2018. *Memorization and the Compound-Phrase Distinction: An Investigation of Complex Constructions in German, French and English*. Berlin & Boston: De Gruyter.

Schneider, Ulrike. 2020. ΔP as a measure of collocation strength: Considerations based on analyses of hesitation placement in spontaneous speech. *Corpus Linguistics and Linguistic Theory* 26(2). 249–274. https://doi.org/10.1515/cllt-2017-0036.

Schmidt, Ben & Jian Li. 2022. wordVectors: Tools for creating and analyzing vector-space models of texts. Manual. http://github.com/bmschmidt/wordVectors.

Schwaiger, Thomas. 2015. Reduplication. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word Formation: An International Handbook of the Languages of Europe*, vol. 1, 467–484. Berlin & New York: De Gruyter.

Schwaiger, Thomas. 2018. The derivational nature of reduplication and its relation to boundary phenomena. In Rita Finkbeiner & Ulrike Freywald (eds.), *Exact Repetition in Grammar and Discourse*, 67–88. Berlin & Boston: De Gruyter.

Southern, Mark R. V. 2005. *Contagious Couplings: Transmission of Expressives in Yiddish Echo Phrases*. Westport: Praeger.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243. https://doi.org/10.1075/ijcl.8.2.03ste.

Stroebel, Liane. 2017. Can Macromania be explained linguistically? Beneath the morphological boundary: A sketch of subconscious manipulation strategies in Emmanuel Macron's political discourses. *Yearbook of the German Cognitive Linguistics Association* 5(1). 57–76. https://doi.org/10.1515/gcla-2017-0005.

Trousdale, Graeme. 2020. Creativity, reuse, and regularity in music and language. *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2021.

Uhrig, Peter. 2020. Creative intentions – The fine line between 'creative' and 'wrong.' *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2027.

Ungerer, Tobias & Stefan Hartmann. 2020. Delineating extravagance: Assessing speakers' perceptions of imaginative constructional patterns. *Belgian Journal of Linguistics* 34. 345–356. https://doi.org/10.1075/bjl.00058.ung.

Ungerer, Tobias & Stefan Hartmann. 2023. *Constructionist Approaches: Past, Present, Future*. Cambridge: Cambridge University Press.

Warriner, Amy Beth, Victor Kuperman & Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45(4). 1191–1207. https://doi.org/10.3758/s13428-012-0314-x.

Weinreich, Max. 1980. *History of the Yiddish Language*. Chicago: Chicago University Press.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*, 2nd edn. Berlin: Springer.

Wiese, Heike & NilginTanış Polat. 2016. Pejoration in contact: m-reduplication and other examples from urban German. In Rita Finkbeiner, Jörg Meibauer & Heike Wiese (eds.), *Pejoration*, 243–267. Amsterdam & Philadelphia: John Benjamins.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1987. Plain morphology and expressive morphology. *Annual Meeting of the Berkeley Linguistics Society* 13. 330–340. https://doi.org/10.3765/bls.v13i0.1817.

Pedro Ivorra Ordines & Carmen Mellado Blanco

Semi-Schematic Patterns and their Social Dimension: A Constructionist Study of the Proverb *Dime con quién andas y te diré quién eres* in Spanish and German

Abstract: Proverbs, as lexically filled phrasemes (also known as substantive phrasemes), offer a new research pathway in the opposition canonical vs non-canonical variants. In the context of Construction Grammar, the role of analogical extensions is of special interest, conceived as a trigger of productivity and schematicity. Against this background, the aim of this paper is twofold. First, to adopt a constructionist approach to the study of the Spanish proverb *Dime con quién andas y te diré quién eres* 'A man is known by the company he keeps' and its prototypical German equivalents *Sage(e) mir, mit wem du umgehst und ich sage dir, wer du bist* and *Zeig (e) mir deine Freunde, und ich sage dir, wer du bist.* To do so, we conducted a corpus-based study (esTenTen18 and deTenTen20, Sketch Engine) with the hypothesis that in both languages, there is a greater quantitative presence of modified instances of the canonical form(s) compared to the canonical uses. This allows to identify the canonical form or forms in each language, as well as to establish the different degree of variability of the proverbs in Spanish and German and to determine which semantic clusters typically tend to occupy the X slot. Second, to explore if the traditional social purpose seen in standard usage of proverbs remains applicable when changes are made, or if new communicative or persuasive functions associated with textual genres in which the non-canonical forms occur can be identified.

1 Introduction

Proverbs, as lexically filled phrasemes (also called substantive phrasemes), and snowclones[1] present a new avenue of research when it comes to the opposition of canonical vs non-canonical variants, i.e. between the citation form of a fixed expression and the innovations deviating from such a canonical form (Hartmann and Ungerer 2023; Ivorra Ordines in press; Mellado Blanco 2024). In the context of Construction Grammar, the role of analogical extensions is particularly intriguing, as they are seen as triggers for productivity and schematicity. Regarding phraseme modification, we align with other works (Mellado Blanco 2020a: 25) in adopting a constructionist perspective to analogy because it can effectively account for: (i) the replacement of certain lexical items in idiomatic expressions with others in discourse as with the constructional idioms [(*no*) PRON *importar* NP[DET N[taboo object or of little value]] ⇔ 'not care at all' (e.g., *no me importa un pepino/comino/carajo* 'I don't give a damn') (Mellado Blanco 2020b) or [DET ADJ *de* DET[possessive] N] ⇔ 'critics or insult with respect to the propositional content' (e.g., *el inútil de su hijo* 'the useless of his son') (Esteban-Fonollosa 2023), (ii) the potential of creativity as with the constructional idiom of vehement rejection [*por* PRON *como si* X] ⇔ 'I don't care at all' (e.g., *por mí como si comes alimentos que no dan sombra*, lit. 'for me as if you eat food that doesn't provide shade') (Ivorra Ordines 2023), (iii) the development of new phrasemes based on existing ones as with the constructional idiom *estar a un click de* 'to be one click away of' (Mellado Blanco 2023), and (iv) the emergence of semi-schematic patterns as with the semi-schematic construction [*kein*(*e*) N$_1$ *ist auch ein*(*e*) N$_1$] ⇔ 'no N1 is also a N1' (e.g., *keine Reaktion ist auch eine Reaktion* 'no reaction is also a reaction') (Mollica 2018) (see Ivorra Ordines 2023: 111; Mellado Blanco 2022: 9; Mellado Blanco and Ivorra Ordines 2023: 26; Mellado Blanco, Mollica and Schafroth 2022: 6; Rasulic 2010; Stumpf 2018: 171, 2022: 180).

This paper is structured as follows. Section 2 discusses the interplay of form, meaning and function in both well-established constructions and idiomatic expressions, acknowledging the significance of context and genre in linguistic anal-

---

[1] In the early 2000s, the term 'snowclone' emerged as a result of a naming contest initiated by Geoff Pullum on the linguistics blog *Language Log* in 2004. Pullum urged the community to come up with a suitable term to describe a "a multi-use, customizable, instantly recognizable, timeworn, quoted or misquoted phrase or sentence that can be used in an entirely open array of different jokey variants by lazy journalists and writers" (Pullum 2003; see Ivorra Ordines in press; Mellado Blanco 2024). Following Hartmann and Ungerer (2023), snowclones are characterized by (i) the existence of a lexically filled source construction, (ii) partial productivity, and (iii) extravagant formal and/or functional characteristics.

ysis –social meaning, i.e., the social and/or stylistic significance (see Leclercq and Morin 2023: 10)– in the constructional description, and recognizing that genres themselves can be seen as conventionalized linguistic structures, thus expanding the scope of conventional pragmatics within grammar. The first objective of this work is tackled in Section 3, that is, a constructionist analysis of the Spanish proverb *Dime con quién andas y te dire quién eres* (lit. 'Tell me who you walk with and I will tell you who you are') 'A man is known by the company he keeps' and its German equivalent *Sage(e) mir, mit wem du umgehst und ich sage dir, wer du bist*. For this purpose, we started with the hypothesis that in both languages, the modified uses of the canonical form(s) quantitatively outnumber the canonical uses. To validate the hypothesis, we conducted a corpus-based analysis using esTenTen18 corpus for Spanish and deTenTen20 corpus for German (both from Sketch Engine). The second aim is discussed in Section 4. Through the corpus study, we aim to gain insights into identifying the canonical form of forms in each language. Additionally, we seek to establish the degree of variability within the proverbs and discern which semantic clusters typically fill the X slot. In this context, another sub-aim is to investigate whether the usual social function observed in canonical uses of the proverb remains valid in instances that undergo modifications or whether new illocutionary or perlocutionary functions linked to the textual genres in which the modifications appear can be detected. Section 5 summarizes the main findings.

2 Theoretical Background: Constructions all the Way and its Relation to Genre

The constructionist perspective to language seeks to comprehensively explain linguistic data within a language while rejecting the transformational aspect of the generativist tradition. This approach, exemplified by studies such as Fillmore, Kay and O'Connor (1988) and Kay and Fillmore (1999), is considered a maximalist theory, even though not all constructions in a language's grammar are identified. These studies aimed to highlight structures previously overlooked due to their irregularity and apparent lack of productivity. In a constructionist approach, both well-established constructions and peripheral elements like stock phrases and idioms are equally important. This perspective abolishes the center-periphery distinction and focuses on language as a whole, shedding light on structures that do not conform to logical statements, and often carry additional pragmatic nuances. For instance, the *let alone* construction is sensitive to pragmatic considerations, such as Gricean maxims like providing sufficient information and maintaining

relevance. The analysis of constructions incorporates pragmatics into the grammatical theory, addressing the question of how speakers' understanding of context is systematic and conventional, making it an inherent part of grammar and constructional description.

The theoretical status of grammatical constructions – as flexible resources (Östman 2015) – is emphasized in this context, that is, conventionalized pairings of meaning and form, in articulating both language-specific and cross-linguistic generalizations, effectively explaining both regular and semi-schematic patterns. The meaning –the semantic pole of any construction–, whether it is related to words or sentence structure, is defined with reference to frames. In this context, Construction Grammar encompasses Frame Semantics and draws from valuable insights found in the frame semantic theory (as highlighted by Fillmore 1982, 1985; Ziem 2014; among others). Consequently, information related to pragmatic, discourse, textual, or register features associated with a particular linguistic form can be incorporated into the meaning aspect of the corresponding construction, alongside purely semantic information (as mentioned by Fried and Östman 2004; Goldberg 1995, 2006, 2019; Nikiforidou 2015, 2016, 2018; Östman 2005; among others). The meaning pole, for its part, stands for "all the conventionalized aspects of a construction's function, which may include not only properties of the situation described, but also properties of the discourse in which the utterance is found [. . .] and of the pragmatic situation of the interlocutors" (Croft and Cruse 2004: 258). In harmony with this reasoning, "constructions are abstract form-meaning-function constellations; form, meaning and function are inseparable in usage" (Östman 2015: 20).

In this context, the emphasis of construction grammarians on meaning and their exploration of encoding idiomaticity has inevitably brought Construction Grammar close to the phraseological tradition, as can be attested in the numerous contributions in the last few years (Ivorra Ordines 2021; Mellado Blanco 2022; Mellado Blanco, Mollica and Schafroth 2022; Mellado Blanco, Ivorra Ordines and Esteban Fonollosa 2024). With encoding idioms, in contrast to decoding idioms, we are dealing with expressions that language users may or may not comprehend without prior exposure, but they would not recognize it as a conventional[2] way of expressing its intended meaning (Fillmore, Kay and O'Connor 1988: 504–505; see

---

2 Identifying thematic regularities as constructional characteristics undeniably constitutes an expansion of the existing approach, although it is a well-founded expansion, as these characteristics are conventionally part of the structure. Conventionality, indeed, "is the most important, in fact necessary, feature for assigning idiomatic/constructional status" (Antonopoulou and Nikiforidou 2011: 2596). On top of that, taking discoursal properties into constructional description avoids any criticism on cognitive theories ignoring social or discoursal properties (see Geeraerts 2005; Langacker 2008).

Croft and Cruse 2004: 225 or Gries 2008: 14). To put it differently, while these expressions conform to the grammatical and semantic rules of a language, still exhibit idiomatic characteristics by linking a specific form to a particular meaning.

As evident from the discussion in the previous section, the incorporation of encoding idiomaticity can apply to various types of formal-pragmatic combinations, significantly broadening the scope of conventional pragmatics and, consequently, constructional pragmatics. In this sense, constructional analysis has arguably expanded its scope to include conventionalized discourse phenomena into the realm of grammar. The arguments for linking specific forms to identifiable contexts is in harmony with two main factors (Nikiforidou 2018): (i) there are frequency-based considerations, where a particular formal pattern may be more prevalent in certain contexts than in others, and (ii) there are stricter criteria, such as the impossibility of a pattern (or some of its features) appearing outside of specific genre contexts. While the latter criteria might provide a more compelling argument for including genre specifications in grammatical descriptions, the former aligns well with usage-based frameworks and are naturally encountered in broader contexts like registers.

Such step is justified by one of Charles Fillmore's fundamental principles regarding Construction Grammar, which asserts that a constructionist approach should align with our understanding of cognition and social interaction, and that textual analysis should not be dismissed from the outset (Antonopoulou and Nikiforidou 2011; Nikiforidou 2009, 2010, 2018, 2021; Nikiforidou and Fischer 2015; Östman 2005, 2015; Östman and Fried 2005; among others). This involves addressing and potentially formalizing the systematicity and regularity of discourse patterns, particularly focusing on their cognitive foundations that shape the interpretation of all related texts. It becomes crucial to acknowledge such conventional associations when dealing with discourse-based-licensed patterns that cannot be derived from other language constructions and might even be considered theoretically implausible. The most primary obstacle when attempting to expand constructional analysis beyond the sentence lies in the way of handling context systematically while ensuring it remains plausible.

Previous attempts to address genre within the framework of Construction Grammar have explored three distinct approaches (see Nikiforidou 2019; Nikiforidou and Fischer 2015). The first approach involves incorporating genre-related information as a feature within the grammatical description. For example, Nikiforidou (2015) conceived empathetic narration as a unique form of narration that deviates from the typical past narrative, which is usually tied to the narrator's perspective and the context of conversation. The second approach entails providing a detailed description of how a particular situation enables specific language functions. In this case, the connection to a genre is not an inherent part of the construc-

tion itself; instead, it emerges from the functional requirements of the situation. For example, Matsumoto (2015) discusses that the stand-alone construction in Japanese is genre-sensitive, in that the variation with a noun (referred to as the stand-alone noun-modifying construction) is closely linked to specific contexts of verbal activities and practices. The third approach involves describing genres as complex signs. This approach is relevant when examining recipes (Östman 2005), football chants (Hoffmann 2015; Hoffmann and Bergs 2018), Instagram posts (Fried and Aarestrup 2021), Knock Knock jokes (Hoffmann and Bergs 2023), among others, where the components and structure have been conventionalized. Taking this last approach even further, Hoffmann and Bergs (2018, 2023) follow up the idea of "constructions all the way *down*" (Goldberg 2006: 18) and, on top of that, add that it is also constructions all the way *up*, since even larger linguistic structures such as genres can be conceived as constructional templates as long as they are frequent enough (see Nikiforidou 2018 for a discussion).

3 Corpus Extraction and the Notion of Cognitive Canonical form

In the approaches to the phenomenon of variation, the notion of *base form* has been quite prominent because it allows to distinguish what is fixed in a phraseme from what is updated in discourse according to syntactic idiosyncrasies. A widely used term to refer to this base form in the Spanish phraseological tradition is *forma citativa* 'quotative form' (Wotjak 1998) or *forma canónica* 'canonical form' (Mellado Blanco 2020a), terms that refer to the paradigmatic form that a phraseme should have outside of context and typically appears as the lemma in dictionaries. Indeed, this is a fairly common practice in the study of phraseological variation, as expressed by Langlotz:

For purely practical reasons, I will equate a given base-form with the idiom's citation-form in idiom dictionaries. I take it for granted that lexicographic practice attempts to record only highly familiar lexicalised constructions belonging to the langue of a given variety (i.e. those units that are entrenched in the mental lexicons of most speakers). Dictionary citation-forms therefore approximate the present view of a usage-based default construction. (Langlotz 2006: 178; italics in the original)

According to the results obtained from the corpus analysis[3], the Spanish proverb shows a clear canonical form, that is, *Dime con quién andas y te dire quién eres*, not only because it is the form that is usually encountered in phraseological dictionaries (see DFDEA or *Refranero Multilingüe*[4]), but also because it is the most frequent type in the corpus with 497 instances (Pfeiffer 2017: 26; see also Petrova 2011). From the corpus esTenTen18, we started with the lexically filled second part of the proverb, which is *y te diré quién eres* 'and I will tell you who you are', together with the filtering of the lemma *decir* 'tell' KWIC -10 -1. The search yielded a total amount of 1111 instances, taking also into consideration those constructs that belong to the canonical form. Such a query would lead to the semi-schematic pattern [*Dime* X$_{[indirect\ interrogative\ clause]}$ *y te diré quién eres*] ⇔ 'people are defined by the entity referred to in X'. Other possibilities in terms of queries would involve leaving *y te diré* 'and I will tell you' lexically specified, offering two potential slots. This would result in a more abstract construction[5], which could have a notable impact on its degree of productivity, schematicity and entrenchment (see Goldberg 2006, 2019).

Out of the 1111 instances, 497 instances appear in the canonical form and 614 occurrences in the modified form, i.e. 55.26%. Adopting the productivity cline proposed by Barðdal (2008), token frequency is a relevant factor for speaker's choice of model items in lower-level constructions. This reasoning would explain how the canonical, and highly frequent type *con quién andas* 'who you walk with', is responsible for the construction's extensibility, i.e., the productivity of a slot conceived as the likelihood that a slot of a construction must be renewed with a new lexical item based on such highly frequent type (Goldberg 2019: 60). Based on the friendships that people cultivate (semantic constraint), we identified 35 analogical extensions (80 occurrences). Among these, *con quién vas* 'who you go with' appears 19 times, *con quién te juntas* 'who you hang out with' appears 14 times, and *quién es tu amigo* 'who your friend is' appears 4 times, as the most common analogical extensions

---

3 For our study, we used the macro-corpora esTenTen18 and deTenTen20 (Sketch Engine) for Spanish and German, respectively.

4 This resource is a distinctive multilingual compilation of popular proverbs in Spanish, offering equivalents in various languages, including German, Catalan, French, Galician, Ancient Greek, Modern Greek, English, Italian, Portuguese, Russian, and Basque. This includes potential variations and synonyms, as well as hypernyms and proverbial antonyms. You can access this resource at https://cvc.cervantes.es/lengua/refranero/.

5 We are referring to a more schematic construction such as [X *y te diré* Y] ⇔ 'X determines or has a great influence on Y'; e.g., "**Cuéntame de qué presumes y te diré de que careces**" '**Inform me on your boosts and I shall reveal your lacks**' (esTenTen18, 66327358). For reasons of purpose of this contrastive study, this more schematic form will not be taken into account. This may be the subject of future studies.

(see examples 1-2). The slot productivity is further evidenced by the significant number of unique instances (hapax legomena), with 300 nonce instances out of 1111 occurrences (27% of potential productivity; see Baayen 2009: 209).

(1) **Dime quiénes son tus amigos y te diré quién eres**. Todo hombre se alía con su propia imagen, y se aparta de su disimilitud. (esTenTen18, 8919352416)
   '**Tell me who your friends are**, and I will tell you who you are. Every man allies himself with his own image, and turns away from his dissimilarity.'

(2) Conocer personas diferentes a tu círculo: al parafrasear el dicho popular **dime con quién estás y te diré quién eres,** se expresa con mayor claridad que el conocimiento y los contactos son determinantes para el futuro. (es-TenTen18, 9426829755)
   'Getting to know people outside your circle: paraphrasing the popular saying "**tell me who you are with I will tell you who you are**", it becomes clearer that knowledge and contacts are decisive for the future.'

The German proverb, for its part, presents a canonical form with significant formal variability, making it challenging to determine exactly what the prototypical form is. Furthermore, if there is no evidence of a single, commonly recognized canonical form, the concept of modification loses strength and is difficult to implement (see Pfeiffer 2017). Using the deTenTen20 corpus (Sketch Engine), we started with the most frequent form of the second part of the proverb, which is *und ich sage dir, wer du bist* 'and I will tell you who you are'[6]. The number of occurrences of this search is 810. In addition to the sequence *un dich sage dir, wer du bist* 'and I will tell you who you are', we have observed similar forms such as *dann sage ich dir, wer du bist* 'then I will tell you who you are' and, to a lesser extent, *und weiß, wer du bist* 'and I know who you are', which we have omitted from the analysis for the sake of simplification. The analysis based on the sequence *und ich sage dir, wer du bist* 'and I will tell you who you are', which primarily aims to decipher the variability and creativity in indirect interrogative sentences, is just one of the possible approaches that could have been taken. It would also be possible to study variability in both clauses by searching for *und*

---

6 In this case, the comma is included in the search. Occurrences without a comma, that is, the phrase *und ich sage dir wer du bist* have not been considered for the study (amounting to 201 occurrences).

ich sage dir 'and I will tell' to describe the pattern [X, *und ich sage dir*, Y]. Other potential semi-schematic patterns that could be the subject of study in future research include [*Sage mir* X, *und ich sage dir*, Y] 'tell me X and I will tell you Y' or [*Zeige mir* X, *und ich sage dir*, Y] 'show me X and I will tell you Y'[7].

Out of 810 occurrences (462 with *sagen* 'to tell', 348 with *zeigen* 'to show'), 612 instances exhibit modifications in the first part of the proverb (75.56%) and barely a quarter (198 occurrences, 24.44%) appear in a canonical form, which we will refer to as flexible because they all have the same meaning, paraphrased as 'people are defined by the friendships they cultivate', and in most cases, the same moralizing intention. The four most common variants are:

(1) *Sag*(*e*) *mir*(,) *mit wem du umgehst*(,) *und ich sage dir, wer du bist* 'Tell me who you associate with, and I will tell you who you are' (49 occurrences).

(2) *Zeig*(*e*) *mir deine Freunde*(,) *und ich sage dir, wer du bist* 'Tell me your friends and I will tell you who you are' (39 occurrences).

(3) *Sag*(*e*) *mir*(,) *mit wem du gehst*(,) *und ich sage dir, wer du bist* 'Tell me who you go with, and I will tell you who you are' (22 occurrences).

(4) *Sag*(*e*) *mir*(,) *wer deine Freunde sind*(,) *und ich sage dir, wer du bist* 'Tell me who your friends are, and I will tell you who you are' (18 occurrences).

There are other constructs that can be also considered part of the flexible canonical form, for example:

– *Sag*(*e*) *mir*(,) *mit wem du verkehrst*(,) *und ich sage dir, wer du bist* 'Tell me who you socialize with, and I will tell you who you are' (5 occurrences).

– *Sag*(*e*) *mir*(,) *mit wem du dich umgibst*(,) *und ich sage dir, wer du bist* 'Tell me who you are surrounded yourself with, and I will tell you who you are' (5 occurrences).

– *Sag*(*e*) *mir*(,) *wer dein Freund ist*(,) *und ich sage dir, wer du bist* 'Tell me who your friend is, and I will tell you who you are' (4 occurrences).

– *Sag*(*e*) *mir*(,) *mit wem du Freund bist*(,) *und ich sage dir, wer du bist* 'Tell me who you are friends with, and I will tell you who you are' (2 occurrences).

– *Sag*(*e*) *mir deine Freunde*(,) *und ich sage dir, wer du bist* 'Tell me about your friends, and I will tell you who you are' (2 occurrences).

---

7 Outside the scope of our study are the occurrences in which the main verb in the first part of the proverb is neither *sagen* 'to tell' nor *zeigen* 'to show', of the following type: "**Verkleide dich und ich sage dir, wer du bist.** Oder so ähnlich. Im Maskworld Onlineshop bekommst du Glitzerfummel, Superheldenkostüme, Perücken [. . .]." '**Dress up, and I will tell you who you are**. Or something like that. In the Maskworld online shop, you can find glittery outfits, superhero costumes, wigs, and more.' (deTenTen20, 18423047636)

– *Zeig(e) mir dein Umfeld(,) und ich sage dir, wer du bist* 'Show me your environment, and I will tell you who you are' (2 occurrences).

We consider that in these occurrences, there is no modification because the original meaning of the proverb prevails, and therefore, there is no creative intent on the part of the speaker. To put is simply, the constructs presented above rely on two distinct patterns, one with an indirect interrogative sentence and another with a direct object NP in the first part of the proverb:

i.   *Sag(e) mir* + indirect interrogative clause + *und ich sage dir, wer du bist*.
ii. *Zeig(e) mir* + NP + *und ich sage dir, wer du bist*.

