

Research Article

Anna Shadrova*

No three productions alike: Lexical variability, situated dynamics, and path dependence in task-based corpora

<https://doi.org/10.1515/opli-2024-0036>

received May 24, 2024; accepted November 13, 2024

Abstract: Situated language use is influenced by a number of dynamic phenomena that introduce lexical variability and path dependence, such as fluid discourse granularity, priming, and alignment in dialogue. The empirical tradition of usage-based lexicology does not account for such variability. In fact, its primary theoretical approaches appear to presuppose high population convergence on particular lexemes in language production. This is implied in several key concepts of phraseological and constructionist models, notably entrenchment, the principle of no synonymy, and the idiom principle, as well as the dominance of the statistical paradigm in the field. In spite of its relevance for linguistic theory and corpus methodology, this assumption appears to be untested. This study provides an analysis of inter-individual lexical overlap of verbs and nouns in five task-based corpora of (mostly) German. Results indicate that speakers are maximally variable and highly divergent in their lexical use in spite of narrow communicative constraints and group homogeneity. A qualitative analysis links this variability to situational (cognitive, socio-pragmatic, and discourse-level) engagement with the task and material, which results in referential diversity, spontaneous meaning mapping, and abundant word formation. The degree of observed variability raises questions with respect to the stochastic properties and functional mechanics of entrenchment and the role of repetition of identical material, such as lexicalized chunks, in conventionalization. It further emphasizes the need for a better understanding of the distributions that underlie pooled data, without which the validity of frequential extrapolation to individual behavior and system status stands to question. This is of particular relevance to language assessment and other practices of contrastive analysis of speaker productions, e.g., in multilingualism studies.

Keywords: inter-individual differences, lexical productivity, lexicology, task-based corpora, usage-based theory, complex dynamic systems

1 Introduction

If three speakers from a homogeneous elicitation group were to write argumentative essays on the question ‘Do young people today do better/have it better than previous generations?’, how many of the same lexemes would they use? How many lexemes would overlap on average between any three speakers giving directions based on a map of 12 landmark icons (e.g., images of a motorcycle or a toaster)? What about three speakers reporting events from a 40-s video of a minor traffic accident? More generally, how similar or variable are speakers from homogeneous speaker groups with respect to their lexical use in situated contexts? Do they cluster into groups by similarity? Is there a general convergence on a basic common set of vocabulary by topic

* **Corresponding author: Anna Shadrova**, Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik, Unter den Linden 6, 10099 Berlin, Germany, e-mail: anna.shadrova@hu-berlin.de

or prompt? If so, how many or what proportion of lexemes does it include, and are those distributed frequently similarly between speakers?

Surprisingly, we do not know. Although task-based corpora have been prolifically studied for decades, there appears to be no research into within-speaker and between-speaker variability of lexical distributions therein. Perhaps more surprisingly, the existing literature offers divergent and even contradictory predictions on this question.

On the one hand, recent research highlights the abundance of inter- and even intra-individual variability in language production on all linguistic levels (Özsoy and Blum 2023, Verspoor et al. 2021, Tsehay et al. 2021, Zerbian et al. 2022, Alexiadou et al. 2022, 2023, Alexiadou and Rizou 2023, Shadrova et al. 2021, Wiese et al. 2022). Moreover, the field has been familiar with a range of dynamic phenomena in situated language use for a long time, e.g., priming, alignment in dialogue, spontaneous category formation, constructional and metaphor blending, or reanalysis/reinterpretation of linguistic structure. Phenomena of this kind introduce local path dependence, i.e., changes in production that are based on previous decisions within the same text.

On the other hand, key concepts of the usage-based frameworks with a tradition of lexicological study – constructionist, phraseologist, and variationist accounts in particular – appear to presuppose a high degree of between-speaker similarity of contextualized lexical expression. This includes *entrenchment* (Balog 2023, Blumenthal-Dramé 2013, Stefanowitsch and Flach 2017) and the *principle of no synonymy* (Leclercq and Morin 2023, Uhrig 2015, Goldberg 1995), as well as functional explanations of phraseological tendencies in natural language. It also applies to the methodological dominance of the statistical paradigm in the field. However, the proposed mechanics of those models allow for multiple interpretations that, often implicitly, exist side-by-side in the discussion (Sections 2.2 and 2.3). Along with a tendency to limit their scope to linguistic abstractions, this leaves them generally underspecified with respect to production in context.

This lack of theoretical and empirical clarification is unfortunate from a methodological perspective, as an understanding of the underlying distributions is a basic prerequisite for valid inference from corpus data. To fill the empirical gap, this contribution provides an analysis of the lexical overlap between all sets of any three and four speakers in five corpora of L1 German representing common elicitation formats: two corpora of thematically prompted essays (FaLKo, Hirschmann et al. 2022, and Zinsmeister et al. 2012), two dialogue corpora based on map and picture description tasks (BeMaTaC, Sauer and Ludeling 2016, BeDiaCo Belz and Mooshammer 2023), and a corpus of witness reports of events from a short video (RUEG, Ludeling et al. 2024). All elicitations were performed under controlled conditions and feature homogeneous speaker groups in terms of regional distribution, age, and (to a lesser extent) educational and socioeconomic background. One of the corpora is as homogeneous as we can realistically hope for elicited data, as participants were all recruited from the same class of a local high school (the Kobalt essay corpus, see Section 3 for details). Thus, this study is not designed to assess diatopic or diastratic variability between groups of speakers, but intra-situational variability within a stratum of speakers. As the witness report corpus (RUEG) features four elicitations under slightly varying contextual conditions per speaker (formal/informal × written/spoken), it further allows for an analysis of intra-individual variability and the influence of some register aspects. The RUEG corpus further contains parallel elicitations of Russian and English L1, which are included to test for language specificity. In anticipation of the results, the reader is encouraged to make an estimation of overlap for the respective text types for later comparison.

The article is structured as follows. Section 2.1 reviews a number of factors that introduce variability and path dependence in lexical production. Section 2.2 reviews the phraseological and constructionist models with respect to their explicit and latent predictions for lexical production in context. Section 2.3 presents the mathematical entailments of some statistical arguments in the field and the logic they imply for the mechanics of the lexicon. Section 3 presents the corpora in detail and derives hypotheses pertaining to the various influences. Section 4.1 presents the results of an analysis of inter-individual lexical overlap of nouns and verbs for all corpora, and intra-individual overlap between elicitations in RUEG. Section 4.2 offers a qualitative perspective that shows how abundant variability is created by a range of path-dependent phenomena even within narrow communicative constraints. Section 5 ties these results back to the usage-based models of lexical production and relates them to the common practice of extrapolating to individual behavior from cumulative corpus frequencies, and to mistaken perceptions of lexical identity in linguistic research.

2 Lexical productivity in usage-based linguistics

The term *lexical productivity* is multiply ambiguous. It can refer to the lexemes used in a particular context by (a) an individual speaker or (b) a population of speakers as measured in lexical diversity, lexical richness, etc. (Dewaele and Pavlenko 2003, De Smet 2020, Song 2008, see below for more references). It can also refer to the types that occur in a corpus – either (c) all lexemes or (d) only words newly coined from word formation processes, such as compounding or derivation, or other constructional activity (*type productivity*, e.g., Baayen and Lieber 1991, Fernandez-Dominguez 2010, Demo 2022).

The study of individual productivity is a study of speakers (psycho-centric). It aims to describe factors and limitations in speaker systems or behavior and is prominent in the context of language dynamics research (L1 and L2 acquisition, attrition, lifetime change, etc.). The study of population productivity is a study of language (lingua-centric). It is of a lexicographic or morpho-lexicographic nature, as it aims to model processes and constraints at the morpholexical and lexicosyntactic interfaces of language as a social ontology (Silvennoinen 2023), rather than the lexicon in use as a cognitive/individual phenomenon.

The focus of usage-based lexicology has been primarily lingua-centric with contributions from various theoretical branches, including traditional approaches to phraseology ('words that go together'), productivity studies of particular morphemes and semi-filled constructions, and variationist perspectives on the distribution of various linguistic items across genres, registers, or social and regional strata (Section 2.1). To date, there appears to be close to no corpus-linguistic research that connects evidence of individual behavior with population patterns of lexical productivity in adults (see Anthonissen 2020; Dąbrowska 2020 for similar observations). Rather, individual lexical productivity has been studied almost exclusively in measures of overall lexical or phraseological *diversity*, *richness*, or *sophistication*, which largely¹ abstract from lexical content, i.e., the actual lexemes used and their semantic or morphological properties (Paquot et al. 2019, 2022, Vandeweerd et al. 2022, Hu et al. 2022, Jiang et al. 2023, Luckman et al. 2020, Rubin et al. 2021, Lei and Yang 2020, Kim et al. 2018, Kalantari and Gholami 2017, Gebril and Plakans 2016, Kyle and Crossley 2015, 2016, Schmid and Jarvis 2014, Jarvis and Daller 2013, Fergadiotis 2011, McCarthy and Jarvis 2010, Meara 2005, and many others).² Only very recently have corpus-linguistic studies begun to look into the productivity patterns of individual adult L1 speakers (De Smet 2020, Neels 2020, Fonteyn and Nini 2020). These studies still focus on particular constructions rather than considering a speaker's overall engagement of the lexicon in solving a communicative task; and are lexicographic relative to an individual speaker's lexical system, not descriptive of their lexical behavior.

This lopsided tradition is reflective of a wider research gap within the usage-based framework, which, in spite of its particular emphasis on the concrete-form realizations of linguistic signs (words and chunks) (Bybee 2013, Ellis 2008, Wulff 2020), its explicit aspiration towards cognitive plausibility (Ibbotson 2013, Arppe et al. 2010), and its empirical principle of viewing language as observable and represented in corpora and linguistic behavior, appears to produce little to no research into situated production.

There is no intrinsic problem with a division between the study of the individual and the population, or behavior vs system, at least where research is concerned with minimal hypotheses about the potentials of human cognition. For instance, we can conclude that (some) speakers must *be able* to coin new words from certain patterns if we find those words in the corpus. However, no conclusions can be drawn about speaker behavior from the same data, for instance, that speakers will *typically* coin new words in a given situation – only a comparison between speakers in the same situation can yield insight on this matter. *Typicality* here goes hand in hand with *generalizability*. This is due to a simple, but powerful, methodological constraint: corpus data does not contain abstractions – it consists entirely of individual speaker productions. It is crucial to know if, when, and to which degree speakers make use of particular realizations in order to determine justifiable

¹ Some complexity measures take into account certain classes of word formation, e.g., from Latinate or Greek affixes, or consider the number of syllables or other form-based aspects of lexemes, which correlate with some semantic properties.

² Some of the measures are developed primarily as diagnostic heuristics rather than for lexicological or linguistic purposes. Nevertheless, they are corpus-based measures of lexical productivity.

generalizations and abstractions from data. Thus, usage-based linguistics cannot limit the scope of its research to the social ontology of linguistic abstractions without also understanding speaker behaviors, unless it were to cease its empirical commitment.³

The empirical tradition of lexicology in usage-based linguistics has implicitly accepted at least partial equivalence between population averages and individual lexical production. This is methodologically reflected in the domineering statistical paradigm that views corpus frequencies as linguistically meaningful, e.g., for the gradability of phenomena like idiomaticity or fixedness (e.g., Wulff 2008), and as representative of average speaker behavior (van Trijp 2024, Partington 2017). Corpus averages are also commonly accepted as proxies for speaker input in constructionist and phraseological approaches, which model the acquisition and production of linguistic patterns as a reflection of frequency-based observations (Bybee 2001, Ellis 2002, Tomasello 1995, Tomasello 2009, Stefanowitsch and Flach 2017). This includes the transmission of detailed lexical distributions for fine-grained semantic differentiation, implying a relevant degree of transitivity of lexical features between a speaker's cumulated input and average individual behavior (Section 2.2).⁴

The problem with averages over population data is that they come at a loss of information about their underlying distributions, which can be problematic. While averages may adequately represent the data in certain cases, they can be rather misleading in others. For instance, the average number of liters of coffee drunk per person per year may predict an average person's coffee consumption. Although some people drink much more or none at all (those are the tails of the distribution), for a large chunk of the population, within a range of variance, predictions would approximate reality. Unlike this, the mean pigmentation density of a zebra neither predicts its black nor its white stripes well; while the average darkness of a 24-hour period adequately describes two predictable, albeit brief stretches of each day, and only under certain geographic and seasonal constraints.