Within the 348 occurrences with *zeigen* 'to show', only 41 instances can be considered canonical, accounting for just 11.2% within the pattern *Zeig(e) mir* + NP + *und ich sage dir wer du bist* 'show me + NP + and I will tell you who you are'. In contrast, in the pattern with *sagen* 'to tell', the percentage of the canonical occurrences is 33.9%, leading us to conclude that the allostruction [*Zeig(e) mir* NP *und ich sage dir, wer du bist*] stands out due to a very high degree of entrenchment, given the low number of canonical constructs it licenses compared to non-canonical ones (Bybee 2013).

Out of the 198 canonical occurrences, 25 of them deviate further from the prototype because they exhibit a certain playful intent and creativity (see examples 3–4), or they feature less common vocabulary, such as *Sag(e) mir, mit wem du Umgang pflegst, und ich sage dir, wer du bist* 'Tell me who you associate with, and I will tell you who you are'. These 25 occurrences are all unique and can be placed along a continuum between the classic canonical forms and the modified ones.

(3)   **Zeige mir deine 5 besten Freunde und ich sage dir, wer du bist**. Hier gilt das Gesetz des Durchschnitts der 5 besten Freunde. (deTenTen20, 16218779753) '**Show me your 5 best friends, and I will tell you who you are**. Here, the law of the average of the 5 best friends applies.'

(4)   Da beißt sich die Katze in den Schwanz, oder ein anderes Sprichwort abgewandelt: **Sag mir, wer Deine Partner sind und ich sage Dir, wer Du bist** . . . Aber theoretisch ist Ihr Gedanke natürlich richtig. (deTenTen20, 16588011681) 'That's a catch-22 situation, or to put it differently, **tell me who your partners are, and I will tell you who you are** . . . But theoretically, your thought is, of course, correct.'

The idea of flexibility of a canonical form is connected to the theory presented in other works (Mellado Blanco 2020a) that we have termed *fijación cognitiva* 'cogni-

tive fixedness', which does not always imply lexical fixedness. Indeed, lexically filled phrasemes, of any type (idioms, formulas, proverbs), constitute cognitive units that consist of a structure and meaning and are stored in our mental lexicon. However, the meaning of these units is not tied to specific words but to a certain notion with diffuse profiles, hence the lexical flexibility observed in instances of the construction in corpora. The canonical form found in dictionaries is, in most cases, just one of the possible realizations of the phraseme, and not always the most frequent one. On top of that, depending on the dictionaries, the canonical form can be divergent, so taking the canonical form from dictionaries for conducting phraseological variability studies is not always the appropriate method (see Mellado 2020a: 23; Pfeiffer 2018). Lexicographically, the ideal situation is for this form to match the most prototypical discursive realization (see above Langlotz 2006: 178). It could be argued that the canonical form of phrasemes is based on the generalization of instances of a prototypical idealized pattern (Bybee 1998: 428). In this sense, "[i]dioms do not function as words but constitute linguistically precoded cognitive micro-models" (Langlotz 2006: 289), which also explains that, when necessary, they can adapt to the context and undergo creative modifications without losing their idiomatic nature. Against this background, it is worth highlighting the discrepancies between the canonical forms of both languages. On the one hand, the canonical form in Spanish relies on, following Barðdal (2008), a highly frequent type which serves as a model of productivity by means of analogical extensions. On the other hand, the canonical form in German does not rely on one only highly frequent type, but medium-to-low-frequent types that are bound together by the same abstract cognitive core, according to the semantic connectedness of the types following Van Wettere's (2021) reasoning (see also Goldberg 2019 on the notion of *coverage*, or Mellado Blanco and Ivorra Ordines 2023). The latter form could be conceived as a kind of "cognitive canonical form".

Finally, there are 300 hapaxes legomena in Spanish and 68 nonce instances in German, where originality, expressivity, and creativity are often directly related to the length of the instance (examples 5–6). Some nonce occurrences stand out for their sharp sense of humor and disphemic nature (examples 7–8):

(5) **Dime a qué noticias de la red le das importancia, qué tan confiables son tus fuentes, qué decides compartir y cuestionar, y te diré quién eres**. (esTenTen18, 7985605409)
    '**Tell me which internet news you prioritize, how reliable your sources are, what you choose to share and question, and I will tell you who you are**.'

(6) **Zeige mir die Stelle, an der ich behauptet habe, du hieltest StarCraft für gewaltverherrlichend und ich sage dir, wer du bist** [. . .]. (de-TenTen20, 1353865372)
     '**Show me the place where I claimed you consider StarCraft to promote violence, and I will tell you who you are**.'

(7) Estos cruces verbales en defensa de los árboles marcados por una mirada tan sesgada en su espacialidad muestran que Macri, Cristina y Foster comparten el mismo paradigma espacial y afectivo de una nación [. . .]. **Dime qué tipos de árboles te preocupan y cuáles ignoras y te diré quién eres**." (esTenTen18, 11511346806).
     'These verbal exchanges in defense of the trees, marked by such a biased perspective on their spatiality, demonstrate that Macri, Cristina, and Foster share the same spatial and affective paradigm of a nation [. . .] **Tell me which types of trees concern you and which ones you ignore, and I will tell you who you are**.'

(8) Kleine Pinkologie NUR FÜR MÄNNER: **Sag' mir, wie du pinkelst, und ich sage dir, wer du bist!** Schamhafter Typ: Kann nicht pinkeln, wenn jemand zusieht, und tut so, als ob er schon fertig wäre. (deTenTen20, 11467300580)
     'Small Pinkology ONLY FOR MEN: **Tell me how you pee, and I will tell you who you are**! Shy Guy: Can't pee when someone's watching and pretends he's already finished.'

As a way of summarizing, after the corpus-based analysis of both proverbs, we validated our initial hypothesis that the high index of variability of the first part of the proverb supports the existence of a type of snowclone, or following Stumpf's (2016) terminology, of a *Modellbildung* (and not a *Modifikationsmuster*), both in Spanish and German. Indeed, the number of modified instances outnumber the number of occurrences corresponding to the canonical form(s); more precisely, the modified forms account for 55.26% in Spanish while in German 75.56%. For this reason, we confirm the existence of the productive semi-schematic construction [*Dime* X[indirect interrogative clause] *y te diré quién eres*] in Spanish and the allostructions[8] [*Sag(e) mir* X[indirect interrogative clause] *und ich sage dir, wer du bist*]

---

8 Cappelle defines allostructions as "(truth-)semantically equivalent but formally distinct manifestations of a more abstractly represented construction" (Cappelle 2009: 187). Regarding the questions whether the allostructions are semantically and pragmatically equivalent and how this may conflict with the "principle of no synonymy" (or as recently reformulated "principle of no equivalence"), see Leclercq and Morin (2023).

and [*Zeig*(*e*) *mir* NP *und ich sage dir, wer du bist*] in German. Comparing the variability index between the two languages, it is evident that German is much more prone to lexical substitution, and the German allostructions, particularly [*Zeig*(*e*) *mir* NP *und ich sage dir, wer du bist*], exhibit a much higher degree of entrenchment than the Spanish semi-schematic pattern. The results regarding the number of hapaxes (27% in Spanish and 8.39% in German) demonstrate that the construction in Spanish is more productive in terms of nonce instances, which has a great impact on the schematicity of the construction. In any case, we are dealing with a model considered a "pattern of coining" rather than a "construction proper", so its nature is different from that of "constructional idioms" (Taylor 2016: 464; see Ivorra Ordines in press or Mellado Blanco 2024). This conditions the relationship between these types of constructions and textual genre, as we will see in the following section.

4 Structure and Meaning of the Constructions

The form of the Spanish proverb aligns perfectly only with the German allostruction that includes the indirect interrogative clause, depending on the verb *decir* 'to tell' in Spanish and *sagen* 'to tell' in German (in the imperative form in both languages). It is interesting to note the conditional-consecutive value of the proverb's structure, since $V_{[imperative]}$ *y*/*und* $V_{[future]}$ should be interpreted as 'if X, then Y'. This means that conceptually the proverb in both languages is based on the logical formula 'if A, then B', which can also take other forms in proverbs, such as the German [*Wer* A, *der* B] (e.g., *Wer wagt, der gewinnt* 'Who dares wins!') or Spanish [*Quien* A, B][9] (e.g., *Quien a hierro mata, a hierro muere* 'Those that live by the sword die by the sword'; *Quien no arriesga, no gana* 'Nothing ventured, nothing gained').

The proverbs under study can be used in diverse situations, and this variability affects how they are understood and their significance. As a result, a part of their meaning is shaped by the specific contexts and situations in which they are employed. To put it in Fillmore's (1975) terms, to comprehensively understand a sentence, it is essential to have some information about the context in which it was employed. In numerous instances, grasping the meaning of a sentence entails

---

9 In this case, we are dealing with one of the formulas or "deep semantic models" of proverbs that Permjakov describes in his works (see Grzybek 2000), which are common to many languages (see Mieder 2004: 6–7). These patterns, unlike snowclones, are "Modelle der Analyse" (Fleischer 1997). See Mellado Blanco (2024) for further details.

being aware of the range of situations in which it can be suitably spoken and understanding the impact they have in that particular context. Following this line of reasoning, the proverbs in question have distinct denotative and pragmatic meanings depending on whether it appears in one of its canonical forms or in modified uses. In the former case, the denotative meaning is paraphrased as 'people are defined by the friendships they cultivate'.

From a pragmatic point of view, it is necessary to differentiate between the social function of proverbs and their illocutionary functions (Burger 2015: 110–116). The social function, in the case of non-modified proverbs, carries a moralizing tone, while the proverb serves to convey an idea commonly accepted as an absolute truth by the linguistic community. This instructive and moralizing value has lost its strength over time in our Western world, and proverbs are no longer regarded as a source of authority. According to Mieder (2012: 147), the decline in didactic intent seems to coincide with the fact that new generations do not appreciate the moralistic tone of proverbs. In the proverbs under consideration, the use of the imperative mood of the verbs *decir* 'to tell' (*dime* 'tell me' in Spanish) and *zeigen* 'to show' and *sagen* 'to tell' (zeig(*e*)/sag(*e*) 'show me/tell me' in German) aligns with the instructive and moralizing nature of these phrasemes (Manero Richard 2000: 347) and also serves a clear generalizing and appellative function.

Concerning the illocutive functions, the canonical form of the proverb is often used as a warning, advice, or reproach to disapprove of or discourage certain friendships. Following a corpus-based methodology, in this usage, the most common illocutionary acts in both languages, related to the speaker's intention, are:

– Prototypical directive illocutionary acts: WARNING, ADVICE, REPROACH.
– Prototypical expressive illocutionary act: CRITICISM.

Furthermore, in its canonical form, the proverb is used as a mechanism of cohesion of the discourse, fulfilling the discursive function of ARGUMENTATION (examples 9-13). Regarding the usage of domain of the canonical forms of the analyzed proverbs, politics predominates in both languages. In these contexts, the speaker criticizes coalitions between political parties.

(9)   Por otra parte, yo no acuso a VOX de ser proabortista. Sí lo son los partidos con los que Vidal Quadras quiere aliarse. Y advierto aquello de **dime con quién andas y te diré quién eres**. Todavía no ocurre eso . . . pero por si acaso. (esTenTen18, 6314498987)
'On the other hand, I am not accusing VOX of being pro-abortion. The parties with which Vidal Quadras wants to ally are. And I caution against that

old saying **tell me who your friends are, and I will tell you who you are**. It has not happened yet . . . but just in case.'

(10) Y el PSOE empeñado con la reacción del PP y Ciudadanos, **dime con quién andas y te diré quién eres**. (esTenTen18, 3796800194)
'And PSOE is determined by the reaction of PP and Ciudadanos, **tell me who you walk with and I will tell you who you are**.'

(11) **Zeige mir, wer deine Freunde sind und ich sage dir, wer du bist**, lautet ein geläufiges Sprichwort. Im Falle der Zeugen Jehovas dürfte die Antwort demnach recht eindeutig ausfallen. (deTenTen20, 10271137475)
'The common saying goes, **show me who your friends are, and I will tell you who you are**. In the case of Jehovah's Witnesses, the answer is likely to be quite clear.'

(12) Es gibt ein Sprichwort: „**Sage mir, mit wem Du gehst und ich sage Dir, wer Du bist.**" Ich hätte etwas mehr Taktgefühl von Herrn Bisky erwartet. (deTenTen20 7113102448)
'There is a saying: "**Tell me who you walk with, and I will tell you who you are**." I would have expected Mr. Bisky to show a bit more tact.'

(13) **Zeige mir Deine Freunde, und ich sage Dir, wer Du bist**. Meine Botschaft an Kanzler Kurz und Vizekanzler Strache lautet: Ihr habt die falschen Freunde! (deTenTen20, 24814120)
'**Show me your friends, and I will tell you who you are**. My message to Chancellor Kurz and Vice Chancellor Strache is: You have the wrong friends!'

The modifications found predominantly fall within the indirect interrogative clause in the case of the Spanish proverb and within the indirect interrogative clause or the NP, in accordance with the schemata (i) for Spanish and (ii) and (iii) for German[10].
(i)   *Dime* + indirect interrogative clause + *y te diré quién eres.*

---

[10] It is worth noting that the pattern with *zeigen* 'to show' exhibits greater variability (88%) with regards to the canonical constructs, while the pattern with *sagen* 'to tell', the use of non-canonical forms is lower (66.1%).

(14) **Dime que Converse llevas y te diré quién eres**. Si Converse logra desta-carse es gracias a la renovación de sus colecciones en sus zapatos clásicos. (esTenTen18, 4473805197)

'**Tell me which Converse you are wearing, and I will tell you who you a**re. If Converse manages to stand out, it's thanks to the renewal of their collections in their classic shoes.'

(ii) *Sag*(*e*) *mir* + indirect interrogative clause + *und ich sage dir, wer du bist.*

(15) **Sag mir, wie Du schläfst, und ich sage Dir, wer Du bist** und ob Deine Be-ziehung glücklich ist [. . .]. (deTenTen20, 112131)

'**Tell me how you sleep, and I will tell you who you are** and whether your relationship is happy.'

(iii) *Zeig*(*e*) *mir* + NP + *und ich sage dir, wer du bist.*

(16) "**Zeig mir dein Büro und ich sage dir, wer du bist**": Um zu verstehen, wie ein Unternehmen tickt, hilft oft ein Blick in dessen Arbeitsumgebung. (deTenTen20, 112131)

'**Show me your office, and I will tell you who you are**: To understand how a company operates, a glance into its work environment often provides insight.'

The patterns under study constitute semi-schematic constructions with lexically filled items and empty slots that must be renewed in the discourse, placing them halfway in the lexicon-grammar continuum. What is interesting from a Construction Grammar approach is the possibility of specifying the structural, semantic and pragmatic properties, and constraints of their slots. Such a description can provide a solid empirical analysis of the properties of such items (Mellado Blanco and Ivorra Ordines 2023). Generally speaking, the slots in both languages are filled by means of a wide variety of semantic classes, which is an indicator of the productiv-ity of the constructions (a pattern's degree of openness; see Goldberg 2006: 93).

The results in Figure 1 and Figure 2 indicate that food and drinks are predom-inant in both languages, with a very similar percentage in both cases (21%). In this regard, it seems plausible that the semi-schematic patterns have specialized in both languages to emphasize the relationship between what/how a person eats and their personality. However, social networks and likes and preferences are more frequently mentioned in Spanish, whereas house decoration and telephone/ computer games are more prevalent in German.
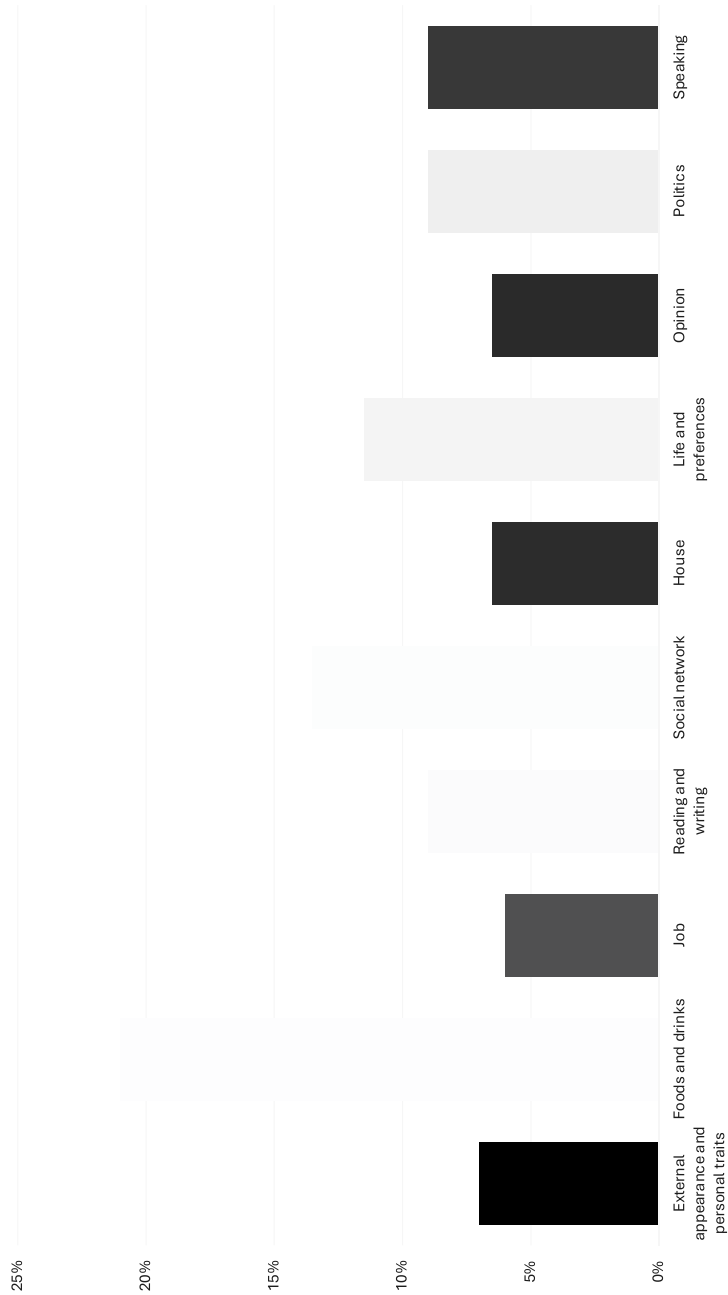
Figure 1: Semantic clusters of the creative potential of the semi-schematic construction in Spanish.
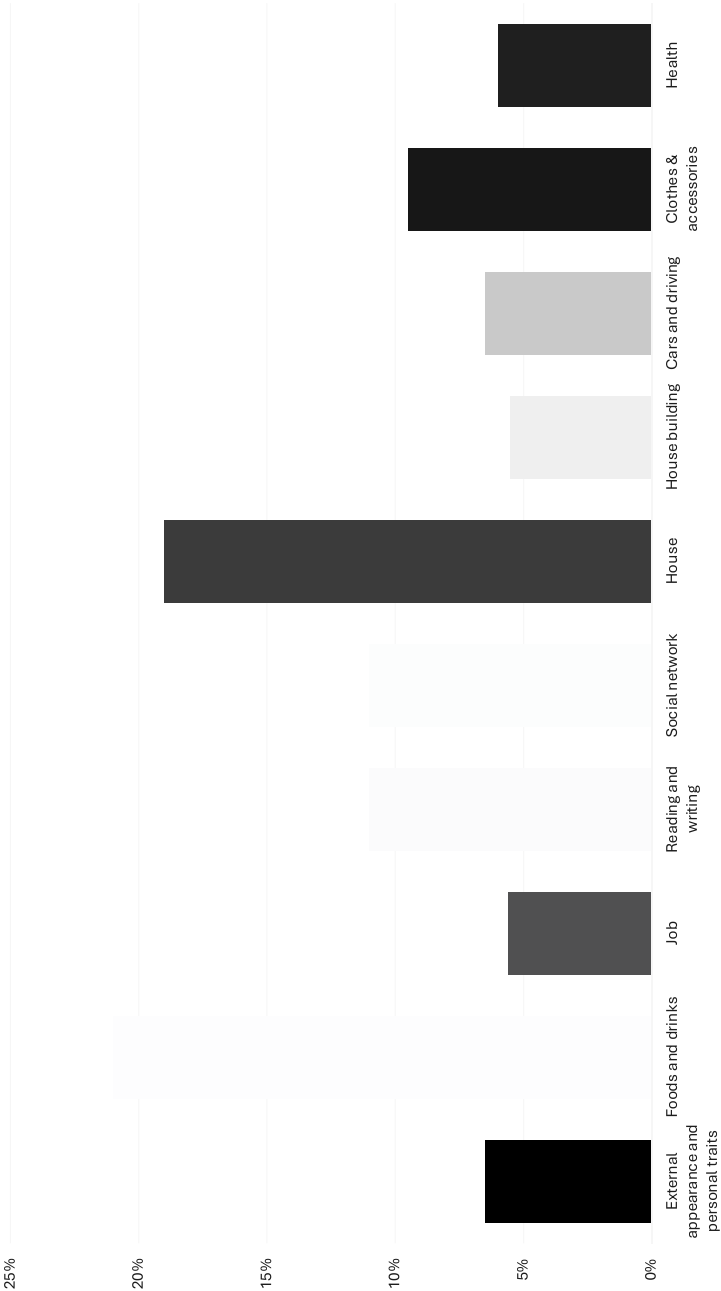
Figure 2: Semantic clusters of the creative potential of the semi-schematic constructions in German.

If the denotative meaning of the unmodified proverb is 'people are defined by the friendships they cultivate', the meaning of the modified instances depends on the semantic slot filler X in the semi-schematic constructions, in line with the different lexical-semantic clusters to which most instances of the construction can be attributed:

i.   [*Dime* X$_{[indirect\ interrogative\ clause]}$ *y te diré quién eres*]: 'people are defined by the entity referred to in X'.
ii. [*Sag*(*e*) *mir* X$_{[indirect\ interrogative\ clause]}$ *und ich sage dir, wer du bist*]: 'people are defined by the entity referred to in X'.
iii. [*Zeig*(*e*) *mir* X$_{[NP]}$ *und ich sage dir, wer du bist*]: 'people are defined by the entity referred to in X'.

In this way, following the typical process of analogy that characterizes the emergence of semi-schematic constructions (patterns emerging from proverbs) from lexically saturated phrasemes, if X refers to food, the meaning of the instance will be 'people are defined by the food they consume'. Conversely, if the slot filler alludes to clothing, the meaning will consequently be 'people are defined by the clothing they wear'.

Regarding the pragmatic potential of the proverb, in the occurrences found in the corpus, the moralizing use of the proverb, through its canonical forms, coexists with modified forms, which are the majority, i.e. slightly more than three-quarters of the total in German and half of the instances in Spanish. In line with Burger (2015: 147), our corpus analysis provides evidence of a shift in the pragmatic functions of the proverb when it appears in a modified form: "[d]as Sprichwort hat im Laufe der Jahrhunderte einen deutlichen Wandel seiner Funktionen erlebt. Insbesondere der einstmals dominierende didaktische Aspekt ist in den Hintergrund gerückt zugunsten einer spielerischen und kreativen textuellen Verwendung."[11]

Unlike what happens in the canonical moralizing use of the proverb, modified instances, in certain contexts, serve a perlocutive function of PERSUASION. Indeed, one of the breeding grounds for modified forms is popular media (book titles) and advertising, and it is precisely in these areas where the speaker/writer resorts to creative forms with the ultimate goal of being more effective in persuading the listener/reader to purchase a product, whether it be a book (examples 17-19), or any other product advertised in commercials (examples 20-22).

---

11 'Over the centuries, the proverb has undergone a clear transformation in its functions. In particular, the once-dominant didactic aspect has receded into the background in favour of a playful and creatively textual use' (the translation is ours).

(17) **Dime cómo te castigaron y te diré quién eres**. Como un grito de alerta llega este libro que aborda la manera como se está educando a los hijos, [. . .]. (esTenTen18, 15396077206)
'"**Tell me how you were punished, and I will tell you who you are**". This book arrives as a warning, addressing the way children are being raised.'

(18) **Dime lo que no comes y te diré quién eres**. Todos prometen lo mismo: pérdida de peso, aumento de la energía y, sobre todo, la salud de un roble. (esTenTen18, 814117731)
'**Tell me what you do not eat, and I will tell you who you are**. Everyone promises the same thing: weight loss, increased energy, and above all, the health of an oak tree.'

(19) Neues Buch über Luzerner Hobbygärtnerinnen und Hobbygärtner „**Zeig mir deinen Garten und ich sage dir, wer Du bist**". (deTenTen20, 184880024)
'New book about hobby gardeners in Lucerne: "**Show me your garden, and I will tell you who you are**.'

(20) **Dime qué zapato llevas y te diré quién eres**. [. . .] En B&W tenemos un amplio catálogo de zapatos de moda para mujer donde puedes encontrar el calzado perfecto para ti. (esTenTen18, 4455486530)
'**Tell me what shoe you are wearing, and I will tell you who you are**. [. . .] At B&W, we have an extensive catalog of fashionable women's shoes where you can find the perfect footwear for you.'

(21) „**Sage mir was Du ißt und ich sage Dir, wer Du bist**". Auf zum Spielplatz! Eis-Lieferservice in Wien Der Frühling ist endlich da [. . .]. (deTenTen20, 8744670250)
'"**Tell me what you eat, and I will tell you who you are**." Off to the playground! Ice cream delivery service in Vienna. Spring is finally here.'

(22) In kaum einem Raum gilt das Motto „**zeige mir, wie du wohnst und ich sage dir, wer du bist**", ausgeprägter als im Wohn- und Essbereich. Ein gut sortiertes Einrichtungshaus und Möbelladen [. . .]. (deTenTen20, 15912060148)
'In hardly any room does the motto "**show me how you live, and I will tell you who you are**" apply more than in the living and dining area. A well-stocked furniture store and home decor shop.'

The fact that these semi-schematic patterns are linked to specific textual genres indicates that they evoke what Schmid (2020) calls pragmatic associations that are

mediated by contextual factors. Pragmatic associations indeed contribute to the routinization and schematization of these semi-schematic constructions in three different ways (Schmid 2014: 275):

(i)   Their effectiveness in communication as persuasive mechanisms can lead to frequent usage, promoting gradual chunking process both within individual minds and across a community of speakers.

(ii) Repeated pragmatic associations can reinforce the symbolic connections of previously chunked elements, thereby increasing their stability within the speech community and granting the status of "social gestalts" (Feilke 1996).

(iii) Pragmatic associations can strengthen the productive use of schemas related to recurring situations and communicative intentions. Being linked to a specific discursive function (such as popular media and advertising) enhances their frequent and productive usage, contributing to contextual entrenchment.

What we observe from these semi-schematic patterns is that the emergence of novel associations between form and meaning (in the process of constructionali-zation[12]) has not only been accompanied by shifts in schematicity, productivity and compositionality (Traugott and Trousdale 2013; see Ivorra Ordines in press or Mellado Blanco 2024), but also by shifts in the specialization of their instances in terms of textual genres. More precisely, the textual genres in which these partly filled patterns of the proverb most frequently appear, with a clear predominance of the first category, are as follows:

(a)   Self-help manuals, personality tests, popular texts on health and psychology. The message conveyed through the pattern is that there are traits of the human being that inevitably relate to certain personality characteristics, whether it is how a person wears their hair, their color preferences, what they eat, what they read, how they write, etc. Within these genres, the sub-genre of book titles or popular articles stand out. By choosing a modified proverb, the title fulfills its characteristic appellative function as a lure to generate curiosity in the reader and encourage them to read.

(b)   Advertising texts for homes/gardens and clothing/shoes, including accessories and complements.

---

12 While the fundamental trigger for the emergence of these types of constructions is, from a cognitive perspective, analogy (see Ivorra Ordines in press), in these schemata generated by lexical substitution, it is not entirely clear whether, as Langlotz (2006: 205) suggests, we are also dealing with a phenomenon of contamination, wordplay or even lapsus. In our opinion, this issue should be resolved on a case-by-case basis, depending on the creative intent of the speaker (see Mellado Blanco 2020a).

In this context, the notion of "discourse patterns", introduced by Östman (1999) and further elaborated in Östman (2005), is prevalent to account for the interdependence between clusters of grammatical features and certain types of texts or genres. Such is the case of headlines, as examined by Östman (2005), whose morphosyntax in structures like *Dog saves baby* or *Mother drowned baby* are fully conventional and expected, with the absence of articles (both definite and indefinite) forming an integral part of the convention. Östman argues that such conventions can be theoretically formalized in terms of discourse patterns, "by specifying a construct as [dp headline], we indicate that articles do not unify in the manner they would by default" (Östman 2005: 140), which would allow to preserve certain unification-based formalism principles from linguistic theory (as proposed by Kay and Fillmore 1999; see Hoffmann 2022 or Sommerer 2023 for an example of a constructional template). The goal is not to replace the insights generated by discourse analysis but to incorporate those that align with the conventions well-established in grammatical theory. This means that information regarding the pragmatic, discoursal, textual, and register characteristics linked to a specific form can be incorporated into the meaning aspect of the corresponding construction, in addition to purely semantic details (Fried and Östman 2004). In Construction Grammar, the term "meaning" encompasses all the conventionalized elements of a construction's function. This may comprise not only attributes of the situation described by the utterance but also aspects of the discourse context in which the utterance is situated (Croft and Cruse 2004: 258; see Antonopoulou and Nikiforidou 2011).

When the usage of domain is not that of headlines or book titles, nor is it advertising, modified forms serve to cohere the discourse and have the illocutive function of ARGUMENTATION (examples 23–25). This is the function that the proverb in its canonical form also fulfills, as is mentioned above (see examples above).

(23) Son de un mal gusto tremendo, riéndose de las enfermedades de sus pacientes. **Dime de qué te ríes, y te diré quién eres**. Hay cosas con las que simplemente no se juega. (esTenTen18, 6545066559)
'They have a terrible sense of taste, laughing at their parents' illnesses. **Tell me what you laugh at, and I will tell you who you are**. There are things that you simply should not joke about.'

(24) –La cuestión es como cobras a tus clientes. **Dime cómo cobras y te diré quién eres**. Por eso el replanteamiento del modo de cobrar. (esTenTen18, 2453197000)
'–The question if how you charge your clients. **Tell me how you charge, and I will tell you who you are**. That is why the reconsideration of way to charge.'

(25) **Sage mir, worüber du lachst, und ich sage dir, wer du bist**. Kinder unter Zehn lachen am liebsten über Witze, in denen gepupst wird. (deTenTen20, 12352035041)
   '**Tell me what you laugh about, and I will tell you who you are**. Children under ten prefer to laugh at jokes in which someone farts.'

Considering the semantic clusters mentioned in Figures 1 and 2, an interesting phenomenon is the emergence of micro-constructions that are originally phraseological quotes, as they have a known author (similarly as how snowclones work following Hartmann and Ungerer 2023; see Ivorra Ordines in press or Mellado Blanco 2024). In this respect, diffusion plays a role in how conventions (emergent micro-constructions and potential analogical extensions) spread within a speech community or its subgroups, across various genres and types of text, and across different activities and situations (Schmid 2020). Driven by the diffusion process, successful emergent micro-constructions gradually extend their influence beyond their original regional and social boundaries within the speech community, expanding into new areas of discourse and application. A common pathway of diffusion "goes from use in a specialized community related to topics of shared interest to its members to use by more and more members of the larger community in more and more contexts" (Schmid 2020: 94). Three different originally phraseological quotes were found in both corpora:

– The expression "Dis-moi ce que tu manges, je te dirai ce que tu es" 'tell me what you eat, and I will tell you who you are' in French, appearing in *Physiologie du goût*, one of the most famous literary works by the French gastronome Jean-Antelme Brillat-Savarin (1825–1926), a work considered one of the foundational texts of gastronomy. In such a case both the Spanish calque "Dime qué comes y te diré quién eres" 'Tell me what you eat and I will tell you who you are' (43 occurrences in esTenTen18) and the German calque "Sag(e)/Zeig(e) mir, was du isst, und ich sage dir, was du bist" 'Tell me what you eat and I will tell you who you are' emerge (14 occurrences in deTenTen20). Cluster: food and drinks.
– The expression "Dis-moi ce que tu lis, je te dirai ce que tu es" 'Tell me what you read, and I will tell you who you are' by the French magistrate, lawyer and historian Pierre de la Gorçe (1846–1934). In this example the Spanish calque "Dime qué lees y te diré quién eres" 'Tell me what you read, and I will tell you who you are' (14 occurrences in esTenTen18) and the German calque "Sag(e)/Zeig(e) mir, was du liest, und ich sage dir, wer/was du bist" ''Tell me what you read, and I will tell you who you are' (5 occurrences in deTenTen20). Cluster: reading and writing.

– The expression "Sag(e)/Zeig(e) mir, wie du baust, und ich sage dir, wer du bist" 'Show me how you build, and I will tell you who you are' by the German writer, dramatist, journalist, and translator Christian Morgenstern (1871–1914) (Duden 12). This emergent micro-construction appears 8 times in deTenTen20. Cluster: house building.

This phenomenon demonstrates the entrenchment of the semi-schematic constructions, capable of generating new micro-constructions based on the token frequency of certain types (see Ivorra Ordines and Mellado Blanco 2021; Mellado Blanco and Ivorra Ordines 2023). It is worth noting that these micro-constructions are supported by a specific author, who most likely was not the person who coined the form but only reproduced it, and as a public figure, contributed to its dissemination. From this perspective, the process of diffusion facilitates change in spatial, social, and stylistic dimensions of conventionality, that is, "spread from urban centres to rural areas or vice versa, from higher and middle social classes to lower ones or vice versa, from colloquial to formal contexts or vice versa, or from technical genres to everyday ones and vice versa" (Schmid 2020: 201).

5 Conclusions

The corpus-based study carried out revealed that the Spanish proverb *Dime con quién andas y te diré quién eres* and its prototypical German equivalents *Sag(e) mir, mit wem du umgehst, und ich sage dir, wer du bist* and *Zeig(e) mir deine Freunde, und ich sage dir, wer du bist* tend to exhibit variability in the interrogative clause and in the NP (in German), albeit to varying degrees. In line with Stumpf's proposal (2016), the fact that more than half of the occurrences found in the respective searches feature a modification of the canonical form in the first part of the proverb confirms our initial hypothesis that the proverb has developed a pattern closely related to the notion of snowclone following Hartmann & Ungerer (2023), that is, a semi-schematic construction through lexical substitution (Ivorra Ordines in press; Mellado Blanco 2024). This means that we cannot speak of occasional modifications but rather systematic variability. The semi-schematic constructions in German are shown to be highly productive by means of non-canonical forms with 77.56% of the instances alluding to modified forms compared to the 55.26% in Spanish. In German, to be more precise, the pattern with *zeigen* 'to show' exhibits a considerable variability (88%) compared to the canonical constructs. On top of that, both constructions can be seen as productive in terms of semantic dispersion considering the wide variety of semantic clusters,

while the number of hapaxes is relatively high in the construction in Spanish in comparison with the German ones (27% and 8.39%, respectively).

Concerning the slot fillers of both constructions, there is a parallel development in the semantic fields of modifications in Spanish and German. However, while the lexical-semantic of the slot fillers show some overlap, there are also discrepancies, with varying degrees of prototypicality among those that do not align. Notably, the cluster of 'food/drink' stand out among the clusters in both languages.

German also exhibits a higher index of variability in its canonical form, not allowing us to determine just a single form but rather two (with an interrogative clause and with a direct complement), and within each form, various lexical variants are detected in the corpus. This leads us to reinforce the idea of cognitive fixedness rather than lexical fixedness in the German proverb, with a flexible lexicalized core referring to people's friendship. Moreover, the presence of two distinct syntactic patterns accounts for the coexistence of two allostructions in German, unlike in Spanish. We also confirmed the existence of the productive semi-schematic construction [*Dime* X[indirect interrogative clause] *y te diré quién eres*] in Spanish and the allostructions [*Sag(e) mir* X[indirect interrogative clause] *und ich sage dir, wer du bist*] and [*Zeig(e) mir* NP *und ich sage dir, wer du bist*] in German.

Regarding the meaning of the constructions, we noted that there are differences in modified uses compared to the canonical form(s). While the denotative meaning is similar ('people are defined by X' with X corresponding to the slot filler of the construction [*Dime* X *y te diré quién eres*]), the pragmatic meaning varies. In this way, the social function by means of which the proverb is perceived as a moral authority, which has been present throughout the centuries, is no longer evident in the modified uses of the proverb, yet it is exploded. A new perlocutionary function has indeed emerged, that of PERSUASION, for the purpose of argumentation, especially in textual genres such as advertising and popular science books, particularly in the subgenre of book titles, which are primarily aimed at promoting the purchase of the work. As for illocutionary functions, in other literary genres (such as journalist language: articles on politics, in particular), similar functions are observed in the canonical forms. These functions include ADVICE, REPROACH, and CRITIQUE.

Finally, this study has highlighted that social and contextual usage constraints are essential characteristics of the constructional templates, which allows to extend the domain of conventional form-meaning associations –focusing on entrenched formal and discourse-pragmatic properties. Additionally, we have shed light on the connection between lexical features and specific speech acts, potentially allowing a comprehensive examination of idiomaticity within discourse patterns, which encompass various levels of productivity and fixedness.

References

Antonopoulou, Eleni & Kiki Nikiforidou. 2011. Construction grammar and conventional discourse: A construction-based approach to discoursal incongruity. *Journal of Pragmatics* 43. 2594–2609.

Baayen, Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An international handbook*. 2nd volume, 899–919. Berlin: De Gruyter.

Burger, Harald. 2015. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: Schmidt.

Bybee, Joan. 1998. The emergent lexicon. *Chicago Linguistic Society* 34. 421–435.

Bybee, Joan. 2013. Usage-based Theory and Exemplar Representations of Constructions. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 49–69. Oxford: Oxford University Press.

Cappelle, Bert. 2009. Can we factor out free choice? In Andreas Dufter, Jürg Fleischer & Guido Seiler (eds.), *Describing and Modeling Variation in Grammar*, 183–202. Berlin: De Gruyter.

Croft, William & Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

DFDEA: Seco, Andrés, Olimpia Andrés & Gabino Ramos. 2004. *Diccionario fraseológico documentado del español actual: locuciones y modismos*. Madrid: Aguilar.

Esteban-Fonollosa, Maricel. 2023. *El inútil de su hijo*. La construcción [DET$_{[det/demost]}$ ADJ *de* PRON$_{[pos]}$ N] en español. *Romanica Olomucensia* 35(2). 299–311.

Fauconnier, Gilles. 1975. Pragmatic Scales and Logical Structure. *Linguistic Inquiry* 6(3). 353–375.

Feilke, Helmuth. 1996. *Sprache als soziale Gestalt: Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt: Suhrkamp.

Fillmore, Charles J. 1975. *Santa Cruz lectures on Deixis 1971*. Bloomington: Indiana University Linguistics Club.

Fillmore, Charles J. 1982. Frame Semantics. *Linguistics in the Morning Calm*, 111–137. Seoul: Hansin.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2). 222–254.

Fillmore, Charles J. 1989. Grammatical Construction Theory and the Familiar Dichotomies. In Rainer Dietrich & Carl Graumann (eds.), *Language Processing in Social Context. An Interdisciplinary Context*, 17–38. Amsterdam: Elsevier Science Publishers.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64(3). 501–538. http://pdfs.semanti cscholar.org/23e2/7224928a008c5be84a7ea1452e0c627fe380.pdf (accessed 17 July 2023).

Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. Tübingen: Niemeyer.

Fried, Mirjam. 2009. Construction grammar as a tool for diachronic analysis. *Constructions and Frames* 1(2). 262–291.

Fried, Mirjam & Morgan Aarestrup. 2021. Relationships between Construction Grammar(s) and Genre. Evidence from an Analysis of Instagram Posts. *Journal of Pragmatics* 183. 87–104.

Fried, Mirjam & Jan-Öla Östman. 2004. Construction Grammar: a thumbnail sketch. In Mirjam Fried & Jan-Öla Östman (eds.), *Construction Grammar in a Cross-Language Perspective*, 11–86. Amsterdam & Philadelphia: John Benjamins.

Geeraerts, Dirk, 2005. Lectal variation and empirical data in Cognitive Linguistics. In Francisco Ruiz de Mendoza Ibáñez & Sandra Peña Cervel (eds.), *Cognitive Linguistics. Internal Dynamics and Interdisciplinary Interactions*, 163–189. Berlin: Mouton de Gruyter.

Goldberg, Adele E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg Adele E. 2019. *Explain me this: Creativity, Competition and the Partial Productivity of Constructions*. Princeton: Princeton University Press.

Gries, Stefan. 2008. Phraseology and Linguistic Theory: A Brief Survey. In Sylviane Granger & Fanny Meunier (eds.), *Phraseology: An Interdisciplinary perspective*, 3–25. Amsterdam & Philadelphia: John Benjamins.

Grzybek, Peter. 2000. *Die Grammatik der sprichwörtlichen Weisheit von G. L. Permjakov. Mit einer Analyse allgemein bekannter deutscher Sprichwörter*. Essen: Scheinder Verlag Hohengehren GmbH.

Hartmann, Stefan & Tobias Ungerer. 2023. Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns. *Journal of Linguistics*. 1–36. https://www.cambridge.org/core/journals/journal-of-linguistics/article/attack-of-the-snowclones-a-corpusbased-analysis-of-extravagant-formulaic-patterns/B6A3C6D3E5B4C3BBEBB8332901565A63 (accessed 28 July 2023).

Hoffmann, Thomas. 2015. Cognitive sociolinguistic aspects of football chants: The role of social and physical context in Usage-based Construction Grammar. *Zeitschrift für Anglistik und Amerikanistik* 63(3). 273–294.

Hoffmann, Thomas. 2022. *Construction Grammar. The structure of English*. Cambridge: Cambridge University Press.

Hoffmann, Thomas & Alexander Bergs. 2018. A Construction Grammar Approach to Genre. *Cognitextes: Revue de l'Association française de linguistique cognitive* 18. https://journals.openedition.org/cognitextes/1032 (accessed 3 August 2023).

Hoffmann, Thomas & Alexander Bergs. 2023. Constructions all the way! Text types as constructions. In Anja Hennemann & Felix Tacke (eds.), *Diskurstraditionen – Konstruktionen – Genres*, 255–277. Bonn: Bonn University Press.

Ivorra Ordines, Pedro. 2021. *Les construccions comparatives intensificadores de la lletjor en català, espanyol, anglés i francés des de les gramàtiques de construccions. Un estudi basat en corpus*. Barcelona: Universitat Pompeu Fabra dissertation.

Ivorra Ordines, Pedro. 2023. *Por mí como si te operas*. Constructional idioms of rejection from a constructionist approach. *Yearbook of Phraseology* 14. 89–120.

Ivorra Ordines, Pedro. In press. *Dime con quién te juntas y te diré quién eres*. Substantive idioms from Construction Grammar. In Torsten Leuschner, Anaïs Vajnovszki, Gauthier Delaby & Jóhanna Barðdal (eds.), *How to do things with corpora – Methodological issues and case studies*. Berlin: Springer Nature.

Ivorra Ordines, Pedro & Carmen Mellado Blanco. 2021. *Más tontos que el novio de la Chelo*. La intensificación de la estulticia en foros y chats por medio de comparaciones creativas: una aproximación desde la Gramática de Construcciones. [Special Issue: *La intensidad en las lenguas románicas como estrategia comunicativa*]. *Estudios Románicos* 30. 39–58. https://revistas.um.es/estudiosromanicos/article/view/471241/308431 (accessed 3 August 2023).

Kay, Paul & Charles J. Fillmore. 1999. Grammatical Constructions and Linguistic Generalizations: The *What's X Doing Y?* Construction. *Language* 75(1). 1–33. https://pdfs.semanticscholar.org/0b8f/b343f979bf1bc48eed705cd1ada4075140de.pdf?_ga=2.208618582.572245901.1612553016-61201396.1604511679 (accessed 4 August 2023).

Langacker, Ronald W. 2008. *Cognitive Grammar. A basic introduction*. Oxford: Oxford University Press.

Langlotz, Andreas 2006. *Idiomatic Creativity. A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam & Philadelphia: John Benjamins.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Lambrecht, Knud. 2004. On the interaction of information structure and formal structure in constructions: The case of French right-detached *comme*-N. In Fried Mirjam & Jan-Ola Östman (eds.), *Construction Grammar in a Cross-Language Perspective*, 157–199. Amsterdam & Philadelphia: John Benjamins.

Leclercq, Benoît & Cameron Morin. 2023. No equivalence: A new principle of no synonymy. *Constructions*. Special Issue: *35 years of Constructions* 15(1). 1–16. https://constructions.journals. hhu.de/article/view/535/584 (accessed 15 September 2023).

Manero Richard, Elvira. 2000. La lingüística implícita del refranero español de contenido metalingüístico. *RILCE: Revista de Filología Hispánica* 16(2). 343–361.

Manero Richard, Elvira. 2007. El refranero español de contenido metalingüístico: refranes sobre tipos de texto y actos de habla. *RILCE: Revista de Filología Hispánica* 23(2). 417–434.

Matsumoto, Yoshiko. 2015. Partnership between grammatical construction and interactional frame. The stand-alone noun-modifying construction in invocatory discourse. *Constructions and Frames* 7(2). 289–314.

Mellado Blanco, Carmen. 2020a. La desautomatización desde el prisma de la Gramática de Construcciones: un nuevo paradigma de la variabilidad fraseológica. *Nasledje* 45. 17–34.

Mellado Blanco, Carmen. 2020b. (*No*) *me importa un comino* y sus variantes diatópicas. Estudio de corpus desde la Gramática de Construcciones. *ELUA*. Anexo VII. 89–111.

Mellado Blanco, Carmen (ed.). 2022. *Productive Patterns in Phraseology and Construction Grammar. A Multilingual Approach*. Berlin: de Gruyter.

Mellado Blanco, Carmen. 2023. From idioms to semi-schematic idioms and vice versa. The case of [*a un paso de* X]. In Inga Hennecke & Evelyn Wiesinger (eds.), *Constructions in Spanish*, 103–128. Amsterdam & Philadephia: John Benjamins.

Mellado Blanco, Carmen. 2024. The ways of biblical quotations are mysterious: Change of register and snowclones in Spanish and German from a constructivist point of view. In Saša Babič, Anna T. Litovkina, Fionnuala Carson Williams & Christian Grandl (eds.), *"Standing on the shoulders of giants". A Festschrift in honour of Wolfgang Mieder on the occasion of his 80tj birthday. Proverbium Online Supplement* 3: 2024. 495–512. Osijek: Faculty of Humanities and Social Sciences at the University of Osijek. https://naklada.ffos.hr/knjige/index.php/ff/catalog/book/18 (accessed 5 September 2023).

Mellado Blanco, Carmen & Pedro Ivorra Ordines. 2023. *Casi palmo de la risa*. A corpus-based study of a Spanish constructional idiom with the causal preposition *de*. [Special Issue: *35 years of Constructions*]. *Constructions* 15(1). 1–34. https://constructions.journals.hhu.de/article/view/537/ 587 (accessed 5 September 2023).

Mellado Blanco, Carmen, Pedro Ivorra Ordines & Maricel Esteban Fonollosa (eds.). 2024. [Special Issue: *New approaches to Construction Grammar(s)*]. *Quaderns de Filologia. Estudis Lingüístics* 29. Valencia: Universitat de València.

Mellado Blanco, Carmen, Fabio Mollica & Elmar Schafroth (eds.). 2022. *Konstruktionen zwischen Lexikon und Grammatik: Phrasem-Konstruktionen mo-nolingual, bilingual und multilingual*. Berlin: de Gruyter.

Michaelis, Laura & Knud Lambrecht. 1996. Toward a construction-based theory of language function: the case of nominal extraposition. *Language* 72. 215–247.

Mieder, Wolfgang. 2004. *Proverbs: A Handbook*. London: Greenwood Press.

Mieder, Wolfgang. 2012. *Proverbs Are Never Out of Season: Popular Wisdom in the Modern Age*. New York: Oxford University Press.

Mollica, Fabio. 2018. Über das Sprichwort *Keine Antwort ist auch eine Antwort* und die Phrasem-Konstruktion [*Kein*(*e*) N1 *ist auch ein*(*e*) N1]: formale und semantisch-pragmatische Eigenschaften. In Martina Nicklaus, Nora Wirtz, Wiebke Langer, Marcella Costa & Karin Ewert-Kling (eds.), *Lexeme, Phraseme, Konstruktionen: Aktuelle Beiträge zu Lexikologie und Phraseologie. Festschrift für Elmar Schafroth*, 205–222. Frankfurt: Peter Lang.

Nikiforidou, Kiki. 2009. Constructional Analysis. In Frank Brisard, Jan-Ola Östman & Jef Verschueren (eds.), *Grammar, Meaning and Pragmatics*, 16–32. Amsterdam & Philadelphia: John Benjamins.

Nikiforidou, Kiki. 2010. Viewpoint and construction grammar: The case of past + now. *Language and Literature* 19(3). 265–284.

Nikiforidou, Kiki. 2015. Grammatical constructions and cross-text generalizations. Empathetic narration as genre. *Constructions and Frames* 7(2). 181–217.

Nikiforidou, Kiki. 2016. Genre knowledge in a constructional framework: Lexis, grammar and perspective in folk tales. In Ninke Stukker, Wilbert Spooren and Gerard Steen (eds.), *Genre in Language, Discourse and Cognition*, 331–359. Berlin: Mouton de Gruyter.

Nikiforidou, Kiki. 2018. Genre and constructional analysis. *Cognitive Perspectives on Genre* 25(3). 543–576.

Nikiforidou, Kiki. 2021. Grammatical variability and the grammar of genre: Constructions, conventionality, and motivation in 'stage directions'. *Journal of Pragmatics* 173. 189–199.