To be epistemologically sound, usage-based theories need to provide testable hypotheses on whether or not – and when and where – they expect distributional equivalence between individual speaker behavior, individual system manifestations, observable population patterns, and population abstractions, and what this equivalence would look like in terms of lexemes and their numbers and distributions in context.

Full equivalence is unlikely to be theoretically intended or plausible, as speakers are free to express novel meaning and also experience the situational dynamics of local communicative and cognitive interaction, while populations do not.⁵ However, the impact and proportion of these dynamic phenomena has decisive effects for the logic of the linguistic arguments as well as the applicability of statistical models, which fundamentally rely on the reproducibility of the same approximate distributions from the same approximate factor combinations (Section 2.3). The following section outlines the types of dynamics and path dependence that can be expected.

2.1 Factors of variability and path dependence

Corpus linguistics has viewed variability primarily through one of two lenses: (1) as a function of language dynamics on a personal or population-based trajectory of language change, including variability in L1 or L2

³ Language is sometimes conceptualized as a 'phenomenon of the third kind' (Dąbrowska 2020) or as directed by 'an invisible hand' (Keller 2014). Those terms, borrowed from economist Adam Smith, are used to describe symbiotic, pluralistic, and emergent human systems, in which individual agents act in accordance with individual needs, but their cumulative decision-making results in the emergence of directed movement for the entire system (e.g., increased financial well-being for all), as if it was guided by an invisible hand. Keller integrated the idea into historical linguistics in particular, arguing that linguistic study should focus on the emergent properties of the system, not the individual production. However, even if language was correctly identified as such as system, the language we study in corpora is not that system. There is no direct access to the linguistic system at higher levels of abstraction.

⁴ Transitivity here refers not to the syntactic or semantic property of a verb, but the mathematical property of transcendence of a feature between levels of an order or a set within a transitive relation, for instance, if $a < b$ and $b < c$, then $a < c$; or if set $S1 \subset$ (subset of) set $S2$ and set $S2 \subset$ set $S3$, then $S1$ also $\subset S3$ (which means that all elements included in $S1$ are also included in $S3$).

⁵ This does not refer to types of situations and dynamics, e.g., *speaking to a friend to agree on a new time to meet*, but their particular manifestations with situational parameters in context (which friend, which activity, etc.).

acquisition (Ädel 2015, Kachkovskaia et al. 2017, Dirdal 2022, Kerz and Wiechmann 2020, Lowie and Verspoor 2019, Vyatkina et al. 2015, Lowie and Verspoor 2015, Lowie 2017, Verspoor et al. 2021, Larsen-Freeman 2019) or language attrition (Martin Villena 2023, Matos and Flores 2024), and language contact situations and diachronic change (Hilpert 2017, Hernandez-Campoy 2021, Karjus et al. 2020). Since language development and change are not the subject of this article, these perspectives will not be further discussed. Or (2), as a function of situational, regional, and social language stratification or differentiation. This branch includes the variationist approaches to corpus linguistics (Cacoullos and Travis 2019, Ludeling 2017, Szmrecsanyi 2017, Biber et al. 2016), in which variability is attributed to diatopic (e.g., regional/dialectal) and diastratic differences (age, level of education, etc.), or intra-individually functional, inter-situational aspects, such as differentiation by register (Pescuma et al. 2023, Szmrecsanyi 2019, Biber 2012). Recent psycholinguistic work further suggests that individual grammars may not cover the full syntactic range of a language community, but vary in accordance with factors such as exposure to written language or level of reading engagement (Dąbrowska 2014, 2018, 2019, Petre and Anthonissen 2020).

All of these factors may fluctuate to some degree over a lifetime, but are expected to be largely stable and controllable in elicitations. The variationist model can therefore be described as a factorized function of language production, in which each factor that contributes to the choice of a particular linguistic expression is generalizable beyond the current situation (either to the speaker or to the type of situation). This includes an amount of free or random variation that emerges from ambiguities within the system itself, such as parallel structures and their reanalysis in language contact and beginning language change. The variationist model is therefore a typology of *systemic* variation relative to the system of a speaker or the language itself – differences in linguistic expression as we find them in corpora are viewed as classificatory of underlying, causal differences in features or functional choices.

More recently, the field has seen an increase in attention towards intra-situational inter-individual differences in language production within identifiable, factorially homogeneous strata of speakers, i.e., speaker groups that are controlled by age, regional, and societal aspects such as urbanicity and socio-economic factors. (Özsoy and Blum 2023, Tsehay et al. 2021, De Smet 2020, Zerbian et al. 2022, Alexiadou et al. 2022, 2023, Shadrova et al. 2021, Wiese et al. 2022). Naturally, it is impossible to control for all potential factors of variation that come to mind. Some of the variability observed in these studies may thus still be attributable to stable and predictive, but currently unaccounted factors.

Beyond this, there is evidence to suggest that a considerable amount of variability may be linked to effects of local path dependence, some of which are conceptually well known in linguistics. In a path-dependent system, what has previously happened and influences what is going to happen later. Path-dependent phenomena are ubiquitous in human cognition, for instance, in priming (the persistent activation of concepts once they are introduced and the temporary co-activation of related concepts, Baayen and Smolka 2020, Kidd 2012, Melinger and Dobel 2005, Arias-Trejo and Plunkett 2013, Kaschak et al. 2014, Lucas 2000, Hutchison et al. 2008), categorization (e.g., family resemblance, in which members are assigned to a category based on the features of other, previously assigned members of the same category, Wittgenstein 1953, Rosch and Mervis 1975), or the ongoing construction of concepts and mental models (in which categories or paradigms, once established, define their own idiosyncratic constraints and plausible heuristics therein, see Kuhn 1970, Lakoff 1987, 1999, Tversky and Kahneman 1974, Neth and Gigerenzer 2015). In dialogue, path dependence is created by alignment or convergence between speakers on (morpho-)syntactic, phonological, semantic, lexical, and conceptual levels (Pickering and Branigan 1998, Garrod and Pickering 2007, Watson et al. 2004, Steels and Loetzsch 2006, Dobnik et al. 2015, Branigan et al. 2000, 2007, Reitter and Moore 2014, and others). In a similar sense, an ongoing process of synchronization between speakers in dialogue and speaker communities allows for temporary and locally stable meaning mapping (such as a locally known way of referring to a certain place), thereby creating idiosyncratic language bundles that may diverge from the larger speaker population. Synchronizations like these have been suggested as initialization points for grammaticalization and language change (Schmidt and Herrgen 2011), although that would first require their stabilization and re-convergence between communities.

Path dependence can also be created via morphological self-priming, morphophonotactic preferences, or collocational associations. For instance, in other analyses of the essay corpora also used in this study, Shadrova

et al. (2021) found clustered occurrences of split particle and prefix verb types within 50 tokens after the first verb of the same morphological type was used. This persistence does not appear to be driven by graphematic similarity or semantics,⁶ thus introducing many new and unrelated lexemes. For coselectional path dependence, Shadrova (2024) found a tendency in German L1 speakers to prefer polysyllabic verb + noun combinations (4+ syllables for both elements). The selection of a polysyllabic verb *or* noun then raises the local probability of also selecting the other, creating a path dependence. Similarly, the choice of one collocate has been argued to increase the local probability of the other in coselectional association (Gries 2013, 2023, Brezina 2018, Laufer and Waldman 2011, Hoey 2005 and others).

Path dependence also follows from the choice of discourse topic on various levels of resolution. Even where macro-level variability is limited by prompt or purpose, speakers are generally free to choose their content for meso- and micro-levels. This may sound trivial, but the choice of the many subdiscourses related to a general thematic field (such as, in the case of *a better life for younger generations?*, the subtopics *education or family life*) and the various levels of granularity (e.g., *amounts of homework in primary school vs how school prepares students for life*) have far-reaching consequences for lexical distributions and between-speaker overlap. This is a common issue with prompt-based elicitations, e.g., in learner corpora. Observable repercussions of path-dependent processes on the frequencies of lexical and morpholexical categories in corpus elicitations, including strong task effects, have been abundantly found in this line of research (Lüdeling et al. 2017, Alexopoulou et al. 2017, Vyacheslavovna and Il'Inichna 2022, Pérez-Paredes and Bueno-Alastuey 2019, Gilquin 2021, Shadrova et al. 2021, Shadrova 2020). We will return to this in Section 4.2.

Path dependence and local non-stationarity are generally intrinsic features of *complex adaptive* or *complex dynamic systems* (CAS/CDS) – multi-agent systems that enter states of dynamic equilibrium through the ongoing ecological or functional adaptation of their agents without centralized control loops.⁷ Usage-based linguistics has a tradition of viewing language as a CAS (van Trijp 2024, Al-Hoorie et al. 2023, Verspoor et al. 2021, Lowie and Verspoor 2019, Lowie 2017, Beckner et al. 2009, Steels 2000). CAS expects a high degree of variability at all times and local convergence of speaker performance is not presumed, although the total system could in principle still converge in its full distributional realization over time/on average, e.g., in cyclical developments. Empirical approaches to CAS hypotheses are relatively young, and no explicit lexicological theories or analyses of lexical production in context appear to have been published to date. In an individual speaker's CAS, an interaction between lexical retrieval, the organization of the mental lexicon, and other situational factors as named above would be expected.

Lexical retrieval has been psycholinguistically shown to be influenced by many factors, including semantic field, semantic similarity, lexical ambiguity, and the set of alternatives available in context (Nickels et al. 2022, Anders et al. 2015, Jolsavi et al. 2013), morphological relationship and phonological and graphematic neighborhood (Li et al. 2024, Vitevitch et al. 2014, Smolka et al. 2014, 2019, Baayen et al. 2006, Luce et al. 2000), multilingualism (Sullivan et al. 2018, Kambanaros et al. 2013), frequency of a word and frequency of co-occurrence with other words (Agustin-Llach and Rubio 2024, Jacobs et al. 2017, Mark Knobel and Caramazza 2008, Ellis 2002), attentional control (Roelofs 2008), and phonological aspects (Vitevitch 2022), among others. This line of research derives properties of the organization of the mental lexicon from measures of word production, association, and recognition from experimental settings. Contextualized production is not usually considered due to the complex challenges of factor isolation in situated context. Although certainly meaningful for situated results, the operationalization of those properties for corpus-based studies is not straightforward. While the process of language production must be mediated by the structure of the mental lexicon, it may be determined by many additional factors. Furthermore, the mental lexicon itself appears to be highly fluid and malleable (even between experiment trials according to Heitmeier et al. 2023), raising the likelihood for situational dynamics to dominate the scene. To derive concrete hypotheses for production in context, a synthesis with the interactional, communicative, and cognitive factors named above as well as a more general

⁶ This is in line with morphological priming studies in German (Smolka et al. 2014, 2019).

⁷ Whether or not CASs and invisible-hand phenomena are functionally and philosophically equivalent lies beyond the scope of this article.

theory of mindwandering or cognitive association in context (which is generally understudied, see Fossa et al. 2019), would be required. This would be an immense undertaking and does not appear to have been tackled yet.