Nikiforidou, Kiki & Kerstin Fischer. 2015. On the interaction of constructions with register and genre. *Constructions and Frames* 7(2). 137–147.

Nikiforidou, Kiki & Mirjam Fried. 2015. Genre and constructional analysis. *Pragmatics and Cognition* 25(3). 543–575.

Östman, Jan-Ola. 1995. Pragmatic particles twenty years after. In Brita Wårvik, Sanna-Kaisa Tanskanen & Risto Hiltunen (eds.), *Organization in discourse*, 95–108. Turku: University of Turku.

Östman, Jan-Ola. 2005. Construction Discourse. A Prolegomenon. In Jan-Ola Östman & Mirjam Fried (eds.), *Construction Grammars: Cognitive grounding and theoretical extensions*, 121–144. Amsterdam & Philadelphia: John Benjamins.

Östman, Jan-Ola. 2015. From Construction Grammar to Construction Discourse . . . and back. In Jörg Bücker, Susanne Günthner & Wolfgang Imo (eds.), *Konstruktionsgrammatik V: Konstruktionen im Spannungsfeld von sequenziellen Mustern, kommunikativen Gattungen und Textsorten*, 15–44. Tubingen: Stauffenburg verlag.

Östman, Jan-Öla & Mirjam Fried. 2005. The Cognitive grounding of Constructional Grammar. In Jan-Ola Östman & Mirjam Fried (eds.), *Construction Grammars: Cognitive grounding and theoretical extensions*, 1–13. Amsterdam & Philadelphia: John Benjamins.

Östman, Jan-Ola & Graeme Trousdale. 2013. Dialects, discourse, and Construction Grammar. In Thomas Hoffmann & Graeme Trousdale (Eds.), *The Oxford Handbook of Construction Grammar*, 476–490. Oxford: Oxford University Press.

Petrova, Oksana. 2011. *Of Pearls and Pigs: A Conceptual-Semantic Tiernet Approach to Formal Representation of Structure and Variation of Phraseological Units*. Åbo: Åbo Akademi University Press.

Pfeiffer, Christian. 2017. Okkasionalität. Zur Operationalisierung eines zentralen definitorischen Merkmals phraseologischer Modifikation. *Yearbook of Phraseology* 8. 9–30.

Pfeiffer, Christian. 2018. Zur Identifikation modifizierter Phraseme in Texten: ein Vorschlag für die analytische Praxis. In Sören Stumpf & Natalia Filatkina (eds.), *Formelhafte Sprache in Text und Diskurs*, 49–83. Berlin & New York: De Gruyter.

Pullum, Geoffrey K. 2004. Snowclones: Lexicographical dating to the second. http://itre.cis.upenn.edu/~myl/languagelog/archives/000350.html (accessed 3 September 2023).

Rasulic, Katarina. 2010. *Long Time, No Buzz*: Fixed Expressions as Constructional Frames. *CogniTextes: Revue de l'Association française de linguistique cognitive* 5. https://journals.openedition.org/cogni textes/356 (accessed 4 September 2023).

Schmid, Hans-Jörg. 2014. Lexico grammatical patterns, pragmatic associations and discourse frequency. In Thomas Herbst, Hans-Jörg Schmid & Susen Faulhaber (eds.), *Constructions, collocations, patterns*, 239–293. Berlin: De Gruyter.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.

Sommerer, Lotte. 2023. *It's what she said, then it's what she said*: a usage-based, constructional analysis of pleonastic conditionals in English. *Corpus Pragmatics* 7. 345–376.

Stumpf, Sören. 2016. Modifikation oder Modellbildung? Das ist hier die Frage – Abgrenzungsschwierigkeiten zwischen modifizierten und modellartigen Phrasemen am Beispiel formelhafter (Ir-)Regularitäten. *Linguistische Berichte* 247. 317–342.

Stumpf, Sören. 2018. Formulaic (Ir-)Regularities in German. Corpus Linguistics and Construction Grammar Approaches. In Natalia Filatkina & Sören Stumpf (eds.), *Konventionalisierung und Variation. Phraseologische und konstruktionsgrammatische Perspektiven*, 149–178. Frankfurt: Peter Lang.

Stumpf, Sören. 2022. Formelhaft (ir-)reguläre Phrasem- Konstruktionen im Deutschen: Theoretische Verortung und korpuslinguistische Analyse. In Carmen Mellado Blanco, Fabio Mollica & Elmar Schafroth (eds.), *Konstruktionen zwischen Lexikon und Grammatik: Phrasem-Konstruktionen monolingual, bilingual und multilingual*, 165–190. Berlin: De Gruyter.

Taylor, John R. 2016. Cognitive linguistics. In Keith Allan (ed.), *Routledge Handbook of Linguistics*, 455–469. London: Routledge.

Traugott, Elizabeth, 2010. Dialogic motivations for syntactic change. In Robert A. Cloutier, Anne Marie Hamilton-Brehm & William A. Kretzschmar (eds.), *Variation and Change in English Grammar and Lexicon*, 11–36. Berlin: de Gruyter.

Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Van Wettere, Nick. 2021. Productivity of French and Dutch (semi-)copular constructions and the adverse im-pact of high token frequency. *International Journal of Corpus Linguistics* 26(3). 396–428.

Wotjak, Gerd. 1998. ¿Cómo tratar las unidades fraseológicas (UF) en el diccionario? In Gerd Wotjak (coord.), *Estudios de fraseología y fraseografía del español actual*, 307–322. Frankfurt: Iberoamericana Vervuert.

Ziem, Alexander. 2014. *Frames of Understanding in Text and Discourse: Theoretical foundations and descriptive applications*. Amsterdam & Philadelphia: John Benjamins.

Ziem, Alexander & Alexander Lasch. 2013. *Konstruktionsgrammatik: Konzepte und Grundlagen gebrauchsbasierter Ansätze*. Berlin: de Gruyter.

Muriel Norde, Francesca Masini, Kristel Van Goethem &
Daniel Ebner

Wannabe Approximatives

Creativity, Routinization or Both?

Abstract: *Wannabe* is a versatile word that has found its way into several languages. A univerbation of *wanna* (< *want to*) and *be*, English *wannabe* is used both as a noun meaning 'a fake person' and as part of a collocation, e.g. *wannabe-gangster* or *Elvis-wannabe*, where it is often depreciative in meaning. This paper presents a corpus-based case study of *wannabe* in English and five other languages *wannabe* has been borrowed into (Danish, Dutch, French, Italian and Finnish). In all languages, *wannabe* collocations are productive and show substantial variation, both between languages and within a single one. This constructional variability of *wannabe* raises interesting questions about creativity and routinization in word formation. We propose an adapted version of the concepts of F- and E-creativity, which distinguishes between F1-, F2- and E-creativity. From the point of view of Construction Morphology, *wannabe* collocations can be formalized as schemas with *wannabe* as a fixed slot and different constraints on the collocate in the open slot. Our case studies show that these schemas are flexible and open to incremental formal or semantic changes (F2-creativity), which may result in either routinization, i.e. adaptation of existing schemas (F1-creativity), or the entrenchment of new ones (E-creativity).

1 Introduction

Wannabe, from the phrase *want to be*, is a versatile word that can be used either as a free form or in combination with other words, both in English and in languages it has been borrowed into. As a free form, it mostly occurs as a noun referring to a disingenuous person, as in (1), or, more rarely, as a predicative adjective, as in (2). Collocates include common nouns, as in (3), proper names as in (4), or adjectives as in (5).[1]

---

1 The examples are from enTenTen20. See Section 4 for a description of the corpora used for this study. All examples are quoted as they occur in the corpus, that is, we did not remove typos or other errors. Also, we do not provide glosses or word-by-word translations, as these are not necessary for this phenomenon.

(1) *My pal Bruno the Singing DJ is a welcome exception. I hope it's him and not some **wannabe**.*

(2) *I don't think she's "**wannabe**" at all she's still unique . . .*

(3) *This drama involves his friend Sumire, an aspiring writer and **wannabe beatnik**.*

(4) *"The world is full of **Madonna wannabes**, I might have even dated a couple," said Britney Spears' ex. "But there is truly only one Madonna."*

(5) *A **wannabe-posh**, soapy, sheer, very smooth (and very synthetic) sort of citrus-floral-musk blend*

While the original meaning of 'to aspire to' persists, the implication is often one of ambitions doomed to fail – try as they might, the *Madonna wannabes* in example (4) can never compare to the real one. This mildly depreciative sense opened the door to more general, often approximative, meanings, as in the adjectival phrase in (5), but also in combination with nouns that do not refer to a human being, such as *BMW wannabe* or *wannabe bestseller*. Examples (1–5) furthermore show that collocational *wannabe* is quite variable: it occurs both with a hyphen or without, and it can either precede its collocate or follow it. More variation still is found in other languages, where *wannabe* can form part of proper compound words, for instance Dutch *wannabepiloot* 'wannabe pilot' or Finnish *wannabekirjailijat* 'wannabe authors'.

In this paper, we present data from contemporary web corpora to compare the various uses of *wannabe* in English and five recipient languages (Danish, Dutch, French, Italian and Finnish). For each language, we analyse the position of *wannabe* in relation to its collocate, its (orthographical) bonding properties, the part of speech of the collocate, frequency distributions, and semantic categories of the collocate. Our analysis shows substantial language-internal variation but also differences between the languages in our set, both in terms of the constructional properties just mentioned and the degree to which the recipient languages are similar to English.

For these reasons, our data lend themselves well to questions concerning the concept of creativity, in particular how creativity can be delineated from routine in word formation. Creativity has often been understood as the "intentional manipulation of linguistic structure" (Bergs 2018: 290), whereby the hearer can discriminate between creative use, i.e. breaking *conventions*, and errors, i.e. breaking *rules* (Uhrig 2020: 1). To distinguish between creativity and routine (or

regular, productive usage), we refer to Sampson's (2016) notions of F-creativity (fixed creativity), defined as "activities which characteristically produce examples drawn from a fixed and known (even if infinitely large) range" and E-creativity (enlarging or extending creativity), i.e. "activities which characteristically produce examples that enlarge our understanding of the range of possible products of the activity" (Sampson 2016: 17). In this dichotomy, F-creativity corresponds to routine, whereas E-creativity corresponds to creativity in Bergs' sense.

However, the difference is difficult to operationalize because there is a 'grey area' where speakers may produce new instances that still follow established patterns of word formation, but which break conventions nevertheless, for instance when they use *wannabe* to modify an adjective or a non-human noun. But when this new usage catches on and type frequency increases, the rule becomes less constrained, such that collocations are no longer restricted to human nouns. Once that has happened, it is no longer justified to speak of breaking conventions. Norde & Trousdale (2024) therefore propose a more fine-grained model which acknowledges two stages of F-creativity. In their model, F1-creativity is used for fully productive patterns (such as de-adjectival nouns in *-ness*), whereas F2-creativity refers to a situation where the old pattern comes to be less constrained (in our case, collocations with an adjective or non-human noun). E-creativity, finally, is the stage when a new rule has been fully established. Since the distinction between the various categories of creativity is gradual rather than absolute, we use the term 'routinization' to refer to emerging patterns.

The remainder of this paper is organized as follows: in Section 2 we discuss how *wannabe* is an example of approximative morphology and we present our research questions in Section 3. In Section 4, we explain how we collected our corpus data and annotated them in Excel. After a survey of the quantitative results of our study in Section 5, we discuss the implications of our study for the debate on creativity versus routinization in Section 6, in particular with respect to the concepts of F1-, F2- and E-creativity. In Section 7, finally, we present our conclusions and offer suggestions for future research.

2 Approximative Morphology

2.1 Modality as a Source of Approximation

Approximation is a concept well studied in various subfields of linguistics, in particular pragmatics and discourse studies, but only recently has it gained attention within morphology (Masini, Norde and Van Goethem 2023). Following Masini &

Micheli (2020: 384–385), we understand APPROXIMATION as a complex functional do-main comprising different (although related) values such as FAKENESS, IMITATION, RESEMBLANCE or INTENTIONAL VAGUENESS. In Grandi & Körtvélyessy's model of evalu-ative morphology (2015: 11), APPROXIMATION involves a shift towards the 'negative' pole of evaluation. Examples from word-formation include compounding ele-ments expressing IMITATION/FAKENESS in Dutch (e.g., *kunstgras* 'artificial grass' or *namaak-wasabi* 'fake wasabi', cf. Van Goethem and Norde 2020), prefixal elements like Italian *simil-* (e.g., *simil-marsupio* 'sort of marsupium/pouch', *freddo simil si-beriano* 'Siberian-like cold', cf. Masini and Micheli 2020) as well as suffixes like English *(-)ish*, which is perhaps the most studied approximative morpheme to date (e.g., Oltra-Massuet 2017; Kempf and Eitelmann 2018; Eitelmann, Haugland and Haumann 2020; Eitelmann and Haumann 2023).

From these studies, it emerges that approximative morphemes may develop out of a variety of sources (Masini, Norde and Van Goethem 2023: 7–10) some of which are better recognized and explored than others. Well-studied sources in-clude words meaning 'fake', e.g. Greek *pseudēs* 'false', adopted in many other lan-guages as the prefix *pseudo-* (cf. Vassiliadou et al. 2023; Van Goethem, Norde and Masini 2023); diminutives (Dressler and Merlini Barbaresi 1994; Grandi 2017); de-gree and quantity items such as *quasi-* in English (e.g., *quasi-particle* cf. Cappelle, Daugs and Hartmann 2023) or *semi-* in Italian (e.g., *semi-official*; cf. Micheli 2023); spatial (proximity) items, such as *near-* in English (e.g., *near-synonym*; cf. Cappelle, Daugs and Hartmann 2023) or *para-* in Italian (from Greek *para* 'beside'; cf. Micheli 2023); similative items, such as English *-like* (*prefix-like*) or Italian *simil-* (Masini and Micheli 2020); relational items, such as the already-mentioned English *-ish* or Dutch *-achtig* (*groenachtig* 'green-ish').

A source that, to our knowledge, has not yet been explored in morphological studies on approximation, is modality. Nevertheless, examples are found in vari-ous languages, such as epistemic adverbs or volitional expressions. The concep-tual step here is from an assessment of speakers' (un)certainty about what they say (epistemic modality) or an aspiration to attain a (still unreal) specific state-of-affairs (volitional modality) to approximation as a process that expresses a quali-tative deviation from a standard or default value, towards the 'negative' pole (cf. Grandi and Körtvélyessy's model mentioned above).

As for epistemic adverbs, English *maybe*, for instance, is found in compound-like expressions conveying approximation, combined with both nouns (6) and ad-jectives (7). The case in (6) is especially interesting since adverbs are not supposed to modify nouns.

(6) *Not at all, I think these history threads are awesome and want to add my little bits of trivia. I'm just not willing to attach specific names to my admittedly fuzzy memory of the **maybe-incident**.* [enTenTen20]

(7) *The best of the bunch is alternate history drama "For All Mankind," from the creator of "Battlestar Galactica," but is it worth the money to see just one **maybe-good** show?* [enTenTen20]

The same applies to Italian *forse* and French *peut-être* 'maybe, perhaps', which are found in combination with nouns (mainly) (8), adjectives (9) and even with full phrases (10) (examples from itTenTen20 and frTenTen20):

(8) [Italian] *Sorella di mezzo ha diciott'anni, una sfilza di fratelli e sorelle, una madre bigotta che la vorrebbe già sposata, il ricordo di un padre morto di depressione, e un **forse-fidanzato**.* 'Middle Sister is eighteen years old, a string of siblings, a bigoted mother who would already like her married, the memory of a father who died of depression, and a **maybe-boyfriend**.'

(9) [Italian] *Da oggi dunque una città ricoperta di rumenta non sarà più colpa di scelte sbagliate, di strategie folli, di priorità errate e di impostazioni politico-ideologiche malate, no, sarà tutta colpa del misterioso incendio **forse-doloso**.* 'Starting today, therefore, a city covered in rubbish will no longer be the fault of bad choices, foolish strategies, wrong priorities, and sick political-ideological approaches, no, it will be all the fault of the mysterious **maybe-fraudolent** fire'.

(10) [French] *Typiquement le message qui fera fuir **un peut-être futur utilisateur de Linux** !*
'Typically the message that will scare off **a maybe-future Linux user**!'

Volitional expressions, too, can be employed as approximating strategies, often carrying a pejorative meaning derived from the 'aspiring' semantics, which implies the non-realization of a desired state. Examples are German *möchtegern* 'would very much like to' in (11), Italian *vorrei-ma-non-posso* in (12) (a lexicalized clause – literally meaning 'I would like but I can't' – used to refer, often ironically, to some aspiration impossible to achieve), and of course English *wannabe* (as in *wannabe popstar*). This *wannabe* construction is the case-study that we are going to discuss in detail in the following sections.

(11) [German] *Die Mehrzahl der Manuskripte fristete ein Schattendasein in den Schubladen frustrierter **Möchtegern-Schriftsteller**.* 'Most of the manuscripts were leading a shadow existence in the drawers of frustrated **wannabe-authors**' [deTenTen20]

(12) [Italian] *In questi giorni di mare e focaccia,* [. . .] *ho rivisto uno di questi teli utilizzato per coprire la vetrina di una **boutique vorrei-ma-non-posso** di paese, sempre in Liguria ovviamente.* 'In these days of sea and flatbread, [. . .] I saw again one of these cloths used to cover the window of a village **wannabe boutique**, again in Liguria of course.' [itTenTen20]

2.2 English *wannabe*

According to the OED (www.oed.com), *wannabe* is a slang (originally U.S.) word created by the univerbation of *wanna* (itself a univerbation of *want* and *to*) and *be* (cf. also earlier *would-be*, used as an adjective). *Wannabe* is registered:

(i) as a noun[2] ("A person who tries to emulate someone else, esp. a celebrity, in appearance and behaviour; a person who wants to belong to and tries to fit in with a particular group of people. Frequently *depreciative*"), when used either independently (13) or after another noun (14);

(ii) and as an adjective[3] ("Desiring or aspiring to be a specified person or type of person; would-be. Also in extended use of a product designed to emulate or rival another. Sometimes *depreciative*"), when preceding another noun (15).

(13) *Many at Cambridge will tell you that the drama world there is filled with ambitious, pretentious, bitchy **wannabes**.* [OED: S. Fry, Fry Chronicles 95, 2010]

(14) *A morbid Madonna-**wannabe** fascinated with tabloid tales of bizarre deaths.* [OED: Washington Post 17 January 1986 (Weekend section) 8/4]

(15) *There was little doubt that Bill Gates was intent on building a **wannabe** Mac for the vast PC market.* [OED: Rolling Stone 18 April 1996 61/1]

---

2 Oxford English Dictionary, s.v. "wannabe, n.", July 2023 (https://doi.org/10.1093/OED/9474186542).

3 Oxford English Dictionary, s.v. "wannabe, adj.", July 2023 (https://doi.org/10.1093/OED/3277987720).

This fluid categorial status tallies with Brinton & Traugott's (2005: 16) gradient view on categories, which is considered as "a factor motivating change, and also as the outcome of changes in usage".

According to the OED, the semantics of *wannabe* is often depreciative (equivalent to "phoney" or "fake") when used as a noun (cfr. (13)) and only "[s]ometimes *depreciative*" when used as an adjective. The first use of *wannabe* recorded in the OED traces back to 1976:

(16) *At 38 she had 21 years of racket life behind her. Whereas Joe, that year, was still a Jimmy Cagney **wannabe***. [OED: New York Magazine 26 July 1976 43/3]

Crucially, univerbated *wannabe* can still function as a verbal phrase, as illustrated in the following example, in which the second and third occurrences of *wannabe* function as verbal phrases syntactically:

(17) *Our children are not **wannabe** gangsters, they **wannabe** loved for who they are, and they **wannabe** great!* [enTenTen20]

3 Research Questions

In this study, we explore various construction types involving *wannabe* in six languages (English, Dutch, Danish, French, Italian and Finnish), with the aim to answer the following three research questions:

RQ1: How are the different construction types with *wannabe* distributed across each language?

RQ2: What are the formal and semantic properties of *wannabe* collocations?

RQ3: How creative are *wannabe* collocations?

We look at *wannabe* both as a free noun and in other constructions, such as combinations with other nouns or predicative usage. Our first research question therefore refers to the distribution across the six languages in our dataset. More specifically, we are interested in the collocational behaviour of *wannabe* cross-linguistically. As combinations of words are construed differently in the languages in our set (three Germanic languages, two Romance languages and one Finno-Ugric language), we avoid the term 'compound'. As we explain in more de-

tail in Section 4, the term 'collocation' may refer to compounds written as one word, such as Danish *wannabehjemmeside* 'wannabe homepage', but also to expressions such as English *wannabe starlet*[4] or French *écrivaine wannabe* 'wannabe author'. Our second research question, then, pertains to formal and semantic properties of *wannabe* collocations of all types. Our third and final research question relates to the topic of this collective volume and our three-way model of creativity outlined in Section 1. For each language in our sample, we consider the various construction types found in that language, in order to assess whether these patterns are F1-, F2- or E-creative.

4 Data and Methods

To answer the aforementioned research questions, we conducted a corpus-based study of *wannabe* formations in English, Danish, Dutch, French, Italian and Finnish. The CQL query [word = ".*wannab.*"] allowed us to extract formations starting or ending with the sequence *wannab*, including different spelling variants such as *wannabe(s), wannabee(s)* or *wannabie(s)*. We downloaded random samples of 1000 tokens per language from the TenTen web corpora at Sketch Engine (Kilgarriff et al. 2014).[5] These samples were then manually checked, and irrelevant occurrences were removed until we obtained 500 relevant occurrences per language. Occurrences considered irrelevant included metalinguistic reference to *wannabe*, references to the Spice Girls' *Wannabe* song, *wannabe* used in URLs and account names (e.g. *wannabe_god*), other strings containing *wannab** (e.g. *wannabike.com*), literally repeated sentences, and *wannabe* used in entire English sentences in the non-English datasets.

We then annotated those 500 occurrences for a diverse set of orthographic, morphological and semantic properties.

First, we annotated the **construction type** and distinguished the following categories:

---

4 English compounding differs from compounding in other Germanic languages in that English compounds may be written as one word (*oatcake*) or two (*cherry cake*) and it is a moot point whether multiword constructions are morphological or syntactic constructions. In spite of extensive literature on the topic, there appear to be no decisive criteria, like primary stress or compositional semantics (see Bauer 2019 for a review). See Section 4 for further discussion on terminology.

5 More specifically, we used the following corpora: enTenTen20 for English, daTenTen20 for Danish, nlTenTen20 for Dutch, frTenTen20 for French, itTenTen20 for Italian and fiTenTen14 for Finnish.

(i)   *Wannabe* **used as a noun** (18–19):

(18)   [Dutch] *Anders denken mensen dat je niet meer bent dan een* **wannabe**. 'Otherwise, people will think you are nothing but a **wannabe**.'

(19)   [Danish] *Resultatet er, at ingen er særlige, men at vi alle bare er kopier og* **wannabes**. 'As a result, nobody is special, we are all just copies and **wannabes**.'

(ii) *Wannabe* **used as an adjective**, either predicative (20) or attributive. To distinguish between attributive use and the "collocation" type (see below), adjectival use was restricted to unambiguous cases including coordination with another adjective (21), modification by a degree adverb (22), or specific adjectival constructions, for instance [*un* N *de* ADJ] in French (23).

(20)   [Danish] *Det er flot og anderledes på den fede måde – ikke bare* **wannabe**. 'It is nice and different in a cool way, not just **wannabe**.'

(21)   [English] *No disagreement on the point about* **actual and wannabe** *plutocrats either (. . .)*

(22)   [Danish] *Hvis man har oplevet La Santa, så forekommer Club Med (som jeg også har boet på en enkelt gang) som* **ren wannabe**. 'If you have experienced La Santa, then Club Med (where I also stayed a couple of times) appears **totally wannabe**.'

(23)   [French] *Ceci est* **un film d'action de wannabe** *(. . .)* 'This is **a wannabe action movie** (. . .)'

(iii) *Wannabe* **used as an 'adverb' or discourse marker** (24):

(24)   [Italian] *Seavessi non scrive per paura della commessa. Su Wikipedia alla voce wannabe, c'è la foto di Seavessi. Voi cosa* **wannabe** *vorreste?.* 'Seavessi [proper name from sentence 'if I had'] does not write out of fear of the clerk. On Wikipedia under wannabe, there is a picture of Seavessi. What **wannabe** would you guys like?'

(iv) ***Wannabe* as part of a "collocation"** (25–26). As already explained in Section 3, we preferred the term 'collocation' to 'compound' because the dividing line between morphological and syntactic formations is rather difficult to draw, even more so in written data that do not allow us to check the prosodic properties of the sequences. Moreover, the spelling criterion is not reliable either to identify true compounding, especially when combining an English loanword (*wannabe*) with native words.

(25) [Danish] *Måske **wannabe -'vampyren'** vil stikke sig selv en metaforisk pæl i hjertet (. . .)?* 'Maybe **the wannabe-'vampire'** wants to stick a metaphorical stake into his own heart?'

(26) [French] *Là par contre, j'aurais accepté sans rechigner que tu me traite de "pseudo-Charles Bronson du dimanche" ou de "**wannabe-Zorro** de l'internet" ou un truc comme ça (. . .)* 'On the contrary, I wouldn't have minded you calling me a "a Sunday pseudo-Charles Bronson" or a "**wannabe-Zorro** of the Internet" or something like that (. . .)'

(v) **Derivations of *wannabe*** (27):

(27) [Dutch] *Dit is wel keihard afrekenen met de **wannabe-ers** a.k.a privé lease-rijders.* 'This does deal harshly with the **wannabe-ers** a.k.a. private lease drivers.'

For the **'collocation' type**, which is the main focus of this study, we annotated some additional properties.

– The first one is **orthographic bonding of the collocations**, which distinguishes between free (28), bound (29) and hyphenated (30) collocations, or a 'hybrid' combination of some of the former spelling strategies (31).

(28) [French] *Soufron, vous n'êtes même pas un bobo, vous êtes ce que j'appelle un **wannabe bobo***. 'Soufron, you're not even a bobo, you're what I call a **wannabe bobo**.'

(29) [Dutch] *Ook beschikte de 37-jarige **wannabepiloot** niet over een vliegbrevet.* 'Also, the 37-year-old **wannabe pilot** did not have a pilot's license.'

(30) [Danish] *Musikken lyder måske mest som hvis nu Robbie Williams' sangskri-verteam havde skrevet en **wannabe-Oasis-sang***. 'The music sounds perhaps mostly as if Robbie Williams' song-writing team had written a **wannabe-Oasis song**.'

(31) [Danish] *Jeg synes, vi skal droppe det der **"nationalsangs wannabee"** fuld-stændig*. 'In my view, we should get rid of this "**national anthem wannabe**"'

– Second, we annotated the **part of speech (POS) of the collocate**, as well as its **lemma form** to facilitate productivity calculations (number of types and ha-paxes). The following examples illustrate collocates in the form of a (proper) noun (32), a noun phrase (33), an adjective (34) and a verb phrase (35).

(32) [Danish] *Sproget – en såkaldt "indre monolog" – lyder, som var det skrevet af en **James Joyce-wannabe***. 'The language – a so-called 'inner monologue', sounds as if it had been written by a **James Joyce wannabe**.'

(33) [Dutch] *Een aanrader voor elke **wannabe journalistieke detective*** 'Highly recommended for any **wannabe journalistic detective**'

(34) [Finnish] *Nonniin se niistä mysteerisistä **wannabe-aistikkaista** ajatuksista*. 'Well, so much for those mysterious, **wannabe-tasteful** thoughts.'

(35) [Dutch] *(. . .) hij wordt een dagje ouder en 'T is natuurlijk volop in zijn **wan-nabe man zijn** periode* '(. . .) he's getting a day older and is obviously fully into his **wannabe manhood** [lit. man be] period'

– Third, we annotated the **word order** of the collocations, distinguishing be-tween three schemas: [X-*wannabe*], [*wannabe*-X] or [X-*wannabe*-X]. For the first two options, see for instance the difference between (32) and (33). We find the third pattern in (36).

(36) [Italian] *Kiefer Sutherland non mi è mai stato troppo simpatico, probabil-mente perchè in questo film interpretava un **vampiro-wannabe-biker**, (. . .).* 'I've never liked Kiefer Sutherland too much, probably because in this film he played a **vampire-wannabe-biker**, (. . .).'

– A final formal parameter included in the annotations is the **presence of in-flection on *wannabe***. We find both native and borrowed (= English) inflec-

tion: see (37) for number and case inflection in Finnish, (38) for English plural inflection for number in Italian. However, since the vast majority of results are in the (uninflected) singular, we decided not to include this parameter in our quantitative analysis.

(37) [Finnish] *(. . .) me täällä ollaan niiden rinnalla ihan **wannabeita***. 'We here are just **wannabes** compared to them.'

(38) [Italian] *questi **wannabees** possono essere noiosi* 'these **wannabees** can be boring.'

When *wannabe* collocates with a noun, we distinguished three **semantic categories**: HUMAN to refer to human beings (living, dead, or fictional) or groups of human beings, such as rock bands, as in (39); ANIMATE to refer to all living creatures that are not human (animals and fantasy figures such as vampires or zombies) as in (40); and INANIMATE to refer to objects, food, abstract concepts etc., as illustrated in (41).

(39) [French] *La vidéo montre des mineurs, parfois adolescents, défiler devant un parterre de **wannabe parents***. 'The video shows minors, sometimes teenagers, parading in front of an audience of **wannabe parents**.'

(40) [Danish] *den får de to af monsteret/ **wannabe zombien** der kommer i slutningen* 'it gets the two of them from the monster/ **wannabe zombie** that comes at the end'

(41) [Dutch] *Gember heeft een eigenschap die het ver boven alle **gember-wannabe** middeltjes plaatst.* 'Ginger has a property that places it far above all **ginger-wannabe** remedies.'

We now turn to the results section, where the focus will be on the collocates in the 'collocation' type.

5 Results

In this section, we present the results of the case studies, starting with the different construction types in which *wannabe* is used and then focusing on the formal and semantic properties of a specific construction: *wannabe* collocations.

5.1 Construction Types

Table 1: Construction types.

Construction type	English	Danish	Dutch	French	Italian	Finnish
collocation	357	342	370	294	340	393
collocation-derivation	1	7	2	0	0	2
derivation	0	4	2	6	2	1
embedded collocation	4	30	11	3	4	34
predicative	2	5	12	6	4	19
wannabe_ADJ	3	1	3	9	12	1
wannabe_N	132	111	100	182	135	50
wannabe_V	0	0	0	0	2	0
wannabe_ADV	0	0	0	0	1	0
wannabe_VP	1	0	0	0	0	0
Total	**500**	**500**	**500**	**500**	**500**	**500**

As shown in Table 1, *wannabe* is used in a variety of construction types in the data samples, but frequencies differ from sample to sample. Collocations with *wannabe*, as in (42), form the biggest part of all language samples (a more detailed analysis of collocational constructions follows in sections 5.2 and 5.3).

(42)　[English] *I am a **wannabe rocker** trapped in folky singer-songwriter schtick.*

Apart from "simple" collocations, all languages also feature the use of *wannabe* in embedded collocations, such as (43), where the collocation *Elvis Presley wannabe* is used to modify the noun *frisure* 'haircut'. Such embedded collocations were found predominantly in Finnish and Danish (and to a certain degree in Dutch) and point to a more synthetic use of *wannabe*.

(43)　[Danish] *(. . .) og derfor er det ikke så underligt, at Cage denne gang er ude i en **Elvis Presley wannabe-frisure**.* 'and therefore it is not so remarkable that Cage this time is sporting an **Elvis Presley-wannabe-haircut**.'

(44)　[Finnish] *Tämä on sellainen **wannabe-feather and fan-kaulaliina** (. . .)* 'This is such a **wannabe-feather and fan-scarf** (. . .)'

The Romance languages, by contrast, show more analytic uses of *wannabe* as independent noun (45) or adjective (46).

(45) [Italian] *Ormai, il web è divenuto solo una vetrina per **wannabe** piuttosto sfigati o – nel migliore dei casi – un mezzo a disposizione di chi già scrive.* 'By now, the web has become only a showcase for rather lame **wannabes** or – at best – a medium available to those who already write.'

(46) [French] *D'entrée de jeu, oubliez tout de suite la selfie prise à bout de bras dans le salon. Ça fait très **wannabe** et vous ne voulez pas ça!* 'Forget about the arm's length selfie in the living room. It looks very **wannabe** and you don't want that!'

Whereas the English dataset almost exclusively consists of the use of *wannabe* in collocations and in nominal use, a more varied range of construction types can be observed in the recipient languages. These include the use of *wannabe* as a verb (47) and the use of *wannabe* or a *wannabe*-collocation as bases for derivation (48–49).

(47) [Italian] *Detto questo:-D io **wannaberei** . . . dire al mio capo (e a mio marito) quanto mi faccia schifo fare il mio lavoro e che mi licenzio* 'That said:-D I **would like to** . . . tell my boss (and my husband) how much I hate my job and that I'm quitting.'[6]

(48) [Danish] *Nogle synes at det helt vildt sjovt at bruge engelske låneord, andre at det lyder **wannabeagtigt**.* 'Some think it is totally cool to use English loanwords, others think it sound **wannabe-like**.'

(49) [Finnish] *Kirjoittajan artikkeli sinänsä on minun mielestäni suurelta osin tyhjänpäiväistä **carriebradshaw-wannabe-maista** läpinää, enkä muuten halua sitä kommentoida [. . .]* 'The author's article itself is, in my opinion, largely empty **Carriebradshaw-wannabe-like** nonsense, and I don't want to comment on it [. . .]'

The use of *wannabe* in a more varied range of constructions can be seen as creative extensions of *wannabe* word formation patterns in the recipient languages; however, the frequencies of construction types only found in the recipient languages (but not in English) are very low.

---

6 This example is peculiar because *wannabe* is used to mean *would like (to)*, with the inflectional suffix *-erei* (1.SG.COND) attached to it, marking its verbal status.

5.2 Formal Properties of *Wannabe* Collocations

5.2.1 Word Order

	English	Danish	Dutch	French	Italian	Finnish
X-wannabe-X	0	0	0	2	4	0
X-wannabe	137	150	65	94	164	33
wannabe-X	225	229	318	201	176	396
n/a	138	121	117	203	156	71
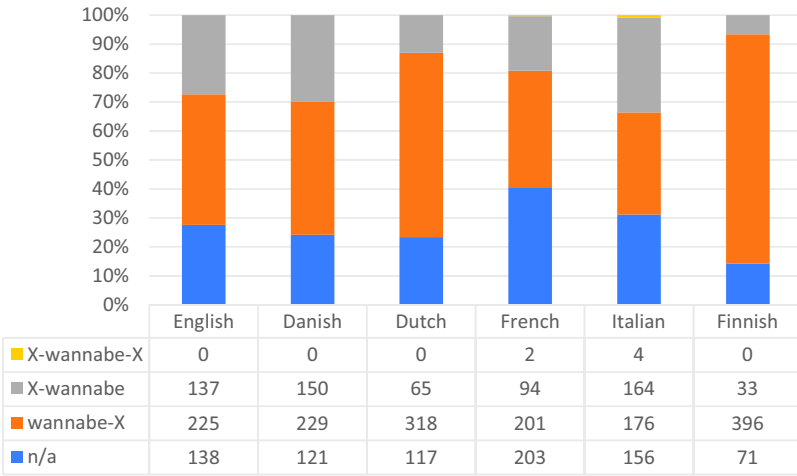
Figure 1: Word order.

Figure 1[7] illustrates the proportions of the three word-order patterns. Since *wannabe* acts as an approximative modifier in the collocation type, we would expect to find [*wannabe*-X] order in the Germanic languages, because of their general Modifier-Head structure, and the opposite order [X-*wannabe*] in the Romance languages, which are characterized by Head-Modifier structure.

    With respect to Germanic (English, Danish, Dutch), we indeed notice a preference for [*wannabe*-X] order, but, especially in English and Danish, also a surprisingly high proportion of the inverse word order (see for instance (50)). The status of *wannabe* in such instances is not crystal-clear: does it function as the head of the sequence or as an approximative marker? This issue needs further investigation.

(50)   [English] *Some have described Departed as feeling like a **Scorsese-wannabe** created it*

---

7 In this Figure and elsewhere, "n/a" stands for "not applicable"; in this case, it refers to *wannabe* constructions that are not collocations and hence do not have word order.

The [*wannabe*-X] order is clearly predominant in Finnish, as illustrated in (51).

(51) [Finnish] *(. . .) onko tää joku median neuvo* **wannabejulkkiksille** *että jos tahdotte saada huomiota niin provosoikaa mahdollisimman paljon.* '(. . .) is this some media advice to **wannabe celebrities** that if you want to get attention, provoke as much as possible'

Whereas the inverse order [X-*wannabe*] was expected in the Romance languages, Italian and especially French seem to show a high proportion of 'pattern borrowing' (Gardani 2020) of the English word order pattern. Moreover, contrasting examples like (52–53) suggest that the word order pattern is quite arbitrary in French.

(52) [French] *Il n'est plus un* **Vador-wannabe**, *il est Kylo Ren.* 'He is no longer a **Vader-wannabe**, he is Kylo Ren.'

(53) [French] *Kylo devient définitivement le* **wannabe Vador** *Fucky Ryan* 'Kylo definitely becomes the **wannabe Vader** Fucky Ryan'

[X-*wannabe*-X] is very rare, it is only found a couple of times in the French and Italian datasets. An example is given in (36) above.

5.2.2 Orthographic Bonding

With respect to orthographic bonding, we do not find clear routinization in this field, but some general tendencies can be detected. Overall, we notice more free uses in English and Romance (see for instance *wannabe rocker* (42) for English and *wannabe bobo* (28) for French) and more bound (especially hyphenated) cases in the other Germanic languages and Finnish (see for instance *wannabepiloot* 'wannabe pilot' (29) for Dutch and *wannabe-aistikkaista* 'wannabe tasteful' (34) for Finnish).

These results are in line with the assumption that Romance languages are more analytic than Germanic languages and that, within the Germanic language family, English is situated at the analytical end of the cline (see, e.g., Van Haeringen 1956; Hüning et al. 2006; Lamiroy 2011; König and Gast 2018; Van Goethem, Norde and Masini 2025).
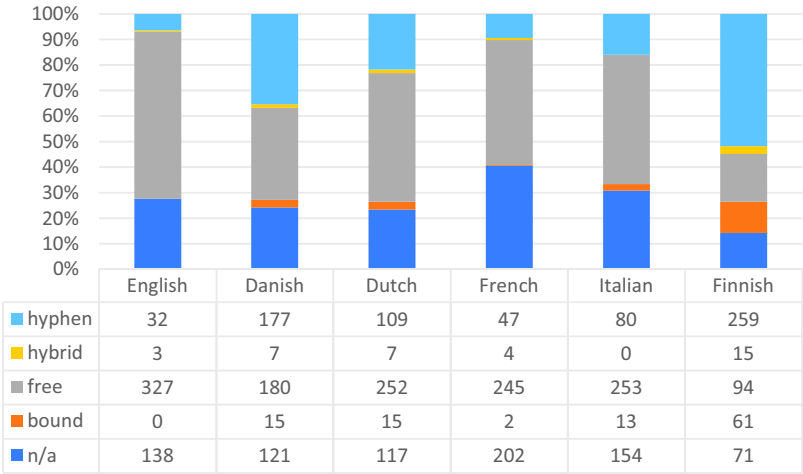
Figure 2: Orthographic bonding.

	English	Danish	Dutch	French	Italian	Finnish
hyphen	32	177	109	47	80	259
hybrid	3	7	7	4	0	15
free	327	180	252	245	253	94
bound	0	15	15	2	13	61
n/a	138	121	117	202	154	71

5.2.3 Part of speech of the collocate

	English	Danish	Dutch	French	Italian	Finnish
VP	0	0	1	0	0	0
V	0	0	3	0	0	0
Pron	0	1	0	1	5	2
N-prop	91	110	97	82	111	50
NP	29	11	14	28	31	3
N/ADJ	0	0	0	4	0	0
N	237	245	254	187	181	352
Interjection	0	0	0	1	0	0
ADJ	5	13	17	2	28	23
n/a	138	120	114	195	144	70
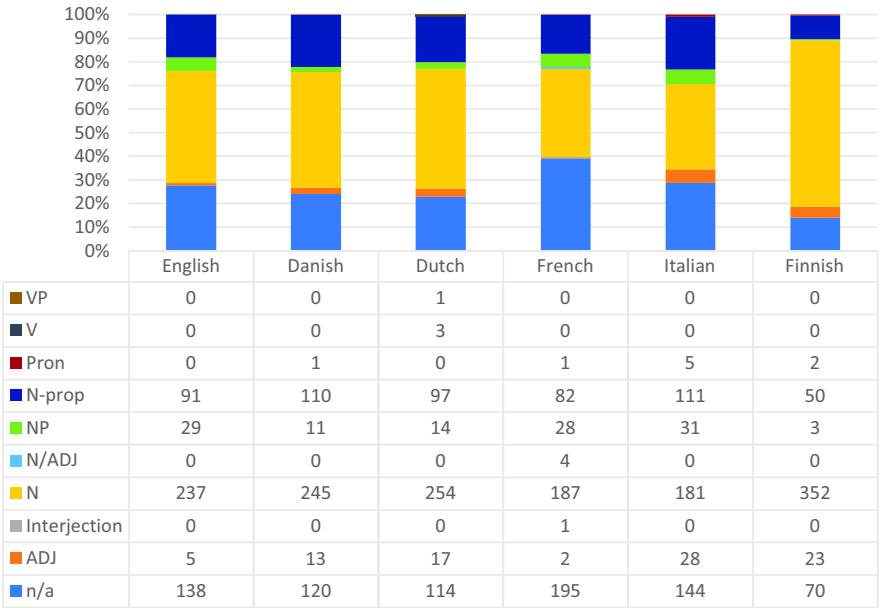
Figure 3: Part of speech of the collocate.

With respect to the part of speech of the collocate, a first observation is that in English *wannabe* almost exclusively combines with nominal collocates (N, Proper N, NP), as in (54).

(54) [English] *Is 'whole hearty wheat' really a burden to good health and **wannabe beach bodies**?*

However, as Figure 3 shows, in the recipient languages we also find creative expansion to other categories, such as adjectives (55), pronouns (56) and verbs (57).

(55) [Italian] *Gli attori dietro gli imbarazzanti personaggi sono altrettanto allucinanti, specialmente considerando una trama **wannabe impegnata** con risvolti drammatici nelle vite di ognuno.* 'The actors behind the awkward characters are equally terrible, especially considering a **wannabe committed** plot with dramatic implications in everyone's lives'

(56) [Italian] *Da anni ormai infatti la kermesse cinematografica fa da vetrina a una pletora di attori, cantanti, celebrità e **wannabe tali**, (. . .).* 'For years now, in fact, the film festival has served as a showcase for a plethora of actors, singers, celebrities and **wannabe such**, (. . .)'

(57) [Dutch] *Ja, dat noemen we **wannabe modden**.* 'Yes, that is what we call **wannabe modding** (moderating).'

Moreover, we find routinization of proper names in all languages (for instance *Scorsese-wannabe* in (50)), if to a lesser degree in Finnish. In Section 5.3, we will have a closer look at the semantics of these proper names.

5.2.4 Productivity

In this section, we only consider the productivity of collocations with common nouns.[8] We consider *wannabe* collocations to be schemas with one fixed slot and

---

[8] As we demonstrated in Section 5.1, many collocations involve proper names. We did not consider these constructions in our productivity measures, because proper names have unique reference, i.e. they describe specific individuals and properties that are associated with them (Ainiala and Östman 2017: 6; Kosse 2021: 1). Adding them to the slot filler types would mean a substantial increase in the number of hapaxes, resulting in a higher potential productivity than is warranted for this schema.

one open slot and we use productivity measures to compare the lexical diversity of the slot fillers in the different language samples. As the figures in Table 2 show, the vast majority of collocations consists of hapaxes (i.e. types that occur only once in the sample). Collocations with a token frequency higher than 1 are rare and have maximum frequencies between 4 and 6 for most languages. Only Finnish has one type (*kirjalija* 'author') with a slightly higher token frequency of 17.

The large proportion of hapaxes results in high productivity ratios, not just the Type / Token Ratio (TTR), but also the Hapax / Token Ratio, or Potential Productivity (PP) and the Hapax / Type ratio (HTR). 'Potential Productivity' (Baayen 2009) can be interpreted as an indication of (potential) vocabulary growth. Especially in large corpora, hapaxes can be seen as a proxy for neologisms (Zeldes 2013: 61), so that a high Hapax / Type ratio suggests that a pattern is relatively new with great potential to expand. The figures in Table 2 suggest high productivity of *wannabe* collocations in all languages.

Table 2: Productivity.

Productivity	English	Danish	Dutch	French	Italian	Finnish
TTR	318/362 = **0.88**	336/380 = **0.88**	330/390 = **0.85**	252/306 = **0.82**	282/340 = **0.82**	345/430 = **0.80**
PP	291/362 = **0.80**	303/380 = **0.80**	292/390 = **0.75**	219/306 = **0.71**	253/340 = **0.74**	305/430 = **0.71**
HTR	291/318 = **0.92**	303/336 = **0.90**	292/330 = **0.88**	219/252 = **0.87**	253/282 = **0.89**	305/345 = **0.88**

5.3 Semantic Category of the Collocate

The last property we investigated is the semantic category of the collocate. As illustrated in Section 4, we used a rather coarse-grained classification, distinguishing three semantic classes: HUMAN, ANIMATE, and INANIMATE.

As we can see in Figure 4, collocates in English belong predominantly to the human category (cf. (58)), whereas the proportion of human collocates in the recipient languages is lower: except French, all other languages display a creative extension to inanimate collocates (animate ones are very marginal everywhere), which is more marked in Danish and Italian, and slightly less prominent in Finnish and Dutch.
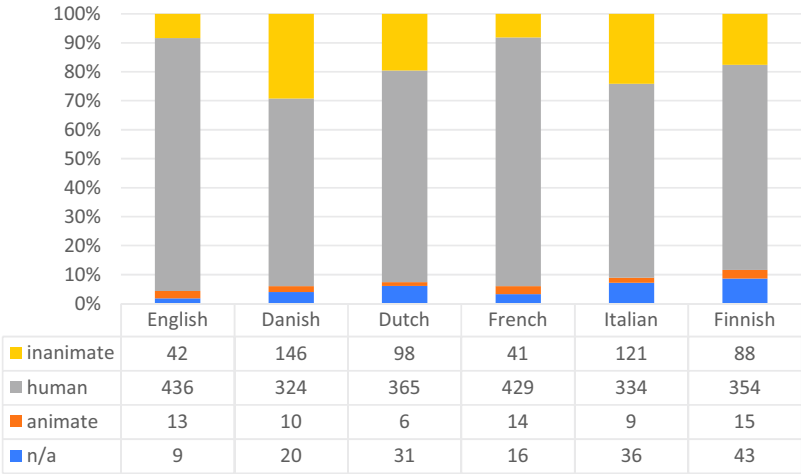
Figure 4: Semantic category of the collocate.

(58) *But I don't agree with the suggestion that every **wannabe-blogger** must opt for Aweber.*

The combination with inanimate entities is an indication of the change *wannabe* is experiencing in these contexts: the original 'aspiring' meaning (cf. (58)), hinting to a possible future state-of-affairs to attain, is not compatible with inanimate entities that, being non-sentient, cannot aspire to anything, strictly speaking. This triggers a more abstract interpretation of 'aspiration' as deviation from a norm or standard, hence the approximative semantics. The contrast can be seen in examples (59–60), which contain the same expression (*wannabe actress*): in (59), Jonni Kendall is not an actress but an aspiring one, whereas in (60) Katrina is an actress (since she is indeed already acting), but not a full-fledged, famous one (because she is acting in sleazy films). This is also evidence that the approximative meaning applies to human nouns too.

(59) [. . .] *Jonni Kendall, a nineteen-year old **wannabe actress**.*

(60) [. . .] *'Sheila Ki Jawani' was supposed to be a raunchy item Bhojpuri kind of song as Katrina's a **wannabe actress** acting in sleazy films* [. . .]

Whereas human collocates are typically common nouns referring to (desirable) occupations (e.g., influencer, actress, blogger, author, etc.) or proper names of celebrities (e.g. Madonna or Elvis as exemplified above), inanimate collocates be-

long to a variety of semantic subcategories, some of which are however recurring, namely: food items (61), or inanimate proper names such as (video) games (62), movie and series titles (63), or brand names (phones, cars, fashion houses, food chains, etc.) (64–65). These inanimate collocations suggest the emergence of specific semantic clusters, which we discuss in Section 6.

(61) [Danish] *Total bagedag, nu også med **marmorkage-wannabe***. 'A full baking day, now including a **marble cake wannabe**.'

(62) [Italian] *Journey è un **mario64 wannabe**, si salta in giro però è lento, scomodo e non succede una cippa* 'Journey is a **mario64 wannabe**, you jump around but it is slow, uncomfortable and nothing happens'

(63) [French] *voici un article qui tente d'expliquer le succès d'audience de Revolution, comparée aux autres **Lost-wannabes*** 'here's an article that tries to explain Revolution's audience success, compared to other **Lost-wannabes**'

(64) [Italian] *La moda costa. Magari non prendevo Gucci, ma meglio H&M di un **wannabe Vuitton***. 'Fashion costs money. Maybe I wouldn't get Gucci, but better H&M than a **wannabe Vuitton**'.