Another underconsidered source of variability is the general diversity in situational perception and interpretation. As a classic linguistic example, Chafe's (1980) pear stories show that speakers focus on different aspects of a story in their narrations of events witnessed in a video (in which a character picks pears, which are then stolen). This results in major differences in narrations and, consequently, divergent lexical expression. As perception and sensory interpretation is fundamentally selective, filtered through an individual's physiological/neurological, cognitive, cultural, social, psychological, and biographical setup, current interest, communicative goals, and other factors, variability in interpretation can be expected with respect to all aspects of a communicative situation (Lamy et al. 2012, Kemmerer 2023, Blomberg and Zlatev 2021).

Unlike globally stratified variation (by register/genre, age, social and regional factors), path-dependent variability is expected to occur inter- as well as intra-individually in near-identical contexts. A person may generally have a particular perceptual, lexical, collocational, etc. preference, and they may also situationally modulate their attention to include or exclude certain aspects or temporarily prime themselves to favor a particular expression.

In corpus linguistics, path-dependent phenomena and other production phenomena have found little attention until now, with the exception of some discussion of persistence/structural self-priming (Gries and Kootstra 2017, Gries 2005, Reitter 2008, Szmrecsanyi 2006). Usage-based lexicology has instead primarily been shaped through debates from phraseological and constructionist models of language. These will be reviewed next.

2.2 Phraseological and constructionist models of lexical production

Phraseological and constructionist views take a largely lexicographic perspective and tend to provide structural, rather than procedural, descriptions. However, as both approaches make claims towards acquisition and processing, they programmatically position themselves as partially cognitive theories of language. Thus, predictions for active language production should be derivable from the interplay of their proposed structures. This is complicated by the fact that some of their core aspects remain vague and underspecified 30 years into their development (Diessel 2023, Ungerer 2024). Some discussion of their underlying principles is hence required for an adequate embedding of the present study, especially in the absence of other, more explicit theories of the lexicon in situated use.

Phraseological and constructionist models of the lexicon agree on some core assumptions, such as the rejection of separability of lexicon and syntax and the importance of lexical material in the acquisition of linguistic structure from input (Bates and Goodman 1999, Boas 2010, 2013, Bybee and Torres Cacoullos 2009, Bybee 2013, Goldberg 2013). In their methodological tradition, both approaches overwhelmingly operate from a *bag-of-words* approach to lexical production, which discards the order and context of words and constructions from the analysis of corpus data. This appears to be a pragmatic choice that serves the simplification of analytical procedures, e.g., for more accessible normalization against the total number of lexemes or to avoid the heavy load of manual annotation of semantic context. However, this practice also carries epistemological weight, as it marks a dimensionality reduction of a linguistic macrostructure such as a paragraph or a text, which is defined by production time, speaker, situation, etc., into an unstructured set of mutually independent words. Although not explicitly stated, the general acceptance of this detachment and isolation of the occurrences of phrases and constructions from the context of their realization implies a model of lexical production as a series of independent actualizations of global generative patterns represented elsewhere (e.g., the mental lexicon).⁸

⁸ The clustered occurrence of identical forms within a small window of text is sometimes accounted for with measures of dispersion (Gries 2022, Egbert et al. 2020). Other interrelationships between lexemes are not typically considered.

Through these similarities in modeling and empirical study, the two models have found some convergence in the postulation of the *phraseological continuum* (Granger et al. 2008; Gyllstad and Wolter 2016), and it is not uncommon for research within the field to be rooted in both accounts or view them as complementary or as simultaneously operating on different levels of linguistic abstraction, e.g., as form- vs schema-based perspectives (Ordines 2023, Mellado Blanco et al. 2022, Pavlova 2020, Ziem 2018, Stumpf 2016, Granger et al. 2008, Granger and Meunier 2008, Benigni et al. 2015). However, for the purposes of this study, their philosophical differences with respect to the role of conventionality and their predictions for lexical variability in context are of particular interest. These differences will be highlighted here in spite of the shared community of the two approaches.

Today's phraseological accounts originated from earlier lexicographic catalogues of idiosyncratically, often non-compositionally combined lexical material, for instance, in the Russian school of phraseology (Zykova 2016, Guliyeva 2016) and in the early computational lexicography of English collocations (Berry-Rogghe 1970, Sedelow 1985). Since then, they have synthesized into a common research paradigm with psycholinguistic and cognitive subdisciplines, which views language production as primarily defined by the reuse of prefabricated lexical material (Siepmann 2012, Siyanova-Chanturia 2015, Wray 2002, 2012, 2013, Pawley and Syder 1983, Ägel 2004, Poulsen 2022, Gyllstad and Wolter 2016, Granger 2005, Schmitt 2004, Conklin and Schmitt 2012, Wood 2015, Frath and Gledhill 2005, Jolsavi et al. 2013, Tavakoli and Uchihara 2019, among others. See Singleton and Leśniewska 2021 and Shadrova 2020 for recent synopses). This programmatic view has perhaps been most prominently formulated in Sinclair's widely cited *idiom principle*, which states that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments" (Sinclair 1991, 110). To Sinclair, communication outside of highly specialized language such as legal registers or poetry can explicitly be attributed almost exclusively to phraseological material (p. 114). Similarly, Wray (2002, 119) formulates 'formulaic processing as the default' and 'construction out of, and reduction into, smaller units by rule [as occurring] only as necessary'. Similar positions were previously formulated by Bolinger (1979), as well as Pawley and Syder (1983), to whom native-like fluency and native-like selection are 'two puzzles of language use': too challenging to solve if language was reconstructed from the lexicon for each utterance.

At first glance, it may appear as though phraseology is primarily a study of 'words that go together', rather than one of which words occur at all. However, functional arguments presented within the field and the scope of its summarizing contributions suggest that phraseology views itself as a philosophical approach seeking "to answer the age-old question of whether human beings are 'primarily like buses, which travel along regular routes' or 'like taxis, which move about freely'" (Ding 2018, 144). Moreover, as recombination is a function of selection, there is a strict entailment of the predictable occurrence of particular lexemes in the realization of the chunks that include them, as well as in the statistical associations of words and their coselected elements (Section 2.3). To simplify, if I predictably choose two words together, each one of them will also occur. The fact that the phraseological paradigm does not view them as individually selected does not change that.

Phraseological accounts do not typically make statements regarding lexical invariability or precise predictions as to which chunks will be reused in similar contexts. Rather, they explicitly postulate only the availability of such formulaic material. However, two much cited arguments suggest functionality of limitation, i.e., the reuse of a small and predictable set of chunks or otherwise coselectionally constrained elements.

First, phraseological material is assigned semiotic status. This means that not words, but identically recurrent combinations are viewed as the basic communicative signs (Wray 2002, Sinclair 1991). This more complex composition of communicative signs is suggested to be of functional advantage due to a limitation of choice in context: rather than picking words from sets of many near-synonyms, which may all carry slightly divergent connotations, a community of speakers can settle on one conventional way of expression. Although this necessarily limits referential precision, as situations never recur in truly identical ways, it is argued that speakers and hearers profit from a limitation of semantic decomposition effort and lower contextual ambiguity (Siepmann 2012, Wray and Perkins 2000). A semiotic view of chunks as primary communicative units has also been prominent in first language acquisition research (Tomasello 2000, 2009, MacWhinney 2014, Bates et al. 1988, Borensztajn et al. 2009, Dąbrowska and Lieven 2005, Ellis et al. 2008, Ellis 2006, and others). This functional argument would fail if there was in fact no limitation of choice in context, i.e., if many chunks

competed for actualization. In fact, the availability of many chunks for the same situation and the incorporability of multiple near-synonyms into their slots would exacerbate the initially proposed problem: a competition between the full range of semantically plausible chunks is simply a more complex and combinatorially larger version of a competition between the full range of semantically plausible words.

Furthermore, the concept of conventionalization by definition implies the limitation of expression to only a subset of the full generative potential. Even if there may be *a few* equally conventional ways to refer to a particular event, expanding this to the entire semantically plausible lexical set undoes the phraseological convention. To say that *in a particular context, any of the semantically available words are equally conventional* is to say that the explanatory power of conventionalization does not extend beyond semantic plausibility.

Second, phraseological accounts of the lexicon have underscored their functional interpretation with a range of cognitive arguments derived from psycholinguistic studies, arguing that formulaic language offers decisive processing advantages in production, perception, and sociopragmatic interpretation by retrieval of lexicalized chunks vs analytical or composite processing (e.g., Conklin and Schmitt 2012, Jiang and Nekrasova 2007, Carrol and Conklin 2020, Schmitt 2004, Weinert 1995, Underwood et al. 2004, Siyanova-Chanturia 2015). The functionality of this minimization of morphosyntactic processing and recombination effort is only given if it is not outweighed by the effort of selection. It is well evidenced that lexical retrieval from large sets of alternatives is slow due to higher inhibition demands (Sullivan et al. 2018, Kroll et al. 2013, Dijkstra 2003, Fargier and Laganaro 2023, Wulff et al. 2016), especially for competing, partially overlapping forms (as would be the case for similar chunks of different lexemes). Thus, if a processing advantage exists for formulaic sequences, it could only persist within a lexically limited space, although perhaps less strictly so than in the semiotic argument.

A global scope of the phraseological view, i.e., the idea that formulaic material is used whenever possible, is further suggested in estimations of the extent of formulaic material, which has been proposed to reach 50% of regular discourse (Conklin and Carrol 2021; Nelson 2018; Erman and Warren 2000) and up to 80% or even 90% for local phenomena (e.g., demonstrative cleft phrases, Calude 2008).⁹ If processing advantages explain phraseological tendencies, they should not be limited to phraseological niches, but expand to cover most everyday situations, as this would provide the largest cumulative advantage (see also Ellis 2008).¹⁰

Constructionist approaches and Construction Grammar (CxG, Goldberg 1995, 2006, 2013, Sag 2012, Boas and Sag 2012, Steels 2006, Croft 2001) are of a more generative nature compared to phraseological accounts. They view the lexicon as a set of schematic templates for higher-level abstractions that allow for the blending of exemplars and constructions with one another.

At the same time, the semantic model of Construction Grammar presumes a distribution of words into exact semantic niches. This is explicitly stated for Construction Grammar in the *principle of no synonymy* (Goldberg 1995, 67): “if two constructions are syntactically distinct, they must be semantically or pragmatically distinct,” and similarly in Croft’s (2001, 111) *principle of contrast* (see also Leclercq and Morin 2023, Uhrig 2015 for more discussion). The principle of no synonymy is equivalent to the structural model of a very detailed semantico-pragmatic map that connects each linguistic form to clearly delineable meaning including socio-pragmatic information on where and when it can be used. This map, termed *Constructicon* in CxG, is modeled

⁹ It should be noted that these estimations often stem from questionable research practices, such as asking annotators if they have seen or heard a phrase before, or cite as their only source observations of recurrent material through corpus concordancers. However, even if annotators were unbiased in their perception, the repetition of material in and of itself is not evidence of phraseological tendencies – it may well be a reflection of simple grammatical operations on common vocabulary. Repetition only becomes meaningful when it reflects a high proportion against its chance of occurrence, i.e., the number of times a particular meaning is expressed or the number of situations that could call for its expression, rather than the occurrence of any other, unrelated material in the corpus. This is difficult to assess and cannot be satisfactorily resolved with statistical measures of collocational strength or similar (Shadrova 2020, 2021, 2022).

¹⁰ Similar views have been formulated in other theories of language production, e.g., Basic Language Theory Hulstijn 2024, 173: “[...] adult native speakers understand and produce utterances predictively and probabilistically. Thus, learners do not create, and language users do not possess a mental, rule-based or principle-constrained grammar [...] This is made possible as a result of massive (earlier) exposure to, and processing of, spoken input. [...] representation does not exist separate from use (speech processing).”

to contain detailed information on the distributional niches for alternations, verb argument structure preferences, and near-synonyms, and is proposed to emerge from exposure to statistical patterns in use (Gries 2010, 2013, 2015, 2019, Kerz and Haas 2009, Wulff and Gries 2021, Faulhaber 2011, Zeldes 2013, 2012, Diessel 2023, Shadrova 2020).