(65) [Dutch] *We sluiten af met een bezoek aan een **wannabee Macdonalds**, Jollibee, hier the place to be voor iedereen, waar ze overheerlijke spaghetti met friet verkopen (. . .)* 'We conclude with a visit to a **wannabe Macdonalds**, Jollibee, here the place to be for everyone, where they sell delicious spaghetti with fries (. . .)'

It is perhaps not surprising to find many inanimate proper names in our data: famous brands and popular products are easily interpretable as something to aspire to (just like celebrities), which explains why *wannabe*'s creative expansion is going in this direction. However, we also find examples that don't comply with this, like (66–67), which better exemplify the approximative semantics of *wannabe*.

(66) [Dutch] *Een zweetdruppel loopt langs mijn **wannabe-bakkebaarden** en een andere teistert mijn linkerwenkbrauw (. . .)* 'A drop of sweat runs down my **wannabe sideburns** and another plagues my left eyebrow (. . .)'

(67) [Finnish] *Kun ei parempaa ole, niin olen sitten vanhoja joulukortteja koonnut* **wannabe-kuuseksi** *oveen.* 'For want of anything better, I assembled old Christmas Cards into a **wannabe-spruce** on the door.'

5.4 Summary and Analysis

In this section, we address our first two research questions and summarize our findings from the preceding sections, concluding with a statistical analysis of the differences between English and each of the recipient languages. The first two RQs are repeated here for convenience:

RQ1: How are the different construction types with *wannabe* distributed across each language?

RQ2: What are the formal and semantic properties of *wannabe* collocations?

Overall, our comparative analysis brought to light substantial variation in *wannabe* constructions. Collocations prevail in all languages, ranging from 294 tokens in French to 393 tokens in Finnish. Conversely, *wannabe* as an independent noun is most common in French (182 tokens) and least in Finnish (50 tokens). When we look at the semantic categories of *wannabe* constructions (both of *wannabe* as a free noun and *wannabe* collocations), HUMAN is the most common category, although the proportion is higher in English than in the recipient languages (except French).

Turning now to *wannabe* collocations, we examined five properties: word order, type of (orthographic) bonding, part of speech of the collocate, productivity, and semantic category. Word order, it turned out, differs from language to language, with a strong preference for *wannabe*-X in Dutch and Finnish, a less pronounced preference in English, Danish and French, and an almost 50/50 distribution in Italian. Where type of bonding is concerned, we found substantial differences – collocations are overwhelmingly non-bound in English, Dutch, French and Italian, whereas Danish and, especially, Finnish prefer hyphens in *wannabe* collocations. The part of speech of the collocate is less varied – nouns form the majority, with proper names coming second and adjectives third. Adjectives are however less common in English than in the recipient languages (except French). Differences are also relatively small in productivity scores, which were invariably high, both in terms of Type Token Ratio, Hapax Token Ratio (Potential Productivity) and Hapax Type Ratio. The high number of hapaxes in particular suggests

that *wannabe* collocations are very productive in all languages, with as yet little entrenchment of specific types.

To conclude our quantitative analysis, we compared the recipient languages to English by performing a Fisher's exact test for all formal and semantic properties discussed in Sections 5.2 and 5.3. The results are given in Table 3, where each cell gives the result of a pairwise comparison of a specific variable in English and one of the other languages.[9] We note that for most of these comparisons the difference is significant, with the exception of French, where we did not find statistically significant differences for part of speech, word order and semantic category.[10]

Table 3: The recipient languages compared to English.

	Danish	Dutch	French	Italian	Finnish
Construction type	***	** / *	*** / **	ns	***
Orthographical bonding	***	***	** / *	***	***
Collocations: Part of Speech	** / *	** / *	ns	***	***
Collocations: Word order	ns	***	ns	** / *	***
Collocations: Semantic category	***	***	ns	***	***

6 Discussion: Creativity or Routinization?

In this section, we return to the concepts of F1-, F2- and E-creativity introduced in Section 1, in order to answer our third and final research question:

RQ3: How creative are *wannabe* collocations?

---

9 The asterisks in the p-value refer to the significance level (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$), ns means non-significant ($p > 0.05$). For these calculations, we only considered values that were found in all languages, disregarding very infrequent patterns (e.g. *wannabe* as an adverb) that had zeros in most cells.

10 One reviewer suggests that we do a Bonferroni correction on the signficance tests to avoid Type I errors (finding an effect when in fact there is none), arguing that we are doing various tests on the same data. The Bonferroni correction is however not uncontroversial (Cabin & Mitchell 2000) and moreover each cell in Table 3 represents a different cross-tabulation. For these reasons, we are not sure whether the correction is necessary, but we did add the values for the Bonferroni corrections nevertheless. In our case, this meant dividing the significance thresholds by 5 (the number of variables in Table 3), which means the following significance levels: *: $p < 0.01$; **: $p < 0.002$; ***: $p < 0.0002$. If the resulting significance level changes as a result of this (which in most cases it did not), the adapted value is given after the /.

To account for the patterns we found in our data, we adopt a constructionist approach to morphology (Booij 2010; Booij 2016; Masini & Audring 2019) and morphological change (Norde & Trousdale 2023). In Construction Morphology, complex words "are not seen primarily as a concatenation of morphemes, but as independent meaningful units within which certain subcomponents (morphemes) may be distinguished on the basis of paradigmatic relations with other words" (Booij 2018: 4–5). Generalizations over complex words take the form of schemas, which express the symbolic relation between the formal and the semantic pole of a morphological construction. Schemas are often partially schematic, with an open slot and a filled one. Two constructions with *wannabe* as a fixed slot are given in (68) and (69): in (68), *wannabe* follows a personal name (e.g. *Madonna wannabe*), whereas in (69) it precedes an adjective.

In these schemas, the double arrow represents the symbolic relation between form and meaning, the small caps subscripts denote part of speech, subscripts $_i$ and $_j$ are lexical indices and {-,␣ } denotes that the collocation can be written either as a single word, with a hyphen, of with a space in between (see Section 5.2.2).

(68) $[[[x]_{PN i}\{,\text{-},␣ \}[wannabe]]_j \leftrightarrow [\text{aspiring to be like SEM}_i]_j]$

(69) $[[[wannabe]\{,\text{-},␣ \}[x]_{ADJ i}]_j \leftrightarrow [\text{approximating SEM}_i]_j]$

Schemas are output-based and flexible, on the usage-based assumption that linguistic knowledge is shaped by usage and keeps on changing across the life-span in what Kemmer & Barlow (2000: ix) have termed a 'feedback loop'. Schemas are furthermore assumed to be organized in a single network (Bybee 2010; Diessel 2019), the construct-i-con (cf. Jurafsky 1992: 8; Goldberg 2019: 36). On this view, there is no distinction between word formation rules and lexical items, because the construct-i-con contains both morphological schemas and specific instantiations. This implies complex and multi-dimensional networks with paradigmatic links between constructions (of varying complexity and schematicity), by virtue of formal and / or semantic similarities. In the case of *wannabe*, not only are the schemas in (68) and (69) linked to each other (because they share a fixed slot), but also to other constructions involving *wannabe*, or to other approximative constructions, to other constructions involving personal names, and so on.

A network approach offers a possible explanation for attested differences between English and the recipient languages. Univerbation of *want to be* took place in English, and *wannabe* can still be used as a verbal phrase (see Section 2.2.). In the other languages, *wannabe* was borrowed both as a noun and in specific collocations, but not as a verbal phrase (for instance, Dutch ᵗ*ik wannabe* (*een ster*) 'I wannabe (a star)' is not found in the nlTenTen20 corpus and sounds decidedly

odd). As a result, the place of *wannabe* in the construct-i-con and the way it is linked to other constructions is different in English. This may account for the observation that we find more collocations with human reference in English than in a language like Danish, where *wannabe* is more often found in combination with non-human collocations than in English. Interestingly though, we do not see network expansion (reflected by high type frequency) overall, but rather the emergence of specific clusters, e.g. food items in Danish, which suggest particular 'pockets of productivity' (Cappelle 2014).

The different network configurations for English as opposed to the recipient languages may also account for the differences in frequency distributions. We noted earlier that the high proportion of hapaxes reflects a schema with a high potential to expand, to include both new construction types and collocations with native nouns. As the recipient languages are not 'hampered' by the link to the VP *want to be*, as explained above, *wannabe* can perhaps more easily be combined with non-human nouns, where the implication 'to aspire to (but probably in vain)' is no longer possible. With inanimates, *wannabe* has acquired a more general meaning of approximation / fakeness but, because these constructions are still linked to *wannabe* constructions with human collocates, the latter cluster is also affected by the semantic shift in the former. As mentioned in Section 2.1, approximation can, in specific contexts, be interpreted as depreciation (Masini, Norde and Van Goethem 2023: 11). We find this with *wannabe* as well – see the two possible interpretations of *wannabe actress* in examples (59–60) above, or (70), where Stockholm is not so much approximating the concept of a capital city (it is the capital of Sweden after all), but it is dismissed as a bad specimen, much less worthy than Göteborg (Sweden's second largest city). The depreciative sense can also be due to an association with the free noun *wannabe* as a 'phoney, disingenuous person' (see example (1)), as appears to be the case in example (71). This example, where *wannabe* is purely pejorative, is found in the context about Vietnam veterans inflating their heroism during the war.

(70) [Danish] *Stockholm er en **wannabe-hovedstad**. I følge Tomas er det selvfølgelig Göteborg, der er den rigtige hovedstad i vores naboland.* 'Stockholm is a **wannabe capital**. According to Tomas, Göteborg is, naturally, the true capital of our neighbouring country'

(71) [English] *"Who'd have thought that thirty years after the war all the records would be available to the public" Exactly, who ever thought that would happen? Certainly not all the **wannabe arseholes** who'd been getting away with their crap for that many years.*

Apart from differences between English on the one hand and the recipient languages on the other, we also found differences within the group of recipient languages. In all languages, we saw both matter borrowing (e.g. *wannabe bad boy* found in the French sample) and pattern borrowing, where *wannabe* collocates with native words (e.g. French *"wannabe"-sauveurs de la terre* 'wannabe saviours of the earth'). When there is more matter borrowing, we may expect stronger similarities to English in terms of word order, semantic category of the collocation and / or part of speech of the collocation.[11] At the same time, typological factors may be at play, particularly word order. Compounds in Germanic languages are generally head-final, which is why differences in word order are non-significant between English and Danish or Dutch. If, however, we find a similar non-significant result for French, which is head-initial, this implies stronger links to collocations that are matter borrowing, and hence (?) a lower degree of creativity in French.

Turning now to our model of F1-, F2- and E-creativity mentioned in Section 1, we review the most common English *wannabe* constructions in Table 4.

Table 4: English *wannabe* constructions and creativity.

	Construction type			Creativity
a	$[[[x]_{PNi}\{-,\_\}[wannabe]]]_j$ *Hitler-wannabe*	↔	$[\text{aspiring to be like SEM}_i]_j]$	F1
b	$[[[wannabe]\{-,\_\}[x]_{PNi}]]_j$ *wannabe Napoleon*	↔	$[\text{aspiring to be like SEM}_i]_j]$	F1
c	$[[[wannabe]\{-,\_\}[x]_{N\{human\}i}]]_j$ *wannabe gangsta*	↔	$[\text{aspiring to be like SEM}_i]_j]$	F1
d	$[[[wannabe]\{-,\_\}[x]_{N\{inanimate\}i}]]_j$ *wannabe road*	↔	$[\text{approximating SEM}_i]_j]$	F2
e	$[[[wannabe]\{-,\_\}[x]_{Ni}]]_j$ *wannabe gangsta / road*	↔	$[\text{approximating SEM}_i]_j]$	F1

---

11 We did not test differences in matter borrowing statistically, because there is no strict demarcation between matter borrowing and native words. Many loans, e.g. *hiphop* or *trendy* in Danish, are not necessarily perceived by speakers as foreign matter, which makes annotation for etymological origin of the collocate extremely difficult. Therefore, we treat matter borrowing as tendencies rather than hard figures.

Table 4 (continued)

	Construction type			Creativity
f	[[[*wannabe*]{,-,␣ }[x]$_{Ni}$]]$_j$ *wannabe arsehole*	↔	[bad specimen of SEM$_i$]$_j$]	F2 (> E)
g	[[[*wannabe*]$_{ADJ}${,-,␣ }[x]$_{Ni}$]]$_j$ *actual and wannabe plutocrats* *I don't think she's wannabe*	↔	[approximating SEM$_i$]$_j$]	F2 (> E)

Starting with collocations with personal names (a and b), we argue that both are F1-creative, i.e. fully productive and conventionalized word formation schemas.[12] Our synchronic data suggest that the same is true for collocations with nouns denoting humans (c). When the semantic category of the noun is not HUMAN, however, as in (d), we argue that this is an F2-creative extension. Because of the phrasal origin of *wannabe* (from *want to be*), which requires a human subject, we reconstruct a HUMAN > INANIMATE pathway. While such inanimate collocates are initially F2-creative, they may eventually lead to the more general schema (e), where the constraint that the *wannabe* collocate is HUMAN has been lifted. However, there is no new schema, so we argue that (e), once conventionalized, continues the F1-creativity of (c). In other words, these are cases of creativity followed by routinization. A different kind of F2-extension is found in (f): here, the semantic change is not in the collocate but in *wannabe* itself: a *wannabe arsehole* is not someone aspiring to be an arsehole, but just a terrible specimen of the collocate. This too is an incremental change, because depreciation is sometimes implicit in approximation, but it does imply that *wannabe* is becoming polysemous and if this goes hand in hand with a category change from noun to adjective, we have a new, E-creative schema. Finally, constructions that are clearly adjectival, like those in (g) can also be seen as F2-extensions from (e), but here, too, one might argue that the category change from noun to adjective is E-creative, especially in predicative constructions where *wannabe* is used independently (i.e. not as part of a collocation).

In the other languages, we see similar patterns, but changes in frequencies suggest that some languages are further advanced than English. Danish, for example, has less HUMAN collocates, which can be interpreted to mean that (e) (without

---

12 Because (a) predates (b) in the OED and has higher type frequency, (b) may originally have been F2-creative, but this requires a further diachronic study.

the HUMAN constraint) is more strongly entrenched in Danish than in English. Since we also find more examples of adjectival usage, we may similarly argue that (f) and (g) are more entrenched and hence further down the E-creative road.

This view on creativity poses two challenges. The first one is that change is incremental, so that the line between F- and E-creativity is not easy to draw (Ungerer and Hartmann 2023: 44). The second challenge concerns the relation between creativity and extravagance (see for instance the papers in Eitelmann and Haumann 2022a). Where Eitelmann and Haumann (2022b: 5) see E-creativity as coming close to extravagance, Hoffmann (Hoffmann 2020) argues that some cases of E-creativity, e.g. the development of English *to be going to* as a future auxiliary, are not deliberate.

7 Conclusions and Outlook

The six languages studied in this paper show substantial variation in *wannabe* constructions at several levels (formal properties, semantic properties and productivity). Overall, *wannabe* in both English and recipient languages is used either as an independent noun or adjective or in collocations, where it tends to acquire an approximative (often derogatory) meaning. However, whereas in English *wannabe* collocates almost exclusively with nominal collocates (nouns, NPs, proper names), in the recipient languages we found expansion to unexpected categories (adjectives, verbs, and pronouns). On the semantic level, English differs from the recipient languages as well: in the latter we find more examples of inanimate collocates, presumably because speakers of languages other than English do not associate *wannabe* with the verbal phrase (*want to be*) it derives from.

The large number of hapaxes in our data suggests that *wannabe* is very productive while at the same time we see new schemata emerging. This is evidenced by tiny mismatches in semantics of either the collocate or *wannabe* itself, but we also found category shifts from noun to adjective. These are all tendencies, and stronger in some languages than in others, but we do argue that they reflect different kinds of creativity: mismatches are (initially) F2-creative, but they may result in either F1-creativity (generalization of an existing schema), or E-creativity (the emergence of a new schema).

Naturally, changes from F2 to F1 or E are gradual and we base our analysis on synchronic data. More data to test our three-way model of creativity is needed, not least diachronic corpus studies. Another interesting angle for future research is intralinguistic competition with other (morphological) markers of approxima-

tion. In any event, emerging phenomena like *wannabe* are especially interesting to be "photographed" to unveil the possible paths of variation in both creativity and routinization processes.

References

The data that support the findings of this study are openly available in AMS Acta, the institutional open access repository of the Alma Mater Studiorum – University of Bologna, at https://doi.org/10.6092/unibo/amsacta/8063.

Ainiala, Terhi & Jan-Ola Östman. 2017. Introduction: Socio-onomastics and pragmatics. In Terhi Ainiala & Jan-Ola Östman (eds.), *Socio-onomastics: The pragmatics of names*, 1–18. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/pbns.275.

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, 899–919. Berlin & New York: De Gruyter Mouton.

Bauer, Laurie. 2019. Compounds. In Bas Aarts, Jill Bowie & Gergana Popova (eds.), *The Oxford handbook of English grammar*, 262–280. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198755104.013.3.

Bergs, Alexander. 2018. Learn the rules like a pro, so you can break them like an artist (Picasso): Linguistic aberrancy from a constructional perspective. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 277–293. https://doi.org/10.1515/zaa-2018-0025.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Booij, Geert. 2016. Construction Morphology. In Andrew Hippisley & Greg Stump (eds.), *The Cambridge handbook of morphology*, 424–448. Cambridge: Cambridge University Press.

Booij, Geert. 2018. The construction of words: Introduction and overview. In Geert Booij (ed.), *The construction of words: Advances in Construction Morphology*, 3–16. Dordrecht: Springer International Publishing. https://doi.org/10.1007/978-3-319-74394-3_1.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Cabin, Robert J. & Randall J. Mitchell. 2000. To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America* 81(3). 246–248.

Cappelle, Bert. 2014. Conventional combinations in pockets of productivity: English resultatives and Dutch ditransitives expressing excess. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the scope of Construction Grammar*, 251–282. Berlin & New York: De Gruyter Mouton. https://doi.org/10.1515/9783110366273.251.

Cappelle, Bert, Robert Daugs & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and 'disproximation.' *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 52–75.

Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108671040.

Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and intensifiers in Italian, German and other languages*. Berlin & New York: De Gruyter Mouton.

Eitelmann, Matthias, Kari E. Haugland & Dagmar Haumann. 2020. From *engl-isc* to *whatever-ish*: A corpus-based investigation of *-ish* derivation in the history of English. *English Language & Linguistics* 24(4). 801–831. https://doi.org/10.1017/S1360674319000340.

Eitelmann, Matthias & Dagmar Haumann. 2022a. *Extravagant morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology*. Amsterdam & Philadelphia: John Benjamins.

Eitelmann, Matthias & Dagmar Haumann. 2022b. Introduction: Extravagance in morphology. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology*, 1–18. Amsterdam, Philadelphia: John Benjamins.

Eitelmann, Matthias & Dagmar Haumann. 2023. *Getting close-ish*: A corpus-based exploration of *-ish* as a marker of approximation and vagueness: *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 76–100. https://doi.org/10.21248/zwjw.2023.1.91.

Gardani, Francesco. 2020. Borrowing matter and pattern in morphology. An overview. *Morphology* 30(4). 263–282. https://doi.org/10.1007/s11525-020-09371-5.

Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press. https://doi.org/10.2307/j.ctvc772nn.

Grandi, Nicola. 2017. I diminutivi come marche di attenuazione e indeterminatezza. In Oana-Dana Balaş, Adriana Ciama, Mihai Enăchescu & Anamaria Gebăilă (eds.), *L'expression de l'imprécision dans les langues romanes*, 162–175. Bucharest: Ars docendi – Universitatea din Bucureşti.

Grandi, Nicola & Lívia Körtvélyessy. 2015. Introduction: Why evaluative morphology? In Nicola Grandi & Lívia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: Edinburgh University Press.

Hoffmann, Thomas. 2020. Speakers are creative, within limits – a response to Peter Uhrig. *Cognitive Semiotics* 13(1). 1–7. https://doi.org/10.1515/cogsem-2020-2028.

Hüning, Matthias, Ulrike Vogl, Ton Van Der Wouden & Arie Verhagen (eds.). 2006. *Nederlands tussen Duits en Engels. Handelingen van de workshop op 30 september en 1 oktober 2005 aan de Freie Universität Berlin*. Leiden.

Jurafsky, Daniel. 1992. *An on-line computational model of human sentence interpretation: A theory of the representation and use of linguistic knowledge*. University of California, Berkeley.

Kemmer, Suzanne & Michael Barlow. 2000. Introduction: A usage-based conception of language. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, vii–xxiii. Stanford: CSLI Publications.

Kempf, Luise & Matthias Eitelmann. 2018. Von "diutisk" zu "dynamisch", von "englisc" zu "anything-ish". "-is(c)h" kontrastiv diachron. *Zeitschrift für Wortbildung / Journal of Word Formation* 2(1). 93–134. https://doi.org/10.3726/zwjw.2018.01.04.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1). 7–36. https://doi.org/10.1007/s40607-014-0009-9.

König, Ekkehard & Volker Gast. 2018. *Understanding English-German contrasts*. 4th edn. Berlin: Erich Schmidt Verlag.

Kosse, Maureen. 2021. "Don't make me pull a Britney": Onomastics and genericness in the PULL A [PROPER NAME] construction. *Colorado Research in Linguistics*. https://doi.org/10.33011/cril.v25i.1133.

Lamiroy, Béatrice. 2011. Degrés de grammaticalisation à travers les langues de même famille. *Mémoires de la Société de linguistique de Paris* 19. 167–192.

Masini, Francesca & Jenny Audring. 2019. Construction Morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 365–389. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199668984.013.25.

Masini, Francesca & Maria Silvia Micheli. 2020. The morphological expression of approximation: The emerging *simil-* construction in Italian. *Word Structure* 13(3). 371–402. https://doi.org/10.3366/word.2020.0176.

Masini, Francesca, Muriel Norde & Kristel Van Goethem. 2023. Approximation in morphology: A state of the art. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 1–26. https://doi.org/10.21248/zwjw.2023.1.100.

Micheli, Maria Silvia. 2023. The emergence of approximative values in the Italian prefixes *semi-* and *para-*. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 156–176. https://doi.org/10.21248/zwjw.2023.1.87.

Norde, Muriel & Graeme Trousdale. 2023. Issues in Diachronic Construction Morphology. *Constructions and Frames* 15(2). 145–159. https://doi.org/10.1075/cf.00070.nor.

Norde, Muriel & Graeme Trousdale. 2024. Creativity, paradigms and morphological constructions. Evidence from Dutch pseudoparticiples. *Linguistics*. https://doi.org/10.1515/ling-2023-0194.

Oltra-Massuet, Isabel. 2017. Towards a morphosyntactic analysis of *-ish*. *Word Structure* 10(1). 54–78. https://doi.org/10.3366/word.2017.0100.

Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, experiment and argument in linguistics and philosophy of language*, 15–26. Berlin: Peter Lang.

Uhrig, Peter. 2020. Creative intentions – The fine line between 'creative' and 'wrong.' *Cognitive Semiotics* 13(1). 1–19. https://doi.org/10.1515/cogsem-2020-2027.

Ungerer, Tobias & Stefan Hartmann. 2023. *Constructionist approaches: Past, present, future*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009308717.

Van Goethem, Kristel & Muriel Norde. 2020. Extravagant "fake" morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458. https://doi.org/10.1515/cllt-2020-0024.

Van Goethem, Kristel, Muriel Norde & Francesca Masini. 2025. The fate of 'pseudo-' words. A contrastive corpus-based analysis. *Languages in Contrast* 25(1). 23–50. https://doi.org/10.1075/lic.22003.van.

Van Haeringen, Coenraad Bernardus. 1956. *Nederlands tussen Duits en Engels*. Den Haag: Servire.

Vassiliadou, Hélène, Francine Gerhard-Krait, Georgia Fotiadou & Marie Lammert. 2023. Pseudo(-) in French and Greek: Categorization and approximation. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 234–262.

Zeldes, Amir. 2013. *Productivity in argument selection: From morphology to syntax*. Berlin & New York: De Gruyter Mouton. https://doi.org/10.1515/9783110303919.

Martin Schäfer

Scare Quotes as Markers of Creative Word-Formation: A Look at English *-ness* and *-ity* Hapaxes

Abstract: Creative modification involves the conscious manipulation of existing patterns in language use. This paper investigates whether consciously used new word formations fulfill this definition of creative modification. Focusing on novel English *-ness* and *-ity* derivations, conscious usage is operationalized as occurrences in scare quotes, and the patterns in the large synchronic ukWaC corpus are compared to known patterns for new *-ness* and *-ity* derivations. Quoted hapaxes do not differ from unquoted hapaxes, but both differ from the patterns found in neologisms. This is in line with the idea that on the way from a hapax to an accepted neologism adherence to standard patterns is of more importance than in ad-hoc coinages, be they conscious or unconscious. As for the role of scare quotes, a closer look at the contexts in which quoted and unquoted hapaxes with phrasal bases occur suggests that their usage is sensitive to the relative creativeness of a new form more than to its departure from general language standards.