Although the principle of no synonymy was originally formulated for abstract constructions, it must extend to words, as a continuity between lexicon and syntax (or abstract and concrete constructions) as form-meaning pairs of more or less identical buildup is one of the core principles of the constructionist theory (Bybee 2002, Bybee and Torres Cacoullos 2009, Ellis 2008, Ellis et al. 2008).

As Uhrig (2015) points out, the creation and sustenance of this map entails global invariability: if speakers are highly sensitive to very fine-grained semantic differentiation, they can only convey their intended meaning by realizing the correctly mapped lexemes. If this map is further created from usage and exposure, it would lose precision from variable use by different speakers in the same semantic situation.

The expected invariability of semantic niches raises questions with respect to the place and role of schema-based innovation. Since schemas only ever occur with lexical realizations in language production (one cannot realize an accusative object without using words), schema-based entrenchment entails token entrenchment within the schema (Schmid 2018), which should stabilize as a phraseological object in the semantic map. On the other hand, innovation through the blending and reanalysis of constructions has traditionally been at the center of constructionist modeling. For instance, a speaker may choose to say “she sneezed the napkin off the table,” although they would certainly have the option to express very similar meaning from existing chunks, e.g., [*she sneezed*][*and*][*caused*][*the napkin*][*to fall off the table*]. It is not very clear when, if, or why a speaker would be pragmatically *required* to apply a particular constructional blend to express meaning in context, i.e., whether particular meanings demand innovative realizations. There are no explicit predictions on the type of situation that would necessitate, justify, warrant, or license the use of the causative construction in *she sneezed the napkin off the table* as opposed to other, e.g., phraseological, variants. In the context of the phraseological continuum, a division between everyday phraseological vs out-of-the-ordinary constructional activity in lexical production could be presumed; or constructional blending may be seen as the generative mechanism contributing to a phraseological potential for conventionalization. Similar questions regarding terminological rigidity and situational prediction vs linguistic abstraction have recently been raised from within the framework itself (Michaelis 2024, van Trijp 2024). Another underdiscussed aspect with repercussions on the lexicon in use concerns the methodological paradigm of both approaches. This will be discussed next.

2.3 The statistical paradigm

Lexicology within usage-based linguistics is marked by the dominance of a statistical paradigm, especially in corpus linguistics (Larsson et al. 2022, Larsson and Biber 2024). This includes in particular a practice of accepting lexicological evidence from probability-based metrics computed from relative corpus frequency. These metrics count the occurrence of words within some particular context and derive conclusions to either the properties of the context in question or the properties of the words themselves in a lexicographic sense. The most influential among them are the following:

- (a) Lexicographic metrics, including measures of collocation or word association, mutual information, odds ratio; or productivity measures of types vs tokens. These are usually transformations of the relative corpus frequency of a lexeme or a combination of lexemes against the total number of words (Stefanowitsch and Gries 2003, 2013, 2023, Evert et al. 2017, Pecina 2010, Shadrova 2020). Context is defined by the co-occurring lexemes or constructions within a small window of the target words (e.g., ‘strong’ + ‘coffee’ within five tokens) or by other words of the same class (e.g., the other nouns for noun productivity). Some of these measures are purely descriptive and do not make explicit probabilistic or inferential claims (e.g., type/token ratio), while others argue from an explicitly stochastic perspective with inferential claims towards generalizable features (e.g., significance testing with the Fisher–Yates exact test for collocation analysis; Bayes’ theorem of conditional probability for ΔP , see below).

- (b) Regression and mixed-effect modeling, a group of stochastic techniques that probabilistically predict outcomes from combinations of influential factors (Gries and Otani 2010, Gries 2015, Siyanova-Chanturia and Spina 2020, Speelman et al. 2018, Barth and Kapatsinski 2018, Bates et al. 2015, and others). A function is inferred from scales or levels of factors and mapped to a distribution of data points to match it as closely as possible. If the inclusion of a factor improves the congruence between distribution and function, it is interpreted as adding probabilistic prediction value to the model. Context is defined as the environment of factors that predict a variable. In corpus linguistics, factors have included speaker attributes as well as syntactic and semantic environment and concrete lexical variables (e.g., to infer a higher rate of realization of ‘that’ in subclauses that complement particular verb lexemes, see Gries 2021).
- (c) Clustering techniques, such as topic modeling, multidimensional analysis, and certain applications of distributional semantics (Biber 2019, Biber and Jones 2009, Bruni et al. 2014, Vogt et al. 2023, Shadrova 2021), which infer from a set of co-occurring words to a higher-level linguistic concept such as meaning, genre, or semantic field/topic. Context is not categorially defined in these techniques but emerges as a latent construct from the clustering procedure itself. For this, words and other features are randomly assigned to groups and iteratively reordered until a mathematical function is maximized or converges. The resulting division of data is viewed as linguistically meaningful. The probabilistic aspect here usually refers to the likelihood that a particular combination of words represents one class (topic, genre, register etc.) over another.

Probabilistic reasoning is often introduced to a model in order to soften deterministic predictions and resolve dichotomies, i.e., to allow for proportional and gradient categories with fuzzy boundaries and to include tendencies and potentialities beyond strictly rule-based processes. However, for the lexicon, the mathematical repercussions of interpreting relative corpus frequencies as probabilities have quite the opposite effect, as they integrate the entire lexicon into a single coherent, mutually interdependent, and globally predictive system.¹¹ This results in more specific predictions to lexical production in context than may be philosophically intended:

- **Generalizability:** in frequentist statistics, probabilities are an objective property of the elements of a stochastic system to approximate expected values over time, i.e., in infinite repetitions of an experiment.¹² The objectivity of that property is often tied to tangible and measurable physical properties. For instance, fair dice are expected to roll an equal number of times onto each of their sides, if – or rather: *because* – those sides are the same size and shape. This means that words, if their frequency is interpreted as probability, have an objective reason to occur at particular rates independently of the situational parameters (as long as the stochastic system remains intact).
- **Stationarity:** since probabilities are inherent features of the variable they describe, they must remain the same across time and space, for as long as they belong to the same stochastic system. The interpretation of relative corpus frequency of lexemes as probability strictly entails that elements will recur, within a limited degree of variance/noise, at approximately the same rates and in the same general distributions in comparable samples. Statistical inference is generally a tool to test the predictable recurrence of stable distributional assumptions. If a system is not stationary, i.e., if today’s corpus frequency says nothing about tomorrow’s, samples are not comparable and any metrics derived from the sample are meaningless for the inference from one corpus to the next or from corpus to language (this does not mean that certain metrics could not still be meaningful for description or analysis). Obviously, language is overall not stationary, as it continues to change. However, the statistical paradigm strictly presupposes stationarity of the frame of reference it chooses, e.g., a collection of texts from a particular time, genre, type of elicitation or other class, to remain interpretable.

¹¹ This is due to the many exemplars of the lexicon and the particularities of their distributions (Shadrova 2021, 2022), as well as the fact that words carry very specific meaning compared to other linguistic structures – predicting one out of two or three similar structural options is philosophically different from predicting one out of 200,000 very diverse lexemes.

¹² Bayesian statistics is only beginning to find attention in corpus linguistics (e.g., Özsoy and Blum 2023). In Bayesian statistics, probabilities correspond to commitments towards subjective distributional beliefs. Those beliefs may be strengthened or weakened through observations of empirical distributions. While this is a more subjective definition, it still reflects a commitment to quantitative predictability of a system from past observations.

- Systemhood: since probabilities are defined as proportions of occurrence of variables within a system, an interpretation of corpus frequency as probability entails the systemhood of all words in a corpus. Similarly, assigning probabilities to lexemes in a speaker's lexicon entails systemhood of their entire lexicon. Words cannot have probabilities independently of all other words, they must be part of a distribution of probabilities that sums up to 1 for the totality of all realizations. If the relative frequencies of words in a corpus are interpreted as probabilities, this means all of the words in the corpus share the same space of potential realization, or simply put, they all compete with one another.¹³
- Mutual (in-)dependence: the interpretation of the rates of co-occurrence of words with other words, constructions, genres, etc. as conditional probability entails a bidirectionality by which the other words, constructions, genres, etc. also predict the word itself. This is due to Bayes' theorem of conditional probability, which states that two variables are either mutually independent or mutually dependent. Thus, words and constructions, genres, etc. in this model are predictive of *each other*.¹⁴ It is mathematically impossible for a lexeme to predict a construction, register, genre, or other linguistic situation under conditional probability without the construction, genre, register, or other linguistic situation also probabilistically predicting the occurrence of the lexeme. This would require other mathematical techniques, such as Markov Chain Modeling, which carry other entailments and are not currently common in linguistic research.

To accept evidence from the metrics mentioned earlier in this section is to accept these entailments for lexical data. This cannot be avoided with a limitation of scope, as metrics would become ill-defined and any analyses based on them invalid if assumptions were not approximated to an acceptable degree.¹⁵ Or, to simplify, within the bounds of some expectable variance/noise, accepting this methodological tradition means that, upon knowing the composition of a corpus, I should be able to predict how many of which lexemes will on average occur.

2.3.1 Practical consequences for the model of entrenchment

These entailments may be of abstract concern as long as they affect purely lexicographic work. However, corpus studies in usage-based linguistics are frequently used with evidential purpose, i.e., to substantiate truth-value hypotheses towards the cognitive and communicative properties of language in individual speakers – this is the case for all corpus-based language acquisition and multilingualism research, for instance. In usage-based theory, this connection is primarily established with the concept of entrenchment, which

¹³ This is the 'bag of words'-problem, which creates major problems for the validity of corpus-wide probability-based metrics. It is obviously an incorrect statement that any word could be realized in any place in a text or corpus (see also Kilgariff 2005, Koplenig 2017). Rather, word production is constrained by syntactic class (the definite article in English is followed by an adjective, an adverb, or a noun, but never an inflected verb) and semantic intention, among other factors. Thus, the entire probability distribution changes for any next word in a text – a feature that is well known and leveraged in computational linguistics for probabilistic taggers and parsers. This means that the lexicon as represented in corpora is already known to be path-dependent, i.e., *non-ergodic* (Dębowski 2018, Medaglia et al. 2011, Weigel et al. 2011). Accepting corpus-wide relative frequency as probability in what is known to be a non-ergodic space is not a logical compromise, but fundamentally epistemologically flawed, as non-ergodicity breaks the connection between the corpus sample and the linguistic abstraction both mathematically and conceptually. Mathematically, it breaks the assumptions of the central limit theorem (CLT), which states that any sufficiently large sample via zeta-transformation can be interpreted as representative of a distribution of the abstract stochastic system iff (if and only if) the system is ergodic and stationary. If those assumptions are not met, there is no mathematical connection between sample and probability, as probabilities are abstract features that can only be inferred, not directly measured. All probability-based concepts, such as significance or effect size, are then ungrounded. Conceptually, it means that only corpora of identical composition and only words within identical slots, i.e., data that represent identical paths, could be compared, provided those paths themselves meet the assumptions of the CLT.

¹⁴ Bayes' theorem: $A|B = P(B|A) \cdot P(B) / P(A)$, in words: the conditional probability of A (construction) given B (word) equals the conditional probability of B (word) given A (construction) multiplied with the probability of B (word) and divided by the probability of A (construction).