1 Introduction

The starting point of this paper are the observations in Filatkina (2018) regarding linguistic patterns that have achieved a high degree of fixedness and conventionalism. According to Filatkina, these can either become subject to variation or to modification by the speakers of a language. In particular, she sees modification as occasional, and "as an irregular, intentional and conscious intervention of a speaker into the form and/or meaning of a pattern directed at the violation of the existing norms" (Filatkina, 2018: 26–27). I will follow Filatkina and refer to modifications thus defined as creative modifications. And while Filatkina developed her defining criteria on the basis of formulaic patterns, I will here strictly focus on word formation. The aim of this paper is to investigate whether consciously used new word formations fulfill this definition of creative modification. In particular, I investigate instances of English *-ness* and *-ity* derivations in scare quotes in a synchronic written corpus, cf. examples (1) and (2). Both examples are from the ukWaC, the corpus also used in the two studies presented in this paper.

(1)   *Flat-bottomed boats have an air of practical simplicity, and a feel of functional **"boatiness"** that contrasts sharply with modern styles.* [ukWaC]

(2)   *[. . .] there can be no adsorbed oxide layers or films of any sort ! This is where the **'plateability'** of metals becomes a factor . Very corrosion-resistant metals such as titanium [. . .]* [ukWaC]

-ness and *-ity* derivatives are chosen because they are both productive and well-described. In addition, since as a pair they constitute an instance of affix rivalry, it is also of interest whether scare quotes are used in different ways for these two affixes.

2 Background and Expectations

The aim of this paper is to investigate quoted hapaxes of -ness and -ity derivatives against Filatkina's (2018) definition of creative modification. Other approaches to creativity in language use are not considered here, and the reader is referred to the relevant literature (for example, Körtvélyessy, Stekauer and Kamár 2022 on creativity in word formation and Eitelmann and Haumann 2022 on extravagant morphology). In order to be able to assess the extent to which the quoted instances deviate from established patterns, the established patterns are introduced in the next two sections. Of particular importance to this study are not only the general patterns, but also the detailed descriptions of *-ity* and *-ness* neologisms in Arndt-Lappe (2014). Since these neologisms were extracted from the Oxford English Dictionary, they can be taken to represent the routinization of new forms and are therefore an ideal foil to assess creative modification. What makes looking at *-ity* and *-ness* so attractive is their similarity: prototypically, both turn adjectives into nouns. In addition, their prototypical semantic effect is very similar, if not identical, forming abstract nouns "with the meaning 'state, quality, condition of –'" (Marchand 1969, for *-ity* on page 312 and for *-ness* on page 334). The differences emerge in the details, with *-ness* being applicable across different parts of speech, and with both suffixes having clear preferences for specific forms of adjectival bases. In addition, on hapax-conditioned productivity measures, *-ness* is more productive than *-ity* (cf. Baayen and Renouf 1996; Plag 2006). Of course, not all hapaxes are new words, but new words typically come into existence as hapaxes. Therefore, I expect in general more new derivatives with *-ness* and will therefore always start with this suffix, using *-ity* mainly for verification and control. This section starts with short overviews of both affixes in sections 2.1

and 2.2, section 2.3 introduces Arndt-Lappe's neologism data, and section 2.4 states the expectations for the current study.

Before looking at *-ness/-ity* in detail, a short note on the status of scare quotes. It is clear that these are used to mark novelty in word-formation (Renouf and Bauer 2000, Kaunisto 2013). But this is not their only function, they also mark meta-linguistic usages or non-literal and/or ironic usages (see Wislicki 2023). These other usages are put aside for the purpose of this paper, with only meta-linguistic usages playing a role: these are used as exclusion criteria in the process of data preparation (see section 3.1).

2.1 *-ness* Fact Sheet

Prototypical occurrences of *-ness* forms are shown in (3) (examples are either directly from British English corpora or other media as indicated[1]):

(3)   a. *It's not* **happiness** *but sorrow that I'm looking for.* [BNC: A18 138]
      b. *If you lose* **consciousness**, *even for a second, then you have suffered brain damage and must withdraw from further competition.* [BNC: A0M 1368]

These are prototypical instances, because the base of the derivatives are the adjectives *happy* and *conscious*, and their meaning is paraphraseable by replacing the forms with *the state of being happy* and *the state of being conscious*.

One and the same form can develop other usages than those resulting from prototypical derivation, cf. (4).

(4)   *His* **consciousness** *delivers paranoid images of aggression and hostility.* [BNC: A05 1609]

Here, *consciousness* refers not so much to a specific state, but to the mind. Established and high frequency forms (cf. also *business* and *illness*) seem especially prone to developing such (additional) meanings. While the usage of an estab-

---

1 Examples from the British National Corpus (BNC) show the BNC-identifier, ukWaC examples come without identifier as the corpus has no sentence-level identifiers. The BNC was accessed via the web- interface provided by Lancaster University: the CQP-edition (Version 4.4) of BNCweb developed by Sebastian Hoffmann and Stefan Evert (cf. Hoffmann et al. 2008; BNCweb is accessible via http: //bncweb.lancs.ac.uk/).

lished form with a new meaning can also qualify as creative modification according to Filatkina's characterization, these are not considered in this study because the focus is on new forms, i.e. new products of word-formation.

The affix *-ness* is not restricted to adjectival bases. Only verbs and bound bases are excluded (Bauer, Lieber, and Plag 2013: 246). Examples for *-ness* on nouns, on phrases, and on prepositions are given in (5)–(7).

(5) *Or has, rather, **animalness** developed through limitation out of humanness that was maturing toward its universality?* [ukWaC]

(6) *I saw them at Hammersmith Odeon in 1978 and my mate and I went to the stage door and met Geddy – he was lovely, talkative and had the long hair we wished we could have. Think he's never lost that **good-bloke-ness**.*
[Guardian; https://www.theguardian.com/music/2018/dec/24/geddy-lee-on-rush-greatest-songs#comment-124085787, accessed 2018/12/25 00:38]

(7) *Principle of **aboveness**: if one thing is above the other, it's perceived as better.* [ukWaC]

For nouns and phrases, Bauer, Lieber, and Plag (2013: 261) note that "*-ness* does not so much denote the state or condition of being the kind of entity denoted by the noun; rather, it highlights or picks out the significant characteristics that make the entity what it is, denoting the abstract quality or state of those characteristics." Accordingly, *animalness* is paraphraseable by "the significant characteristics that make an animal an animal", and *good-bloke-ness* by "the significant characteristics that make one a good bloke".

The exact wording of prototypical meanings differs slightly across accounts. While Marchand has *state, quality, condition of –,* Bauer, Lieber, and Plag (2013) just use *state or condition of.* Arndt-Lappe (2014: 508) operationalizes semantic transparency of *-ness* forms by looking for the signal words 'property', 'quality', 'state' or 'condition' in their definitions. It is not entirely clear whether these different possibilities to paraphrase correspond to different meanings or are meant to correspond to different meanings or rather provide just alternative formulations for the same basic standard meaning. In contrast, Aronoff (1976: 38), in discussing X-ous-ness derivatives (*callous/callousness*), points out that they have exactly three different paraphrases, exemplified with his examples in (8).

(8)   a.   reading 1: the fact that Y is Xous
        His callousness surprised me. = The fact that he was callous surprised me.
     b.   reading 2: the extent to which Y is Xous
        His callousness surprised me. = The extent to which he was callous surprised me.
     c.   reading 3: the quality or state of being Xous
        Callousness is not a virtue. = The quality or state of being callous is not a virtue.
     Cf. Aronoff (1976: 38)

Aronoff (1976: 38, fn5) leaves it open whether these three paraphrases present separate readings or one "tripartite or ambiguous one". Importantly, these readings can be related to one another, and all three appear to be standard interpretations. And as the usage of the same example sentence for the first two readings shows, they are not exclusive in the sense that one context necessarily leads to just one reading. These three paraphrases are not restricted to X-ous derivations, and again, sometimes it is hard to decide which paraphrase fits best, or whether just one of the meanings is intended, cf. *physicalness* in (9).

(9)   *And I remember ski places, those overheated rooms and the books that people leave behind them and the galvanic excitement of* **physicalness***.* [John Cheever: Falconer (1977), p. 79 of the First vintage international edition, october 1991]

In (9), all three different paraphrases seem possible: *the galvanic excitement of the fact that it is physical/of the extent to which it is physical/of the quality or state of being physical.* Ultimately, the context determines the exact readings. Nouns and phrases can sometimes also have state readings, cf. (10).

(10)   *"That's a privilege of* **old manness***"* [Comedians in cars getting coffee, Season 1, Episode 9 I Want Sandwiches, I Want Chicken, at 4:30]

Overall, these differences highlight the general difficulty of giving one fitting overall meaning description, which, in turn, is also related to the same form allowing closely related usages. At the same time, all of these different paraphrases are examples of standard, prototypical readings for *-ness* derivations.

     Morpheme-level and other form-level properties There are little or no absolute restrictions on possible bases for *-ness*, but there are, even within adjec-

tives, its prototypical base category, clear tendencies and preferences for specific patterns at the morpheme level and below. These will be discussed together with the corresponding properties of *-ity* in section 2.2.

2.2 *-ity* Fact Sheet

Prototypical occurrences of *-ity* forms are shown in (11):

(11)   a. *Perhaps in a foul world these men were seeking* **purity**. [BNC: A8F 800]
       b. *It is difficult for a woman to understand a man's sensitivity to any slur on his* **virility**. [BNC: ACS 853]

The bases are adjectives, and paraphrase with *state/condition of* are possible: *Perhaps in a foul world these men were seeking a state of being pure/It is difficult for a woman to understand a man's condition of being sensitive to any slur on his state of being virile*. In contrast to *-ness*, *-ity* is much more restricted, with adjectival bases often taken to be the only free morphemes it combines with (Marchand 1969). Further, more *-ity* forms than *-ness* forms are high frequency lexemes and have idiosyncratic, lexicalized meanings (Bauer, Lieber, and Plag 2013: 257).

Morpheme-level and other form-level properties Early discussion of the morpheme-level and form-level properties of the bases pointed to an aversion of *-ity* to native bases (Marchand 1969: 314) or even claimed that it is restricted to just Latinate bases (Aronoff 1976: 51). Lindsay (2012), investigating google hits for 3256 potential rival pairs, arrives at a more differentiated picture when looking more closely at potential bases in terms of their endings: while *-ness* is overall more productive (in terms of its distribution across bases), *-ity* dominates in some subdomains. For example, bases ending in *-ing, -ish* and *-ful* occur only with *-ness* in Lindsay's data, but both suffixes occur to a considerable extent with bases ending in *-ous/-os* and *-ive,* and *-ity* is dominant for bases ending in *-able, -al, -ic*, and *-ar* (see also Anshen and Aronoff 1981 for experimental support of an *-ity* preference for *-able/-ible* bases). In other words, for some bases, *-ity* is preferred even if they are not strictly speaking Latinate as long as they are formed with a Latinate affix, like in the case of *drinkable/drinkability*, with its Germanic root *drink*.

Lindsay (2012) discusses this distributional difference in terms of a morphological constraint on *-ity,* but Arndt-Lappe (2014) points out that previous studies do not allow for a distinction between a preference for either suffix based on the morphological makeup of the base, or just the form of the base. To take just one

example, is the decisive feature for the status of the adjective *affective* as a potential base for *-ity* or *-ness* just its form, that is, its ending on the string *-ive*, or that it consists of a base and the morpheme *-ive*? Arndt-Lappe here points to the derivative *perspectivity* in her own dataset, since *-ive* in the base *perspective* does not have morphemic status. As this example already illustrates, it is rather difficult to find examples that would allow one to clearly distinguish between morpheme-based and form-based approaches.

2.3 Patterns in *-ness* and *-ity* Neologisms

The study by Arndt-Lappe (2014) on the *-ity/-ness* suffix rivalry provides overviews of the characteristics of *-ity/-ness* bases for twentieth-century neologisms in the Oxford English Dictionary (OED). This is a particular valuable resource, because it presents patterns that show up in new forms that made it into the language. I hypothesize that their patterns are more conservative and in line with prototypical word-formation processes than the typical occasionalism that is formed creatively. Of course, the adoption of neologisms to the pool of established derivatives also changes that pool itself, but the idea is that this change is a very gradual one, precisely because these neologisms tend to be not consciously-formed, creative formations, but result from a routine, plausibly sub-conscious, use of these affixes. In Table 1, I reproduce her Table 3, giving an overview of the syntactic category of the bases in the 344 *-ity* and 220 *-ness* derivatives in her set of twentieth-century neologisms. The category 'minor category' subsumes adverbs, prepositions and pronouns.

Table 1: Base categories of *-ity* and *-ness* derivatives in OED twentieth century neologisms, cf. Table 3 in Arndt-Lappe (2014).

syntactic category	*-ity* derivatives (n=344)		*-ness* derivatives (n=220)	
	n	%	n	%
Adjective	326	94.8	186	84.5
Noun	7	2	14	6.4
bound form	11	3.2	0	0
Phrase	0	0	10	4.5
minor categories	0	0	10	4.5

Table 1 very clearly shows that *-ness* takes a wider range of bases, including not only nouns and adjectives but also phrases and minor categories like adverbs,

prepositions, and pronouns. Of the two most frequent free bases, adjectives dominate for both suffixes, but much more clearly so for *-ity* which only shows 2% of nominal bases while -ness has 6.4% nominal bases.

Of the 512 adjective bases, 42 are without a suffix, while those 445 whose suffixes appear at least ten times distribute across the two suffixes as shown in Table 2. This distribution of bases and morphological classes within the adjective bases gives us a baseline to compare the quoted hapaxes against.

Table 2: Morphological categories of *-ity* and *-ness* adjectival bases in OED twentieth century neologisms, cf. Table 4 and Figure 1 in Arndt-Lappe (2014).

morph class	total	*-ity*	*-ness*
-ar	12	100%	
-ic	55	>95%	<5%
-able	109	>95%	<5%
-al	78	>95%	<5%
-ive	51	>70%	<30%
-ous	24	>60%	<40%
-ed	29		100%
-ing	12		100%
-ish	10		100%
-less	10		100%
-y	55		100%

2.4 Expectations and Hypotheses

The general expectation is that the quoted hapaxes are markedly different from the neologisms, either in their preferences for specific types of bases, or in their semantics. This expectation rests on the assumption that neologisms that are accepted into the language are typically not the product of creative, conscious word formation, but of more routine-usage of existing affixes in accordance with how they are regularly used. For the bases, this difference can emerge on the level of part of speech preferences, or, for adjectival bases, on the level of preferences for specific form patterns.

3 Study 1: Quoted *-ness* Hapaxes in the ukWaC

3.1 Materials and Techniques

All occurrences of *-ness* forms were extracted from the ukWaC, a web-derived 2 billion word corpus of English (Baroni et al. 2009; available via https://wacky. sslmit.unibo.it/). This corpus was chosen because it is, to my knowledge, one of the few freely available and annotated large corpora. Of the material made available, I used the full corpus without annotation, the version with part of speech annotation, and the unigram lemmata frequency list. For all further steps, R and Python scripts were used. These scripts and the final dataset of hits with their contexts is available at https://doi.org/10.6084/m9.figshare.25549474.v1.

To identify the hapaxes of interest, I proceeded as follows:

A. I created a list of all *-ness* derivatives and their bases in the ukWaC. The list of ukWaC unigram lemmata and a second, POS-tagged lemmata list derived from the tagged ukWaC was used to identify all derived *-ness* forms and add their base lemmata. In the process of this, all *-ness* forms that did not have a base in the ukWaC were excluded. Further, forms that contained obvious artifacts (for example, starting with a hyphen) or proper names like *McGuinness* were excluded. In addition, lexicalized high-frequency *-ness* items like *witness* and *business* that often form the head of more complex words were used to exclude such more complex forms (e.g. *expert-witness* or *charity-business*). This yielded a list of 8785 *-ness* lemmata. This list contains overall 3758 hapaxes.

B. Python scripts were used to extract all occurrences of quoted *-ness* forms from the tagged corpus. Any occurrence of a form ending in *-ness* and immediately preceded and followed by either single or double quotes was collected.

C. The lists resulting from steps A and B were cross-checked so that only those items that are true quoted hapaxes remained, that is, lemmata that occurred only once, and that one time in scare quotes. For items in the quoted *-ness* hapaxes for which step A not already provided a POS-tagged base, a base and its POS tag were added manually.

D. The identified target words were used to extract the contexts for the words from the tagged ukWaC corpus for further inspection and annotation of the targets.

E. All quoted hapaxes were checked in their context. In the process of annotating the data, quoted forms were excluded whenever the quotes were used to indicate metalinguistic usages, as in (12), where the quotes indicate different translation possibilities for a specific word.

(12) *There are many reminders of Germany's new Unbefangenheit – a word that hovers untranslatably somewhere between **"unencumberedness", "relaxedness", and "unbotheredness".** [ukWaC]*

Further, misspellings (*agressiveness* instead of *aggressiveness*) and spelling variants (*clueless-ness* besides *cluelessness*) were excluded, as well as one occurrence in the quoted hapaxes of the unrelated *-ness* suffix indicating female gender (*Solomoness*).

As a result, 214 hapaxes in scare quotes were identified, either in single or in double quotes, cf. (13).

(13) a. *Judicious hedge grubbing can open up cast-off fields and create more regular patterns without destroying the overall **'hediness'** of the farm.* [ukWaC]

b. *The central theme of the day is to explore the value of flagships, the concept of **"flagshipness"** and its implications for industry, design and future research.* [ukWaC]

3.2 Results [Study 1]

The distribution of the base POS, against the numbers from Arndt-Lappe (2014, Table 3), is shown in Table 3 (see Table 9 in the appendix for the full list of hapaxes). The coding here reproduces the categories used by Arndt-Lappe, but note that in contrast to her data, the minor category is mostly (4 out of 7) made up of abbreviations and acronyms, cf. (14) for two examples.

Table 3: Base categories of *-ness* derivatives in the ukWaC quoted hapaxes and in Arndt-Lappe's OED 20th century neologisms.

	ukWaC scare quote hapaxes (n=214)		OED neologisms
syntactic category	n	percentage	percentage
adjective	139	65	84.5
noun	59	27.6	6.4
phrase	9	4.2	4.5
minor category	7	3.3	4.5

(14) *And no matter how many court rulings go against MS, the **"IEness"** is so en-trenched into XP that it's difficult for the average Joe User to escape.* (IE = Internet Explorer) [ukWaC]

(15) *. . . certain operations and types of code, while RISC chips have grown in size and complexity while retaining a core **"RISC-ness".*** (RISC = Reduced Instruction Set Computer) [ukWaC]

While the percentage of phrases and minor categories is similar, there is a clear difference in the two dominant categories of bases, nouns and adjectives. This difference is significant (Pearson's Chi-squared test, X-squared = 33.022, df = 1, p-value = 9.112e-09).

For any patterns in the bases themselves, Table 4 shows the distribution of endings of the adjectival bases in the data corresponding in form to the suffixes reported by Arndt-Lappe, with the addition of *-ful*.

Table 4: Endings corresponding to morphological suffixes in the set of adjectival bases in the ukWaC *-ness* derivatives, ordered by frequency.

ending	frequency	percent
-y	39	28.06
other	33	23.74
-ed	31	22.3
-ish	7	5.04
-al	6	4.32
-ing	6	4.32
-less	5	3.6
-ive	4	2.88
-ic	3	2.16
-ful	2	1.44
-ous	2	1.44
-ble	1	0.72

If we just consider the three most frequent patterns, it is noticable that: a) Adjective bases without an ending corresponding to a noticable morphological pattern constitute 24% of the bases. b) Overall, the ranking is similar to the one in Arndt-Lappe's dataset, with the same order of the two most frequent endings, *-y* followed by *-ed*, dominating the distribution.

As far as the meanings of the new forms are concerned, it was impossible to discern any special patterns. Since definitions of the meanings are not indepen-

dently available, the only viable technique was to explore whether paraphrasing with any of the paraphrases discussed in section 2.1 was possible or not. This was always possible. Especially, it was always possible to use one of Arndt-Lappe's signal words for semantic transparency in the paraphrases. For example, *hedginess* and *flagshipness* in example (13) can be paraphrased with *the quality of being hedgy* or *the state of being a flagship*.

3.3 Discussion [Study 1]

The data so far suggests that the main difference between creative word formation in the sense of this paper and standard word formation, for which we used the OED neologism data, lies in the relative preference of creative formations for an unusual base, the noun. But there might be other factors at play that pave the way for a form to become an accepted neologism, and it therefore seems prudent to check first whether the unquoted hapaxes in the ukWaC pattern with the OED neologisms, or with the ukWaC hapaxes, or show yet another pattern. Table 5 shows the distribution across the non-quoted ukWaC hapaxes, against the distributions of OED neologisms and quoted hapaxes from Table 3.

Table 5: Base categories of *-ness* derivatives (in percent).

syntactic category	OED 20th century neologisms (n=220)	ukWaC scare quote hapaxes (n=214)	ukWaC hapaxes unquoted (n = 3544)
adjective	84.5	65	58.9
noun	6.4	27.6	34.5
phrase	4.5	4.2	2.1
minor category	4.5	3.3	4.5

Note that the difference in the percentage of phrases between the quoted and the unquoted hapaxes must be taken with a grain of salt. As explained in section 3.1, the quoted hapaxes were manually checked and the amount of phrases therefore maximized, similar in quality to the OED data. In contrast, the unquoted hapaxes where only checked against existing hyphenated or concatenated forms in the corpus, which likely leads to some misrepresentation of specifically this category, because the phrases are hyphenated or possibly concatenated in the derivatives but occur spaced outside of complex word forms. The differences across the two ukWaC datasets in the two main categories are small and not statistically significant (X-squared = 3.7643, df = 1, p-value = 0.05236), whereas the difference be-

tween the main categories in the unquoted hapaxes and the OED is again significant (X-squared = 72.617, df = 1, p-value < 2.2e-16).

What about the endings of the adjectival bases? In the unquoted hapaxes, there is again the category "other" in second place, with 21.26%. And the two dominant categories again are in the same order, with *-y* with 22.2% before *-ed* with 21.26%, cf. Table 6.

Table 6: Two top endings across the three *-ness* subsets with adjectival bases.

ending	OED neologisms	quoted hapaxes	unquoted hapaxes
-y	55	39	464
-ed	29	31	423

For the two most frequent endings, the difference between the two ukWaC subsets is not significant (X-squared = 0.30135, df = 1, p-value = 0.583), and the difference between the quoted ukWaC subset and the OED set is also not significant (X-squared = 1.53, df = 1, p-value = 0.2161). Only the difference between the OED set and the unquoted ukWaC hapaxes is significant (X-squared = 5.3452, df = 1, p-value = 0.02078). Note, however, that during manual correction of the coding for the ukWAC hapaxes it was noticeable that many bases in *-y* were automatically tagged as nouns. Since the tagging of the full set of hapaxes was not manually corrected, it is likely that the proportion of *-y* adjectives is actually higher, so I do not think this difference means much.

So far then, against the original hypothesis, the comparison of quoted and unquoted hapaxes in the ukWaC corpus seems to show that there is nothing special about the quoted hapaxes. That does not necessarily mean that there are no differences between the two datasets. Two issues that complicate the picture are the use of quotes for morphological explicitness and the role of the surrounding text. We will look at both in turn.

3.3.1 Morphological Explicitness: Avoid Doublets/Ambiguity

In a few cases, the usage of quotes occurs to make clear that other readings than existing ones are intended, without these readings themselves occurring to be particularly special. Two such examples are *bass-ness*, one of the quoted hapaxes, and *pointed-ness*, which occurs quoted but is not a hapax, cf. (16) and (17). In both examples, a hyphen is used in addition. It seems that both means, the quotes and the hyphen work together here: the hyphen indicates that it is a reading that is

different from another possible reading (cf. also Kaunisto 2013 on the usage of hyphens as indicators of special characteristics), and the quotes indicate that this makes it something special and it is not just a typo or spelling-variant.

(16) *[It's] not as aggressive as the Snarling Dogs nor as wide as the Morley but it does help retain the **"bass-ness"** of your bass and it does a good job on those "Pulling Teeth"-style solos* [ukWaC]

(17) *However, one diviner did not use this interpretation. He disregarded the **"pointed-ness"** of the cards, concentrating instead on whether the cards had been turned over;* [ukWaC]

Bass-ness can be contrasted with *bassness*, which does not occur in the ukWaC, but for which we find the dictionary definition *the quality of being low-pitched* (Collins English Dictionary[2]). That is, the base here is the relational adjective *bass*, whereas for *bass-ness*, it is the noun denoting the musical instrument, giving rise to the reading *quality of being a prototypical bass*. Similarly, *pointed-ness* contrasts with the established *pointedness* as *quality of having an end that comes to a point*, cf. (18).