¹⁵ General concerns and doubts around stationarity and ergodicity of lexical distributions in corpus data have lately been raised (Holdaway and Piantadosi 2022, Williams et al. 2015, Piantadosi 2014, Aitchison et al. 2016, Koplenig 2017, Dębowski 2018, Shadrova 2022), although primarily from statistical, not from linguistically-functional considerations.

models a speaker's linguistic representations as neural pathway reflections of input. For this, it presumes the formation of a connection between a situational context and a word in the speaker's mind with every encounter of the word in use, as well as connecting branches to co-activated concepts such as the different layers of constructions the word appears in (Balog 2023, Blumenthal-Dramé 2013, Hilpert and Diessel 2017, Stefanowitsch and Flach 2017). Over time, patterns from language use are understood to grow into a probabilistic meta-structure of coactivation that translates more or less directly to a speaker's lexical production: "repetition of linguistic material is indicative of structure: repeatable structures are evidence for the units of linguistic cognition" (Reitter 2008, 14).¹⁶ At this point, the mathematical entailments and variable interpretations of the linguistic arguments create a field of tension between individual variability and generalized stationarity. Let us first clarify different ways in which a multi-agent (multi-speaker) system can emerge as overall stationary:

- Lexical distributions could be self-similar, i.e., occur in the same shape regardless of scale. In that case, the same expected values would be approximated by individual speakers, smaller and larger groups of speakers, and in corpora. This view provides most validity for corpus-based inference to individual behavior as well as stochastic acquisition from input. Probability in this view is attached to words and structures in the language and mapped to minds through exposure to the language, thus providing full transitivity across levels of abstraction: frequency patterns and phraseological conventionalization spread across the language community via – literally through – the minds of its speakers, as entrenchment in speaker minds is modeled to arise from the structure they pick up, i.e., somebody else's output, which has arisen from their input, which is the cumulative output of their interlocutors, etc. This is the original view of entrenchment introduced by Langacker (1987) and corresponds to the practice of correlating experimental results with corpus frequencies without discussion of underlying variability (see Stefanowitsch and Flach 2017 and references therein). This view implies high lexical invariability in context, not only causally (because everybody picks up the same distributions), but also functionally: if speakers passed on slightly divergent frequency distributions, each construction would emerge with different distributional niches and stationarity across levels would be lost. This is particularly true for rare phenomena, which should be more prone to change, as there is less evidence to stabilize their distributions; and for idiosyncratic coselectional preferences that appear not to rely on semantic analysis, which should destabilize easily in an inter-individually variable space.
- Philosophically obvious, but mathematically unaccounted for, is the fact that distributions from thematically wide-ranging corpora will never be reproduced in speaker productions that are not equally wide-ranging. Thus, full self-similarity of the entire system in local production is impossible – the system must be path-dependent (Dębowski 2018, 2020). However, within those thematic, etc. constraints, lexical distributions between speakers and corpora could still converge if expected values were stationarily anchored in situational frames rather than globally. Overall stationarity would still only arise if the situational frames also occurred at stable rates. In this case, corpus composition in terms of situational parameters would determine the comparability to other samples (as is the premise of task-based corpus elicitations) and words could no longer be claimed to possess particular associations outside of those frames. This would pose problems to lexicographic word association measures derived from corpora that are not controlled for semantic or situational frames.
- Probabilities may also be anchored in each speaker's lexicon, and speakers individually or group-wise might approximate expected values in stable distributions that diverge from one another. This appears to be the view on entrenchment as a fluid and idiosyncratic structure proposed by Schmid (2015, 2020). Transitivity of probability would permeate locally between speakers and would be limited to speaker communities (as suggested in the earlier mentioned concept of meso-synchronization by Schmidt and Herrgen (2011)). Corpus averages would then not generally be representative of individual speakers, and averages to serve as proxies

¹⁶ In spite of its uncontested prominence in the usage-based framework, the notion of entrenchment does not appear to exist outside of cognitive linguistics (see also Stefanowitsch and Flach 2017). From the perspective of learning psychology or neurocognitive studies, it is not entirely clear what it adds to the otherwise existing notions of associative learning, context/contextual/contextualized memorization, long-term priming, operant conditioning (see Sturdy and Nicoladis 2017), or simply learning.

for psycholinguistic concepts would have to be derived from speaker-individual corpora. This scenario would also mathematically imply that speakers or speaker groups talk at precise quantitative rates (within limited variance) about the same things, and that those things differ between speakers. The limitation of the stochastic system to a local space in this view further cuts the stochastic connection between lexicographic objects and speakers, raising questions to the nature of those lexicographic objects: after all, what is the ontological status of an association *in the language* if it is not also present for a significant number of its speakers?

How large and variable the set of predictable words would be per situation or speaker group also depends on the functional interpretation of stochastic prediction in language production:

- Is prediction globally optimized, i.e., is preference always given to the most frequent word in production? This has been suggested for naturalistic language production, for instance, by Hulstijn (2024) (and is also a common algorithmic feature of Large Language Models). In that case, we should find converging lexical sets by topic or situational frame. Those sets should be small per approximate meaning, i.e., avoid the realization of near-synonyms, especially within the same production. Rare words would then be limited to semantic niches, as they would be commonly overridden by hyperonyms and semantically lighter lexemes. Speakers would be expected to be highly invariant within and between productions.
- Is prediction optimized towards semantic precision rather than by frequency? Then, we would find many separate subdistributions, i.e., small sets of lexemes for each semantic subset (paths). This is most in line with the principle of no synonymy. Overlap between the different semantic niches should not generally happen at least in individual speakers, and near-synonyms should not compete for the same slots except for ambiguous contexts.
- Alternatively, lexical production may be locally and temporarily semantically optimized. Global frequency would then play a negligible role. Rather, we would find globally unstable lexical distributions from instant category composition in accordance with situational needs (like in Lakoff's (1987) example of the category of potential birthday gifts for a particular friend). In that case, local variability could be high or low depending on the context. However, global stationarity would become unlikely to ever be approximated, as not only the frequency of occurrence of elements, but also their categorial distributions and their frames of references would keep changing, raising more questions to the ontological status of lexicographically definable semantic niches. A system like this would require a shift of the epistemic focus from form-based calculations to linguistic abstractions. For instance, even if word exemplars did not recur at predictable rates, semantic or pragmatic categories still might. If they did not, non-probabilistic approaches to quantification would be required, such as graph-based metrics (Citraro et al. 2023, Stella et al. 2017, Wills and Meyer 2020, Shadrova 2020, 2024). Cognitively, in a dynamic system in which all parameters – categories, distributions, and frequencies – keep changing between situations, speakers, and communities, global stochastic prediction from past events would not result in high accuracy. This is in line with the fact that local unpredictability is one of the defining features of CDSs. Then, cognitive dissonance should arise for speakers whose precise predictions keep failing, which would seem to add cognitive effort rather than reducing it. In such a landscape, a fuzzy heuristic would be more functional than a precise mapping of semantic niches to neatly defined proportions of co-occurrence.

A note on the problem these implications create at the scale of the lexicon: probability is a direct correlate of the number of choices vs chances of realization. If there are only two verbs to refer to an activity, and one is much more frequent than the other, five or ten situations of exposure might suffice to establish a clear frequency distribution. However, most words in corpora are rare, and even near-synonyms occur in large sets. Take, for example, this incomplete list of verbs for theft: *steal, nick, snatch, filch, thief, grab, rob, pinch, pilfer, shoplift, rustle*.¹⁷ Although they share core semantic aspects, they are not fully synonymous. If any overlap in situated contexts is expected, but stochastic, i.e., if speakers were free to choose between those verbs in the same situation, but not at the same rate, establishing their entire distribution in high resolution

¹⁷ See Dux (2016) for a frame-semantic analysis of some of those verbs and their argument structure preferences.

would take hundreds of realizations. This is unlikely to happen within an overseeable time unless one happens to live in a particular social context that requires frequent communication around stealing, nicking, snatching, filching, pinching, pilfering, and so on.

Even though exposure to a variety of niche meanings may have increased in recent decades with access to mass media, the stabilization of such distributions through contextual repetition at scale for the entire lexicon under locally highly variable conditions appears to be implausible.¹⁸ This depends, to some degree, on the estimation of the amount of recurrence required for the quantitative properties of associative connectivity to not be significantly changed, forgotten, or overwritten (which would create problems for stationarity, generalizability, and predictability). According to estimations by Brysbaert et al. (2016), individual L1 speakers of English know on average 42,000 lexemes at 20 years of age (ranging from 27,000 to 52,000 for the lower and upper 5%), and on average learn another 6,000 – or a new lexeme every other day – until the age of 60. Let us simplify by assuming that 20% of those lexemes suffice to navigate 80% of our everyday situations – this is a common simplification of the Pareto distribution, the continuous variant of the Zipf distribution. In that case, 33,600 lexemes would have to be reinstantiated and their distributional records maintained in 20% of the linguistic interaction time of a young speaker. At the highest density, if we presume that speakers interact linguistically for a maximum of 16 hours a day including inner monologue, 20% would amount to 3.2 hours. Assuming conservatively that it suffices to reuse a lexeme once a year to ensure stability of its distributional properties in memory, this leaves $33,600/365 = 92$ rare words every day, or a new rare word roughly every 10 minutes, for a lifetime. As speakers would not iterate through their vocabulary one-by-one and as lexemes occur in thematic clusters, not randomly (see also Egbert et al. (2020), Burch et al. (2017)), this is obviously an unrealistic simplification. However, breaking the numbers down in this way illustrates the high density that arises from limitations of time and chance of occurrence vs the large numbers of exemplars in the lexicon. The example may appear absurd, but in reality represents only a conservative estimation of the repetition rates required to *reactivate* each lexeme. To achieve higher resolution, i.e., to reliably induce the comparative frequency distributions of the elements of a semantic class, especially within the lower frequency band and at partial semantic overlap, occurrence at higher magnitudes would be required. In naturalistic dialogue, this would also require repeated exposure to the events those rare lexemes are associated with; and at each re-instantiation of events, the most frequent form would compete for realization with several rare ones.

It appears as though a view of the lexicon as one coherent system of stochastically learned or defined semantic niches cannot avoid this practical contradiction: high distributional precision requires many identically mapped instantiations, but more instantiations give rise to more potential variability if not *a priori* constrained by some (phraseological?) force – which itself disturbs the randomness of the stochastic model. This can easily be resolved by dropping the stochastic interpretation of conventionalization. If conventionalization was not stochastic, but primarily semiotically-functional and rooted in salience at instantiation, constant reinforcement and reinstantiation would not appear necessary. In that case, lexical forms would not compete with one another in a system of probabilities relative to the entire lexicon or situational frames, but coexist as an open, ever-growing, and fuzzily related set. In that case, lexical variability in context would be unproblematic and in fact expected. However, global, probability-based corpus metrics of lexical frequencies would cease their mathematical validity as well as their evidential scope, and functional arguments towards cognitive or communicative advantages by global limitation of choice would fail.¹⁹

To summarize, the mathematics of metrics derived from global corpus frequencies entail high degrees of predictable ‘sameness in context’ of lexical production – and the uncontested dominance of such metrics for

¹⁸ Similar criticism has recently been raised with respect to stochastic lexical acquisition in childhood (see Unger et al. 2023).

¹⁹ There is a chance that the instationary and the stationary could coexist in memory. Recent study on the mouse model shows that memories may generally be encoded in several copies, some more stable, others more malleable and temporary (Kveim et al. 2024). If this proves to be a general mammalian cognitive mechanism, it could potentially integrate global stationarity with local instationarity, perhaps also explaining locally partial overlap of the mappings between lexical forms and situational meaning with globally separate lexicographic complexes. This could conceptually solve some of the apparent contradictions in current lexicology, although a precise model of both aspects and their interaction would still be required. For this to be properly integrated into linguistic modeling, research on general cognition and memory would have to be considered at much higher resolution than is current practice in the field.

lexicological evidence implies if not theoretical commitment to these entailments, then at least their practical acceptance. This is especially true for statistical techniques that directly model a probabilistic correlation between word distributions and larger units of meaning, such as topic or genre. Whether the realized lexical sets in context are small and populated with semantically light prototypes, large and diverse, or consist of only a few near-synonymous lexemes, depends on the interpretation of concepts such as predictive processing and entrenchment. In either case, they should be convergent and predictable by situational, semantic, or speaker factors. However, the scale of the lexicon and its functional diversity under the stochastic entailments create apparent implausibilities for linguistic theory and inconsistencies between approaches that are commonly viewed as complementary or convergent. To address these contradictions with a better empirical understanding, we shall now proceed to the corpus analysis.

3 Data and hypotheses

3.1 Data

The present analysis is based on a total of five corpora that were chosen from the currently available corpora of German composed from productions of homogeneous speaker groups in controlled elicitations:

- 1–2 Two corpora of argumentative essays, Kobalt (Zinsmeister et al. 2012) and Falko (Reznicek et al. 2012);
- 3–4 Two dialogue corpora, BeMaTaC (Berlin Map Task Corpus, Sauer and Lüdeling 2016, map descriptions) and BeDiaCo (Berlin Dialogue Corpus, Belz and Mooshammer 2023, discussions of differences between pictures);
- 5 The RUEG corpus (corpus of the *Research Unit Emerging Grammars in Language Contact Situations*, Lüdeling et al. 2024, Wiese 2020, witness reports in four settings – formal written, formal spoken, informal written, informal spoken). Only the RUEG corpus contains parallel elicitations in English and Russian, which are also analyzed in this study.

Table 1 provides an overview of the corpora. Type and token numbers are provided in Table 2 in Section 4.1. All corpora are task-based, i.e., elicited and compiled under strictly controlled conditions. Written prompts are provided in Table 1. Examples of prompt images are provided in Figures 1 and 2. All corpora are small to medium-sized and deeply annotated. Normalizations and annotations have been performed and/or corrected manually in the context of the respective corpus projects (references above and in Table 1). Some corrections have been added by me, and all errors remain my own. All corpora are freely available for download and analysis and can be expanded and enriched with further annotations (see under the respective references).

Some of the corpora were originally designed for L1/L2 or mono- and bilingual contrastive studies, but for this study, only data from monolingually raised L1 speakers were used.²⁰ All participants were residents of a country where the respective language represents the majority societal language at the time of elicitation (Germany, Russia, the US). Speaker groups are homogeneous with respect to age, region, and urbanicity in all corpora. For the essay corpora, speaker groups are also homogeneous with respect to socio-economic status (SES) and education background (Table 1). For instance, speakers in Kobalt are all 12th grade students of the same school in an affluent part of Berlin, all taking the same German class with the same teacher. In that particular case, it can be assumed that speakers also act as mutual interlocutors outside the elicitation, i.e., share a community space of lexical usage or a variety.²¹

²⁰ This is not to imply a model of a ‘true’ or ‘fully’ ‘monolingual native speaker’, as multilingual practices are omnipresent and ubiquitous even in the so-called monolingual speakers (Nurmi et al. 2018, Minnaard and Dembeck 2014, Bunk 2024, Wiese et al. 2022). For pure simplicity, speakers raised in primarily non-multilingual environments will be referred to as monolingual where the contrast is required in this article.

²¹ The German school system is multi-tiered from a young age on and thus stratified to a problematic degree. This has rightfully been the target of intense criticism from educational researchers and social organizations for decades, but for the purposes of linguistic comparison creates incidental access to a highly homogenized speaker community.

Table 1: Overview of the corpora used in the analysis

Corpus	Language	Speakers	Speaker group	Task	Prompt/material	Conditions
Kobalt (Zinsmeister et al. 2012)	German	20	12th grade students (17–18 years old) from a Berlin high school (medium to high SES)	Argumentative essay (90 minutes in class, no aids, handwritten)	<i>Geht es der Jugend heute besser als früheren Generationen? Diskutieren Sie.</i> ('Do young people do better (have a better life) than previous generations? Discuss.')	—
Falko (Reznicek et al. 2012)	German	15	12th grade high school students (17–18 years old) and undergraduate university students from Berlin	Argumentative essay (90 minutes in class, no aids, typed in a text editor without language assistance)	<i>Der Feminismus hat den Frauen mehr geschadet als genützt. Diskutieren Sie.</i> ('Feminism has done more harm than good to women. Discuss.')	—
BeMaTaC (Sauer and Lüdeling 2016)	German	14	Adult L1 speakers of German recruited through the university (students, acquaintances, colleagues)	Instructor is asked to explain a path they see on a map to an instructee who sees a different map. Participants are unaware of the differences between their maps at first	Two sets of two maps of twelve icons, maps within a set differ by one icon (see Figure 2 for an example)	Two sets of two maps (A and B); two roles (instructor, instructee). Instructees speak considerably less
BeDiaCo (Belz and Mooshammer 2023)	German	16	Adult L1 speakers of German recruited through the university (students, acquaintances, colleagues)	Two speakers are asked to find differences between two pictures, each only sees their own picture.	Two pictures (Figure 1)	Diapix <i>Street 1</i> and <i>Farm 1</i> Baker and Hazan (2011)
RUEG (Lüdeling et al. 2024, Wiese 2020)	German English Russian	48 (Ger) 64 (Rus) 67 (Eng)	Adult and adolescent speakers recruited from schools, universities, workplaces, acquaintances, etc.	Speakers are asked to report the events they witnessed to a) the police (written and oral statement) and b) a friend (WhatsApp message and voice message)	Stimulus video (about 40 s, https://osf.io/szfhd) showing the following events: A man and a woman with a stroller walk along a street, the man is bouncing a ball. The ball rolls into the street. A dog chases after the ball, thereby causing their owner to drop a bag of groceries that she was about to load into her car. In order to avoid hitting the dog, a car driver has to abruptly halt, which causes a second car to hit the first. As the man with the ball helps the woman collect her groceries, the two drivers talk and call the police.	Formal written, informal written, formal spoken, informal spoken. The two formal and two informal conditions were elicited in separate elicitations 2 weeks apart. Elicitors were acting more formal in the formal condition, wearing formal clothes, etc., while acting more casual in the informal condition.

Table 2: Speakers and category types and tokens for all corpora

Corpus	Task	Language	Topic	Speakers	Verb tok.	Verb lex.	Noun tok.	Noun lex.
Falko	Essay	German	Feminism	15	1,006	420	1,742	717
Kobalt	Essay	German	Youth	20	994	422	1,966	747
BeMaTaC	Map task	German	Directions	14	541	114	957	201
BeDiaCo	Description/dialogue	German	Diapix	16	1,067	152	2,382	483
RUEG	Witness report	German	Accident	48	3,382	436	5,713	569
RUEG	Witness report	English	Accident	64	4,101	267	6,210	343
RUEG	Witness report	Russian	Accident	67	4,192	564	5,669	490

Verbs only include lexical verbs, i.e., no modal, auxilliary, copula etc. verbs. For Falko and Kobalt, verbs and nouns that occur in the prompt have been excluded from the analysis. In the other corpora, prompt material was not provided in written form.

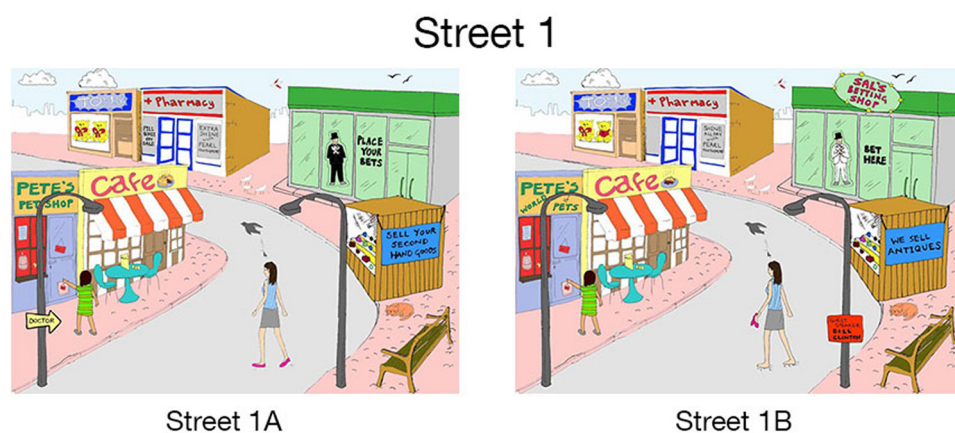


Figure 1: Diapix *Street 1* Baker and Hazan (2011) (reprinted from the public domain with author's permission, available through <https://www.phon.ucl.ac.uk/project/kidLUCID/diapix.php>). Participants are prompted to identify differences between the two pictures, such as the missing sign on the letting shop or the shop keepers different suit colors. Diapixes used in BeDiaCo were translated to German (see Belz and Mooshammer 2023).

3.2 Hypotheses

As virtually no explicit hypotheses towards lexical production in context have been proposed in the literature, they have to be deduced from the structural principles and mechanisms of the approaches discussed in the previous sections. Lexical production is likely shaped to some degree by all of the phenomena reviewed above – conventionalization/distributional features of lexical elements, constructional activity, situational dynamics, and the impact of variationist factors. In order to assess the contribution of each tendency, it is worthwhile to formulate maximal hypotheses for each of the models. Hypotheses are formulated and labeled according to their operationalization based on observable linguistic items in the corpus, not their theoretical status (degree of lexicalization, phraseological weight, etc.).

H1 form-based/holophrastic convergence: Phraseological accounts suggest form-based convergence. If tendencies of between-speaker convergence on the level of surface form dominate, speakers would be expected to use the same words for the expression of similar meaning. They would be more convergent in the elicitation settings requiring description of identical material, i.e., the dialogical and witness report corpora. Speakers should overall avoid derivational word formation if similar meaning can be expressed easily with paraphrase, as newly derived words would always be rarer than frequent words used for

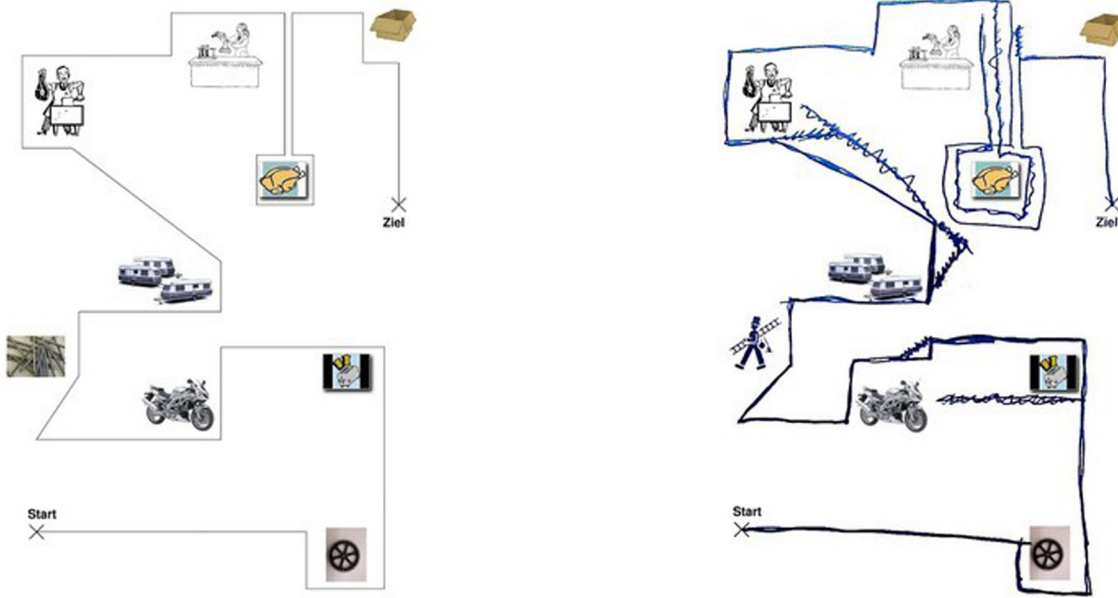


Figure 2: BeMaTaC instructor and instructee maps. Participants are asked to negotiate map directions based on divergent landmarks in their respective maps, such as the missing box at the top right of the instructor map or the pile of nails vs the chimney sweeper to the left of the map (reprinted from the public domain, available through <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/bematac/bematac>).

simple paraphrase.²² Overall, a high degree of overlap between speakers would be expected. As speakers are often encouraged to avoid repetitions of identical lexemes in educational settings, and corpus elicitation might be perceived as educational or testing environments by some, it is possible that speakers consciously add lexemes for variety. This should increase convergence rates, as speakers would then realize a larger proportion of the situationally expected set, allowing for more overlap.