(18) *"Following a brief conversation, during which Hyde-Wollastone exhibited a number of interesting, though seemingly harmless fascinations, notably with toys and origami, the **pointedness** of his remarks concerning his financial influence led this agent to run a standard-Procedure exhaustive background check on him."* [ukWaC]

In contrast, *pointed-ness* in (17) can be paraphrased as *the extent to which the cards where pointed*. While this reading uses a prototypical pattern, it clearly contrasts with the lexicalized standard reading of *pointedness*.

3.3.2 Phrasal Bases and the Role of Context

Yet a different factor emerges if we compare the nine phrasal quoted forms, cf. (19) with unquoted phrasal forms.

---

2 https://www.collinsdictionary.com/dictionary/english/bassness; last accessed 2023-09–14.

(19) *broad-mindness, like-thisness, sugary-sweetness, to-be-heard-of-ness, to-be-looked-at-ness, too-much-ness, up-side-downness, when-whereness, women-only-ness*
[all quoted phrasal *-ness* forms]

The nine formations show a variety of different phrasal constructions, so a first question is whether the set of unquoted phrasal forms is perhaps more limited in phrasal patterns. This does not seem to be the case. There are comparative phrasal bases of different kinds, cf. (20-a), and also patterns not shown in the nine quoted examples, for example verb-object, verb-particle, and preposition-pronoun bases, cf. (20-b). Beside the constructional variety, the phrasal bases in the unquoted set also show more colloquial elements like the constructions with *fuck-*.

(20) a. *too-coolness, too-hotness, cool-as-fuckness, cooler-than-thouness*
      b. *do-it-yourselfness, do-nothingness, fuck-off-ness, fuck-you-ness, up-itself-ness*
          [selection of unquoted phrasal *-ness* forms from the ukWaC]

If we zoom in on the contexts of some examples, we note an interesting difference. Consider the *too*-X-*ness* forms first, with the quoted example in (21), and the unquoted examples in (22).

(21) *He was particularly aware of how much of a break-through it was and large-minded enough to be quite satisfied with the result. Some Rindviecher (block-heads), as he called them, had clearly drawn "the wrong conclusions in their five-finger exercises on the pulse of British public opinion". Evidently, there was the usual "**too-much-ness**" of German eagerness and too little knowledge of what the British are really like.* [ukWaC]

(22) a. *Do you believe in rock n' roll, can music save your mortal soul? Sure, but as Three One G gives it, it can also pulverize you, bear hug your guts into mush, and take your choirboy virginity with no promise of respect or love come morning. But that's all hype and fan-drooling and gushing and who needs that? Facts: Three One G operates completely removed from the flaky, style-over-substance, **too-coolness** laid down by a lot of their peers' labels.* [ukWaC]
      b. *I was outside earlier and it was so hot that someone quite literally burst into flames before my very eyes. Well sort of. All right, not at all. They did sweat a bit, though, and go quite, quite puce, so I was worried. There's nothing to do but sit on the sofa and enjoy the **too-hotness**.* [ukWaC]

An obvious difference is the general style of the context. For the quoted form, the language and style in the immediate contexts is a straightforward, no-nonsense prose style in which the *too-X-ness* form looks and feels out of place. Since, for want of a better word, it is used anyways, the quotes are used too indicate that stylistically this form does not fit. In contrast, the style of the two unquoted *too-X-ness* passages is almost flamboyant, mixing colloquial forms, unusual grammar and inventive language. The *too-X-ness* do not occur out of place in these contexts.

This role of the local context is also apparent if we look at the *fuck-X-ness* phrases, cf. (23) and (24).

(23) *Man, there are so many ideas packed into this 8 track debut EP coming fromThe Bumblebees, a band whose members span continents and whose sound spans genres with such **fuck-you-ness** that the end result works like a dream. Sounding like early Beastie Boys/Beck/Missy Elliot/Timberland/The Hives/NERD and lord knows what else, White Printz is the culmination of a month long one mic/one take recording frenzy that saw each member of the 'bees collective stepping up and bringing to the mix their own individual take on the music in their heads.* [ukWaC]

(24) *And by day, grey against the grey that passes for sky in Cambridge, the cranes are like cranes. Bird-like, long-legged, flapping in the wind. [. . .] The cranes are so improbably close to each other that they interact like some giant sculpture. [. . .] So they're not there as useful engineering tools, but as art. From afar, the cranes seem gracefully flimsy, until you cycle underneath them, when their hulking **fuck-off-ness** gives them the awesome majesty of a super tanker mowing down a dinghy in a shipping lane.* [ukWaC]

Both are embedded in very long sentences, and while the first one is more colloquial in style and contains also other unusual forms, like the forward-slash use in the next sentence, the second one prepares us for the unusual form more via its unusual, wildly creative imaginative prose. As one reviewer remarks, one could even question whether in such a context it is still a case of creative modification in Filatkina's sense, because the existing norms here could be seen as having already been shifted.

Across these examples, the common theme is that quotes are unnecessary when the surrounding text is already clearly creative and therefore full of conscious new combinations. In a way, the creativity of one specific form then becomes more a question of relative creativity in relation to the surrounding context than of creativity against the general standard norms of the language.

3.3.3 Other Factors

In at least one case, the quotes could plausibly simply indicate that attributing a property based on origin and/or ethnicity is considered as not politically correct. This example is *Fulani-ness* in (25)

(25) *that most strikingly sets the Fulani apart from other peoples. It is the ethic or soul of* **"Fulani-ness"**, *and involves complex rules of interaction within Fulani society.* [ukWaC]

This is plausible because we find many instances of common formations such as *Britishness* in quotes, and often the function of the quotes seems to be to distance oneself from this concept, cf. (26) for a very clear example.

(26) *I don't really know where to start with Gordon Brown's half-baked plan to celebrate* **"Britishness"**. *For a start what does Brown mean by* **"Britishness"**? *Does such a thing even exist?* [ukWaC]

All in all, there are therefore a number of suggestive differences between the quoted and unquoted hapaxes despite their overall similarity in their preferences for the part of speech of their bases. Whether these differences still obtain when looking at all unquoted hapaxes in detail must be left open here.

4 Study 2: Quoted *-ity* Hapaxes in the ukWaC

4.1 Materials and Techniques

Data extraction and annotation proceeded parallel to study 1, except that this time, I looked for *-ity* forms. Again, scripts and results are available in the repository accompanying this paper (https://doi.org/10.6084/m9.figshare.25549474.v1). Overall there were 7262 *-ity* lemmata, with 2634 hapaxes, of which 2591 were unquoted, and 43 quoted. In a marked contrast to the *-ness* forms, the quoted hapaxes contained a higher number of new *-ity* forms that were not the result of derivation but of other processes, notably compound formation and blends, cf. the examples in (27) and (28).

(27) *cryptoidentity, cybercreativity, dot-density, launch-on-warning-capability, well-waterquality*

(28) a. *coppertunity, honourtunity, operatunity; ecotricity*
      b. *horsepitality, onfinity*

All examples in (27) are best analyzed as compound formations, with the right-most element being an *-ity* noun, all of which are well-established (*identity, creativity, density, capability, quality*). Similarly, all examples in (28) are best analyzed as blends, from the more conventional formations in (28-a) to those in (28-b).

Since only new derivations are of interest in this paper, these forms were excluded, leaving only 44 quoted *-ity* derivatives.

4.2 Results [Study 2]

Table 7 shows the distribution of the parts of speech of the *-ity* bases in percent, again in contrast to the OED neologisms and with the numbers for the bases of the unquoted hapaxes in the ukWaC corpus already added (see Table 10 in the appendix for the full list of hapaxes).

Table 7: Base categories of *-ity* derivatives (in percent).

syntactic category	OED 20th century neologisms (n=344)	ukWaC scare quote hapaxes (n=43)	ukWaC hapaxes unquoted (n = 2591)
adjective	94.8	74.4	68.4
noun	2	25.6	28.7
bound form	3.2	0	0
phrase	0	0	0
minor category	0	0	2.86

Note that while the quoted hapaxes did not show any bound bases, the annotation routine for the unquoted ukWaC hapaxes was not able to capture bound forms, as, in parallel to study 1, derivatives were only kept if the base occured as a free form by itself in the corpus. The most notable result is that the quoted occurrences descriptively fall in-between the OED neologism pattern and the unquoted hapaxes pattern. If we only consider the two major categories of bases, the differences between the OED data and both ukWaC subsets are significant (OED vs. quoted: X-squared = 46.059, df = 1, p-value = 1.147e-11; OED vs. unquoted: X-squared = 114.24, df = 1, p-value < 2.2e-16). In contrast, the difference between the two ukWaC subsets is not significant (X-squared = 0.32167, df = 1, p-value = 0.5706).

Table 8 shows the distribution of endings across the 32 adjectival bases in the dataset.

Table 8: Endings corresponding to morphological suffixes in the ukWaC -ity quoted hapaxes, ordered by frequency.

ending	frequency	percentage
-ble	18	56.2
-al	7	21.9
-ic	3	9.4
other	2	6.2
-ive	1	3.1
-ous	1	3.1

The numbers are small here, with the two endings *-ble* and *-al* with 56.2% and 21.9% already accounting for 75% of the adjectival bases. Descriptively, the *-ble* and *-al* endings, in that order, are also the most frequent endings in the OED neologisms from Arndt-Lappe's study, where we find overall 109 *-able* endings and 78 *-al* endings, with both endings almost exclusively (>95%) combining with *-ity* (see Table 2 in section 2 of this paper). In the unquoted *-ity* hapaxes, these two endings are also the most frequent, with 722 *-ble* bases and 379 *-al* bases within the adjectival bases. These two similarly dominate the distribution, with 40.6% and 21.3%, with the next most frequent ending, *-ic*, at 12.05% (213 bases). While I do not have the exact numbers for the OED data (they are not reported in Arndt-Lappe 2014), the difference between the quoted and unquoted hapaxes for these two top categories is not significant (X-squared = 0.44424, df = 1, p-value = 0.5051).

4.3 Discussion [Study 2]

The number of quoted hapaxes that are *-ity* derivatives is relatively small. Even so, study 2 shows overall the same pattern as study 1: as far as the part of speech of the bases is concerned, the ukWaC quoted and unquoted datasets, though descriptively less similar than the *-ness* datasets, pattern together against the OED neologisms. One other result that is only peripherally related to the general finding is that of the 43 *-ity* derivatives 26% have noun-bases. This is rather unexpected given that Marchand (1969) claims *-ity* takes only adjectival bases, and in Bauer, Lieber, and Plag (2013, p. 247), it is claimed that *-ity* occurs "only infrequently on nominal bases".

5 Summary and Conclusion

The main aim of this article was to investigate whether consciously formed new word formations fulfill the definition of creative modification "as an irregular, intentional and conscious intervention of a speaker into the form and/or meaning of a pattern directed at the violation of the existing norms" (Filatkina 2018). Using quoted hapaxes as operationalization of consciously formed new forms, this is clearly not the case: When it comes to the distribution of the bases in terms of part of speech and the endings of the adjectival bases, there is no notable difference between quoted and unquoted hapaxes in the huge ukWaC corpus. For both suffixes, the unquoted and quoted hapaxes in the ukWaC corpus pattern together and contrast especially with the part of speech distribution across the bases in the OED. This is in line with the idea that on the way from a hapax to an accepted neologism adherence to standard, routine patterns is of much more importance than in ad-hoc coinages, be they conscious or unconscious. When zooming in on phrasal bases in the *-ness* ukWaC hapaxes, some other reasons for the usage of quotes emerged. When the context itself is standard prose, quotes are used to signal non-standard usages. However, when the linguistic context already is non-standard, no quotes are necessary since an unusual new formation is not unexpected in such a context. That is, it is not so much creativity in an absolute sense but relative creativity that is reflected in the use of quotes. In other cases, quotes seem to be used simply as explicit pointers to a non-standard internal structure in order to avoid ambiguity, or to indicate some distance to the appropriateness of the word-formation product as such.

The aim of this study was modest in focusing on one specific concept of creative modification, and investigating this concept through one of the functions, marking novelty, of one single formal marker, the scare quotes. Further avenues for research and open questions not only concern the interpretation of the data against other concepts of creativity and routine in word formation, but also a closer look of the interplay of scare quotes with other means of marking novelty and with other usages of scare quotes themselves.

6 Appendix

Table 9: Quoted -ness hapaxes in the ukWaC.

nessLemma	basePOS	baseLemma
ieness	abbreviation	ie
mmorpgness	abbreviation	mmorpg
risc-ness	abbreviation	risc
wodkness	abbreviation	wodk
adverbialness	ADJ	adverbial
altness	ADJ	alt
apparentness	ADJ	apparent
astoundedness	ADJ	astounded
audit-mindedness	ADJ	audit-minded
banklessness	ADJ	bankless
belarusianness	ADJ	belarusian
blaséness	ADJ	blasé
blinkerdness	ADJ	blinkerd
blue-whiteness	ADJ	blue-white
boatiness	ADJ	boaty
bohemian-ness	ADJ	bohemian
botheredness	ADJ	bothered
browiness	ADJ	browy
caring-ness	ADJ	caring
cheekyness	ADJ	cheeky
civilianness	ADJ	civilian
clanginess	ADJ	clangy
clawedness	ADJ	clawed
coensiness	ADJ	coensy
commensurateness	ADJ	commensurate
computer-friendliness	ADJ	computer-friendly
curriculum-embeddedness	ADJ	curriculum-embedded
cuspiness	ADJ	cuspy
customer-consciousness	ADJ	customer-conscious
differencelessness	ADJ	differenceless
disproportionateness	ADJ	disproportionate
dopyness	ADJ	dopy
drawlingness	ADJ	drawling
earnedness	ADJ	earned
earth-boundedness	ADJ	earth-bounded
earth-friendliness	ADJ	earth-friendly
elementalness	ADJ	elemental
emotionalness	ADJ	emotional
enculturedness	ADJ	encultured
entrepreneur-readiness	ADJ	entrepreneur-ready
establishedness	ADJ	established

Table 9 (continued)

nessLemma	basePOS	baseLemma
existlessness	ADJ	existless
family-centredness	ADJ	family-centred
far-out-ness	ADJ	far-out
fulani-ness	ADJ	fulani
gaelicness	ADJ	gaelic
gappiness	ADJ	gappy
goddishness	ADJ	goddish
handmadeness	ADJ	handmade
hedginess	ADJ	hedgy
high-level-ness	ADJ	high-level
holier-than-thouishness	ADJ	holier-than-thou
homemadeness	ADJ	homemade
honours-worthiness	ADJ	honours-worthy
hoofedness	ADJ	hoofed
housewifeliness	ADJ	housewifely
hybridness	ADJ	hybrid
insidedness	ADJ	insided
institution-wariness	ADJ	institution-wary
internationalness	ADJ	international
jogginess	ADJ	joggy
joined-up-edness	ADJ	joined-up-ed
kentishness	ADJ	kentish
literalmindedness	ADJ	literalminded
malaysianness	ADJ	malaysian
mappy-ness	ADJ	mappy
mayanness	ADJ	mayan
middlingness	ADJ	middling
mildheartedness	ADJ	mildhearted
mingingness	ADJ	minging
minoan-ness	ADJ	minoan
moddiness	ADJ	moddy
moreishness	ADJ	moreish
n-connectedness	ADJ	n-connected
newbie-friendliness	ADJ	newbie-friendly
non-expansiveness	ADJ	non-expansive
non-naturalness	ADJ	non-natural
non-sweatiness	ADJ	non-sweaty
nouniness	ADJ	nouny
on-demand-ness	ADJ	on-demand
one-footedness	ADJ	one-footed
one-wayness	ADJ	one-way
orientalness	ADJ	oriental
other-worldly-ness	ADJ	other-worldly

Table 9 (continued)

nessLemma	basePOS	baseLemma
overthoughtoutedness	ADJ	overthoughtouted
pearshapedness	ADJ	pearshaped
perfectiveness	ADJ	perfective
picture-friendliness	ADJ	picture-friendly
placedness	ADJ	placed
plunginess	ADJ	plungy
polyexpressiveness	ADJ	polyexpressive
poofiness	ADJ	poofy
post-ness	ADJ	post
prisonerishness	ADJ	prisonerish
proportionateness	ADJ	proportionate
quakerliness	ADJ	quakerly
quasi-faithfulness	ADJ	quasi-faithful
retroness	ADJ	retro
rewatchableness	ADJ	rewatchable
ringy-ness	ADJ	ringy
rogueishness	ADJ	rogueish
rusticness	ADJ	rustic
self-reflectiveness	ADJ	self-reflective
sidgwickedness	ADJ	sidgwicked
skullcaplessness	ADJ	skullcapless
slovenlymindedness	ADJ	slovenlyminded
sorbianness	ADJ	sorbian
sovietness	ADJ	soviet
specklediness	ADJ	speckledy
squashyness	ADJ	squashy
squirreliness	ADJ	squirrely
steppiness	ADJ	steppy
strategicness	ADJ	strategic
stroppy-ness	ADJ	stroppy
stuckedness	ADJ	stucked
student-directedness	ADJ	student-directed
surf-consciousness	ADJ	surf-conscious
tailiness	ADJ	taily
tamilness	ADJ	tamil
tapselteerieness	ADJ	tapselteerie
tellingness	ADJ	telling
terracedness	ADJ	terraced
thinglyness	ADJ	thingly
toastedness	ADJ	toasted
tuscanness	ADJ	tuscan
uncenteredness	ADJ	uncentered
unencumberedness	ADJ	unencumbered

Table 9 (continued)

nessLemma	basePOS	baseLemma
unenglishness	ADJ	unenglish
unfinishedness	ADJ	unfinished
unfoundedness	ADJ	unfounded
unperfectness	ADJ	unperfect
unscottishness	ADJ	unscottish
unsuccessfulness	ADJ	unsuccessful
upmarketness	ADJ	upmarket
verbiness	ADJ	verby
way-outness	ADJ	way-out
wholiness	ADJ	wholy
whylessness	ADJ	whyless
winningness	ADJ	winning
wispiness	ADJ	wispy
woman-centredness	ADJ	woman-centred
wordyness	ADJ	wordy
wyrd-ness	ADJ	wyrd
ambridge-ness	N	ambridge
amiganess	N	amiga
applianceness	N	appliance
arrowness	N	arrow
bass-ness	N	bass
blackpoolness	N	blackpool
bodhisattvaness	N	bodhisattva
button-ness	N	button
colness	N	col
compleatness	N	compleat
coupley-ness	N	coupley
cyrusness	N	cyrus
dwam-ness	N	dwam
flagshipness	N	flagship
futureness	N	future
get-togetherness	N	get-together
giraffeness	N	giraffe
god-almightiness	N	god-almighty
grrness	N	grr
guardianness	N	guardian
happeningness	N	happening
hmongness	N	hmong
imageness	N	image
immortality-ness	N	immortality
john-ness	N	john
junction-ness	N	junction
lancashireness	N	lancashire

Table 9 (continued)

nessLemma	basePOS	baseLemma
lawness	N	law
liftshaftness	N	liftshaft
logosness	N	logos
luvvieness	N	luvvie
madonnaness	N	madonna
maverickness	N	maverick
moneyness	N	money
mumness	N	mum
nicaraguaness	N	nicaragua
nottinghamness	N	nottingham
novice-ness	N	novice
our-ness	N	our
peasantness	N	peasant
pilgrimness	N	pilgrim
plusness	N	plus
poetry-ness	N	poetry
refugeeness	N	refugee
request-scopeness	N	request-scope
rileyness	N	riley
roverness	N	rover
sissiness	N	sissy
snakeness	N	snake
somebodiness	N	somebody
spikeness	N	spike
system-ness	N	system
theloniousness	N	thelonious
toreadorness	N	toreador
turkness	N	turk
twitness	N	twit
vectorness	N	vector
waveness	N	wave
zoneness	N	zone
c-ness	other	c
broad-mindness	phrase	broad-mind
like-thisness	phrase	like-this
sugary-sweetness	phrase	sugary-sweet
to-be-heard-of-ness	phrase	to-be-heard-of
to-be-looked-at-ness	phrase	to-be-looked-at
too-much-ness	phrase	too-much
up-side-downness	phrase	up-side-down
when-whereness	phrase	when-where
women-only-ness	phrase	women-only
beness	V	be
sent-ness	V	sent

Table 10: Quoted -ity hapaxes in the ukWaC.

ityLemma	basePOS	baseLemma
aspatiality	ADJ	aspatial
attritablity	ADJ	attritable
cherubicity	ADJ	cherubic
convincibility	ADJ	convincible
deservability	ADJ	deservable
designability	ADJ	designable
dextrality	ADJ	dextral
discardability	ADJ	discardable
discipline-centricity	ADJ	discipline-centric
dispatchability	ADJ	dispatchable
draftability	ADJ	draftable
equiprobability	ADJ	equiprobable
eurocompatibility	ADJ	eurocompatible
exocentricity	ADJ	exocentric
extra-legality	ADJ	extra-legal
freudianity	ADJ	freudian
imposability	ADJ	imposable
inscriptivity	ADJ	inscriptive
intralinguality	ADJ	intralingual
makeability	ADJ	makeable
meltability	ADJ	meltable
photoreality	ADJ	photoreal
plateability	ADJ	plateable
policeability	ADJ	policeable
post-nationality	ADJ	post-national
preconditionality	ADJ	preconditional
simulatability	ADJ	simulatable
sphericosity	ADJ	sphericous
tackability	ADJ	tackable
unalienability	ADJ	unalienable
vendability	ADJ	vendable
worthwhility	ADJ	worthwhile
arch-enmity	N	arch-enemy
bandity	N	band
chavity	N	chav
hamockuity	N	hamock
liftshafticity	N	liftshaft
motority	N	motor
plosivity	N	plosive
polyvisuality	N	polyvisual
second-year-icity	N	second-year
spirality	N	spiral
tackity	N	tack

References

Anshen, Frank & Mark Aronoff. 1981. Morphological Productivity and Phonological Transparency. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 26(1). 63–72. DOI: 10.1017/S0008413100023525.

Arndt-Lappe, Sabine. 2014. Analogy in suffix rivalry: the case of English -ity and -ness. *English Language and Linguistics* 18 (03). 497–548. DOI: 10 . 1017 / S136067431400015X.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Baayen, R. Harald & Antoinette Renouf. 1996. Chronicling the Times: Productive Lexical Innovations in an English Newspaper. *Language* 72(1). 69–96

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3), 209–226. DOI: 10.1007/s10579-009-9081-4.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

BNC. 2007. *The British National Corpus, version 3* (BNC XML Edition). http://www.natcorp.ox.ac.uk/.

Eitelmann, Matthias & Dagmar Haumann, eds. 2022. *Extravagant Morphology: Studies in rule-bending, pattern-extending and theory challenging morphology*. Amsterdam: John Benjamins.

Filatkina, Natalia. 2018. Expanding the Lexicon through Formulaic Patterns: The Emergence of Formulaicity in Language History and Modern Language Use. In: *Expanding the Lexicon. Linguistic Innovation, Morphological Productivity and Ludicity*. Ed. by Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin & Esme Winter-Froemel, 15–42. Berlin & Boston: De Gruyter Mouton.

Hoffmann, Sebastian, Stefan Evert, Nicholas Smith, David Lee & Ylva Berglund Prytz. 2008. *Corpus Linguistics with BNCweb – a Practical Guide*. Frankfurt am Main: Peter Lang.

Kaunisto, Mark 2013. Scare quotes and glosses: Indicators of lexical innovation with affixed derivatives. In *Selected proceedings of the 2012 symposium on New Approaches in English Historical Lexis* (HEL-LEX 3), 97–106.

Körtvélyessy, Lívia, Pavol Stekauer, and Pavol Kamár. 2022. *Creativity in word formation and word interpretation: Creative potential and creative performance*. Cambridge: Cambridge University Press.

Lindsay, Mark. 2012. Rival suffixes: synonymy, competition, and the emergence of productivity. In: *Morphology and the architecture of grammar: Proceedings of the 8th International Morphology Meeting*. Ed. by Angela Ralli, Geert Booij, Sergio Scalise & Athanasios Karasimos. Vol. 8. University of Patras. Patras, 192–203.

Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word-Formation. A Synchronic-Diachronic Approach*. 2nd, completely revised and enlarged. München: C.H. Beck'sche Verlagsbuchhandlung.

Oxford English Dictionary. 1989. Oxford University Press. 2nd ed. 20 vols. Also available at http://www.oed.com. Oxford.

Plag, Ingo. 2006. The variability of compound stress in English: structural, semantic, and analogical factors. *English Language and Linguistics* 10(1), 143–172.

Renouf, Antoinette and Laurie Bauer. 2000. Contextual clues to word-meaning. *International Journal of Corpus Linguistics* 5(2), 231–258.

Wislicki, Jan. 2023. Scare quotes as deontic modals. *Linguistics* 61(2), 417–457. DOI: doi:10.1515/ling-2021-0083.