- H2 **distributional niches:** If lexical production is primarily guided by the optimization of semantic precision, speakers would be expected to be most convergent in semantically fine-grained distributional niches. Where speakers choose similar sub-discourses in the essay corpora, and where they choose to describe identical details, they would choose identical words. Near-synonyms should be rare, unless they transmit contrasting details in context.
- H3 **distribution by external factor:** If distributional tendencies by external factors dominate, as is suggested by variationist perspectives, speaker productions would follow hypothesis 1 or 2, but vary intra-individually between elicitation conditions (only applies to the witness report corpus RUEG). For instance, near-synonyms would be distributed by elicitation condition (formal/informal, etc.).²³ Conventionalization may also be tied to speaker communities, i.e., present in highly localized varieties. Therefore, the Kobalt essay corpus, whose speakers form a community outside the elicitation context, should be more convergent compared to the other corpora, especially the other corpus of the same format (Falko). It is also possible

²² Hypotheses for compounding are less clear, as compounds only minimally diverge from their parts and can be conceptualized as non-syntactic chunks. However, at the same time, non-lexicalized compounds are underspecified with respect to the relationship between head and non-head, or noun and attribute (*[the man that][holds the][book]* vs *the book man, the book holder*). If local ambiguity is indeed avoided, non-lexicalized compounds should only occur where their object of reference is very clear. Unpredictable or odd forms should be dispreferred relative to paraphrases featuring higher semantic predictability.

²³ Although variability may also occur for particular linguistic forms that are representative of ongoing language change or reflect language contact phenomena or other systemic ambiguities, most linguistic forms would not be affected by these processes. Instead, they should be distributed according to their existing and stable niches.

that speaker cohorts generally divide into smaller varieties. Then, we should find subgroups of speakers that converge less vs more, i.e., multimodal distributions of lexical overlap.

- H4 **schema-based novelty**: If the constructionist aspect of schema-based productivity dominates in lexical production, speakers should show evidence of constructional activity (e.g., blending, near-synonyms through derivational morphology), especially for the transmission of unusual or idiosyncratic meaning. Paraphrases should not outweigh novel word formation. Attempts to avoid repetition (see H1) and to signify creativity may also result in higher constructional activity.
- H5 **situationally-dynamic variability**: If situational dynamics dominate, speaker productions would lexically diverge. This can be due to differences in interpretation of the situational content, discourse paths or task engagement, and cognitive effects, such as priming/persistence, and would result in idiosyncratic lexical distributions. As word formation activity may occur abundantly as a reflection of cognitive engagement, some overlap with H4 is expected. Speaker productions would vary intra-individually about as much as inter-individually and variability would not primarily occur as register-specific.

Although this study is not of primarily contrastive or typological interest, a note on language specificity seems appropriate. Since German is a morphologically rich language that features many productive native word formation processes, productivity can be expected to be overall higher compared to English, and as high as or lower than in languages featuring even richer morphological productivity, such as Russian. Additionally, German has a well-known graphematic idiosyncrasy in the convention of not splitting compounds, which might yield lower rates of convergence where compounds are plentiful. As the witness report corpus (RUEG) allows for the comparison with parallel elicitations in English and Russian, this will be considered under the following hypothesis:

- H6 **language-specificity**: Speakers converge more in languages that allow fewer grammatical operations on lexemes (esp. word formation). English is thus most convergent between speakers, Russian is least convergent, and German is situated between the two.

4 Results

This section is structured as follows: first, quantitative results are reported for all corpora. This includes lexical overlap between all sets of three and four texts in each subcorpus as well as the ratio of shared lexemes relative to each speaker's total lexemes (types).²⁴ Then, only for the witness report corpus (RUEG), intra-individual variability is analyzed by comparing each speaker's reuse of lexemes, i.e., the lexical overlap between their own productions in the four elicitation settings. The second subsection provides a qualitative discussion of three phenomena that explain high variability in spite of narrow task constraints: (1) variable discourse granularity, (2) referential diversity and word formation productivity, and (3) task interpretation.

4.1 Quantitative analysis

Table 2 provides an overview of type and token numbers for each corpus. Speakers are overall remarkably productive across corpora, even those featuring the most communicatively limited elicitations. For instance, German and Russian speakers in the witness report corpus (RUEG) find around 500 different verbs to describe the events of a 40-s video; and in the map task corpus (BeMaTaC), the 14 speakers presented with two maps containing 12 landmark icons each use a total of 114 different nouns to describe them.

²⁴ The term *type* is ambiguous as it can refer to a linguistic type (a lexeme or an inflected verb) or a graphematic type (a misspelled word, an idiosyncratic, e.g., onomatopoeic, transcription). In this article, only lexemes are considered as types. Their number is computed from manually normalized text and manually corrected lemmatization, i.e., without spelling variants and inflections.

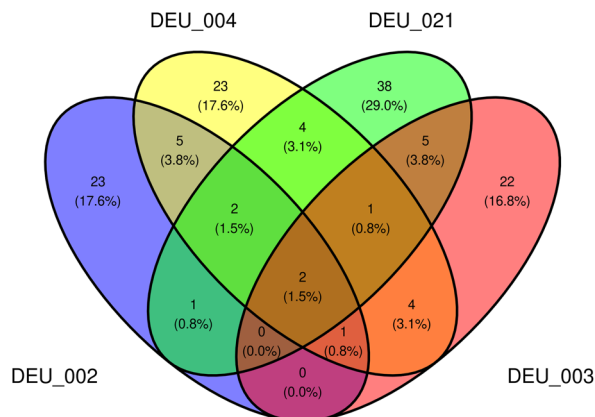


Figure 3: Venn diagram representing the overlap of verb lexemes between four texts in Kobalt (DEU_002, DEU_003, DEU_004, and DEU_021). Two verb lexemes occur in all four texts, 1, 1, 2, and 0 occur in different combinations of three texts, but not all four. 23, 23, 38, and 22 occur only in one of the texts, i.e., they are not shared with either of the other texts in the set. Percentages represent the proportion of the number of lexemes relative to the total number of lexemes in the set, not each speaker's individual set (the 23 lexemes that occur only in DEU_002 make up 17.6% of the total number of verb lexemes in the entire set of four documents). Sets of three texts are equivalent to the overlapping fields of three texts in this diagram.

4.1.1 Inter-individual variability

Between-speaker lexical overlap is not a very commonly used measure in corpus-linguistic research, although a few applications exist (Sinclair and Fernández 2023, Meylan and Gahl 2014, Horton and Spieler 2007).²⁵ Although it abstracts from the lexemes for the comparison between sets, between-speaker overlap requires knowledge of which particular lexemes occur in each text. This makes it a more precise metric of the qualitative similarity of the lexicon between speakers compared to the measures mentioned in Section 2.2, such as between-speaker lexical diversity, sophistication, over- and underuse, etc., which can be computed from only knowing *how many* types, tokens, etc. of each category the corpus contains.

For the inter-individual analysis, between-speaker lexical overlap is computed for each set of three and four texts in the corpus. For example, in a corpus containing texts *doc_1*, *doc_2* ... *doc_n*, the first set of four texts would include documents 1–4, the second documents 1–3 and 5, the third 1–3 and 6, and so on. These three sets would then overlap in three of four texts (1–3). Lexemes are only counted if they occur in all texts of a set. This is visualized for a set of four texts from the Kobalt essay corpus in Figure 3. Only the central field of the Venn diagram is considered in the following analysis for four-text sets, and the four directly adjacent fields for the three-text sets. Only nouns and lexical verbs (no auxiliary, copula, or modal verbs) are considered in the analysis, as they form the most variable, productive, and open lexical classes, and are also most sensitive to content and situational context. Computations are performed separately for verb and noun lexemes and normalized against the total of each category (n overlapping verb lexemes out of m verb lexemes used by a speaker – relative overlap will usually differ between speakers, as three verbs may make up 5% of one production and 10% of another; Figure 4).

Although the corpora are rather diverse with respect to the genre, communicative situation, and even language they elicit, the results for lexical overlap between speakers are remarkably similar and remarkably low (Figures 5 and 6. Violin plots for all distributions have been moved to the appendix for easier reading). The results can be summarized as follows (Figure 7):

- (1) All lexemes are shared at a low rate with medians of 0–2 shared verbs 0–6 shared nouns across conditions and languages. Verbs are overall barely shared.

²⁵ Elsewhere, the term is also used to refer to semantic overlap in the mental lexicon, see, for instance, Benitez and Li (2023), Kalashnikova et al. (2019), and Diesendruck and Markson (2001). This is not the intended meaning in this study.

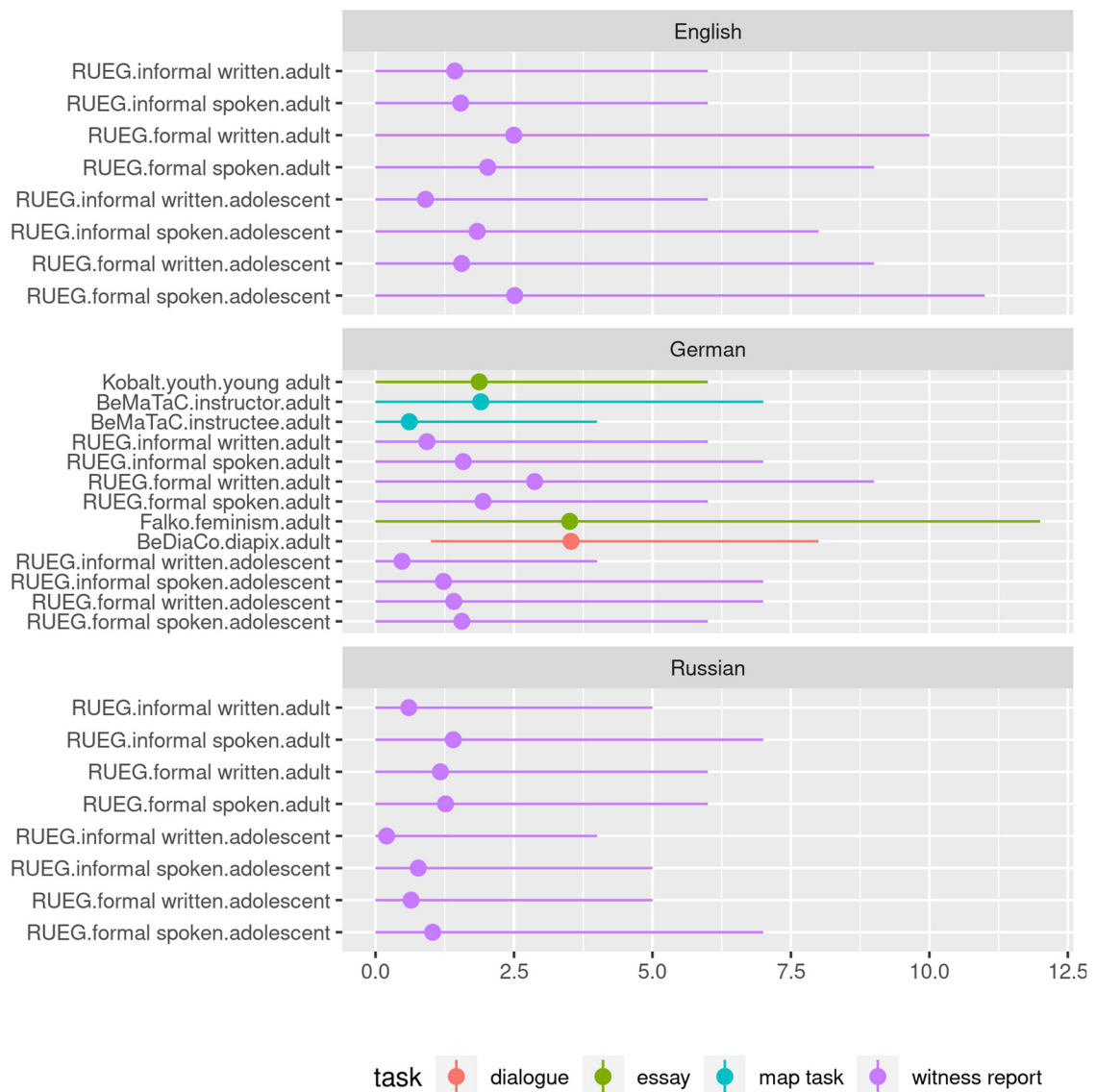


Figure 4: Summary statistics of verb lexeme overlap for sets of three texts from all corpora (mean/minimum/maximum, only matching conditions).

- (2) The ratio of shared lexemes per speaker (lexemes they share out of all lexemes they use) is surprisingly stable across corpora per condition (Figure 8): for the more constrained corpora, it is at median 21% for nouns and 8% of verbs, and for the essay corpora, it is at about 5% for both nouns and verbs. An exception is the diapix corpus, where verbs and nouns are shared to the same degree. This is likely due to the descriptive character of verbs therein, as a person in the picture would be described with reference to their performed activity.
- (3) Even for those maximally similar sets, the ratio of shared lexemes lies below 25% for the essay corpora and below 50% in most of the dialogue conditions. Overlap ratios in RUEG have longer upwards tails (some reaching higher ratios), which is largely an artifact of text length variability (higher overlap between very short and rather long productions).²⁶

²⁶ For instance, the full production of one speaker (USmo69M) from the English subcorpus is: “so what had happened was a man was walking by with a ball and dropped it. There was a woman standing nearby with a dog and the dog must have been excited by the

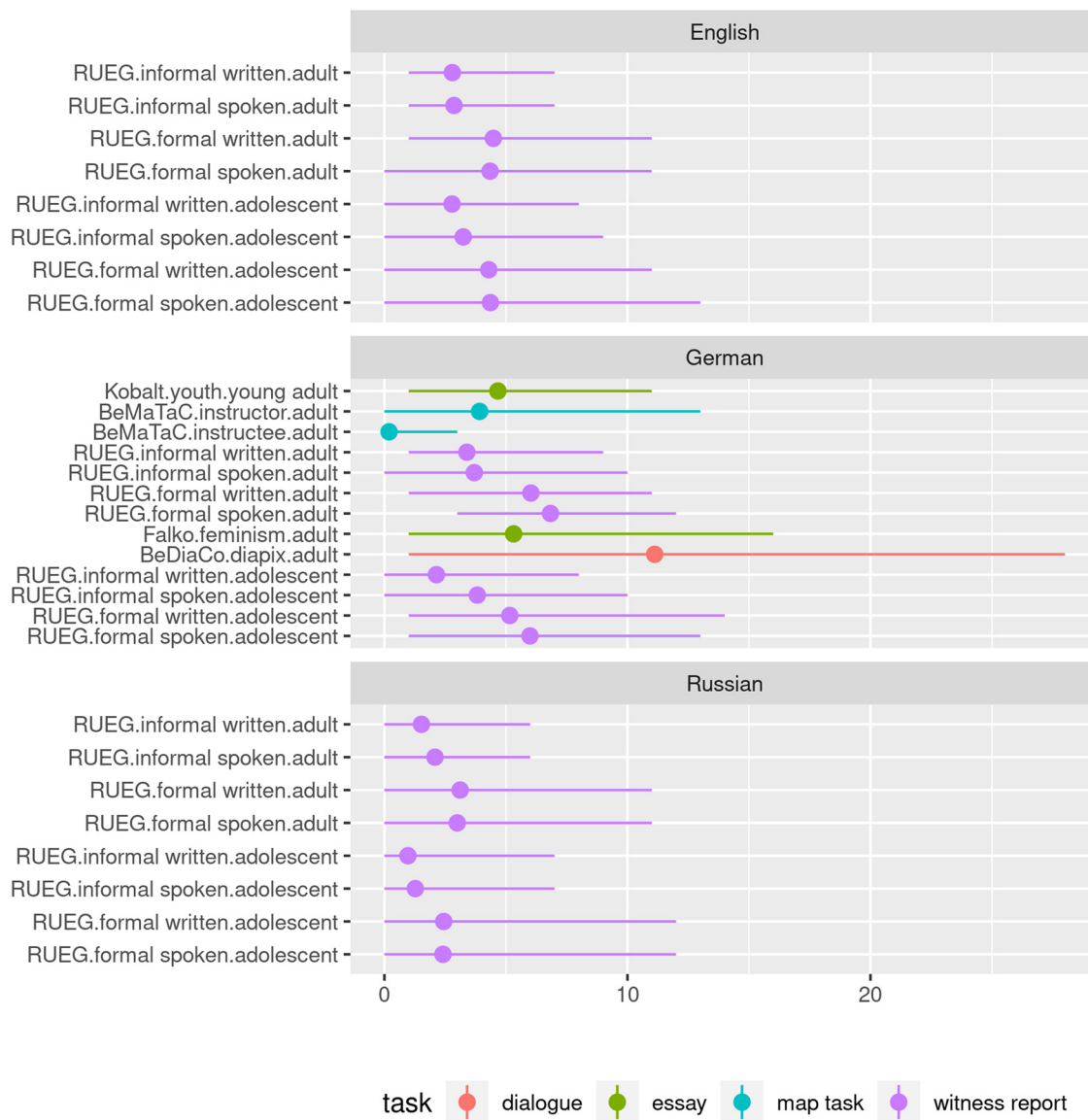


Figure 5: Summary statistics of noun lexeme overlap for sets of three texts from all corpora (mean/minimum/maximum, only matching conditions).

- (4) Four-text sets have shorter tails, i.e., fewer sets with above-average overlapping lexemes, compared to three-text sets. This suggests that there is an upper bound to the qualitative similarity at three speakers even within narrowly constrained communicative situations, and very low probability of more speakers converging on the same set beyond that number. There is no evidence of a clustering of groups by similarity in a multimodal distribution. It does not appear as though speakers can be sorted into communities by lexical similarity based on choice of subdiscourse or conventionalization. Speakers in Kobalt do not converge more, but in fact less, than in Falko, in spite of uniquely forming a speaker community outside the elicitation among the corpora presented. Even in the picture description dialogue, speakers do

ball. The dog ran out and chased the ball in front of two cars. The first car stopped suddenly and the second car behind it hit the first car," yielding the noun lexemes *man, ball, woman, dog, car*. For this speaker, five shared lexemes amount to a ratio of 1. Most speakers produce many more noun lexemes in the same condition (median 14, maximum 55), of which only some occur in the other productions.

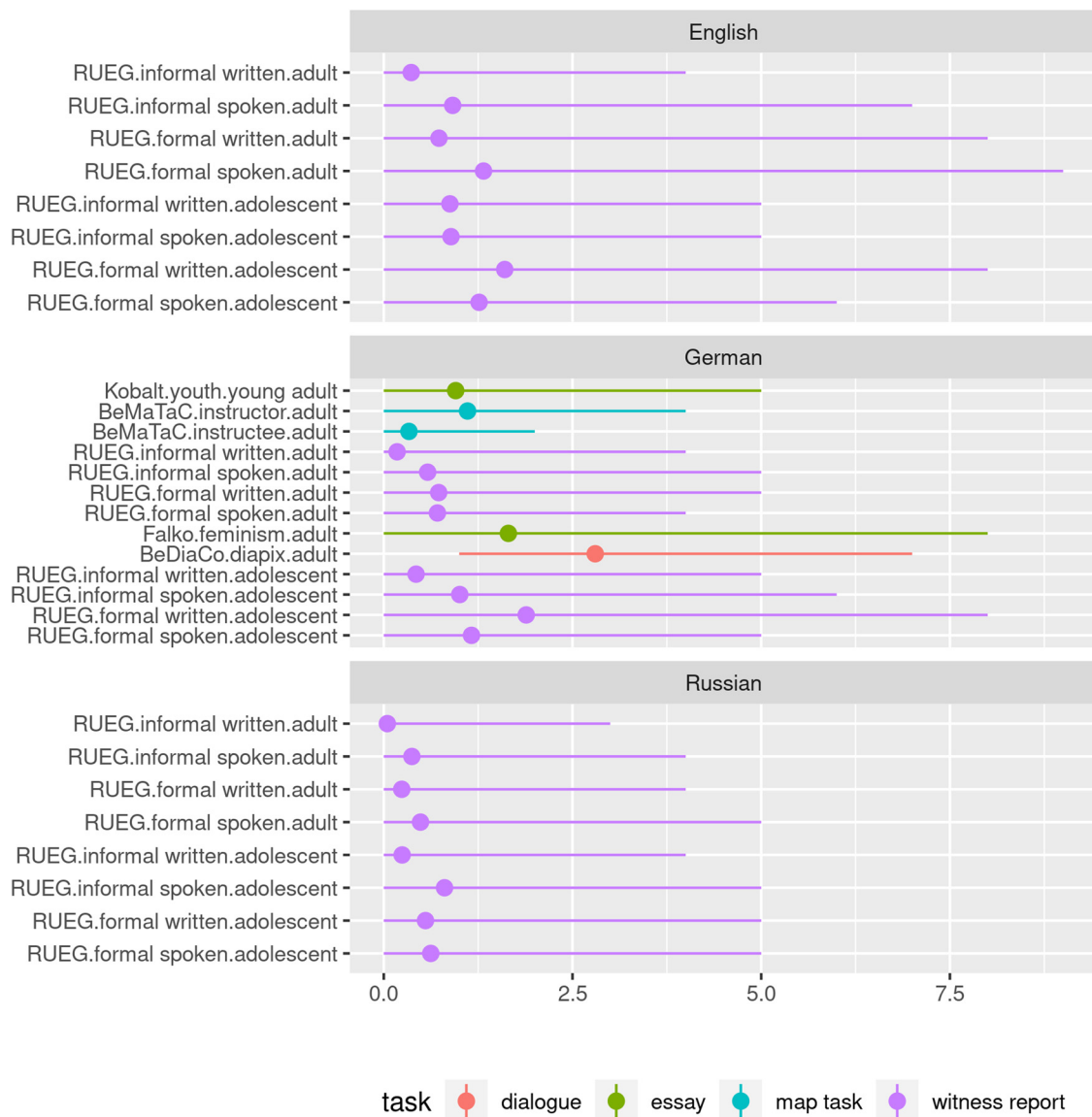


Figure 6: Summary statistics of verb lexeme overlap for sets of four texts from all corpora (mean/minimum/maximum, only matching conditions).

not share more than 25% of noun lexemes per dyad. While a 100% overlap would not be expected, since it is unnecessary for both speakers in a dyad to refer to the same elements in dialogue (one speaker can expand on the previous utterance or confirm it, etc.), it is noteworthy that the ratio of shared lexemes in dialogue is barely higher compared to the ratio of shared lexemes of other, non-interlocuting speakers in the same condition.

4.1.2 Intra-individual variability

As the witness report corpus (RUEG) was elicited in four conditions for each speaker, it allows for an intra-individual analysis (Figure 9). The four contributions create a 2×2 matrix of formality and modality (formal/informal spoken/written). Although speakers might differentiate between register-specific vocabulary,

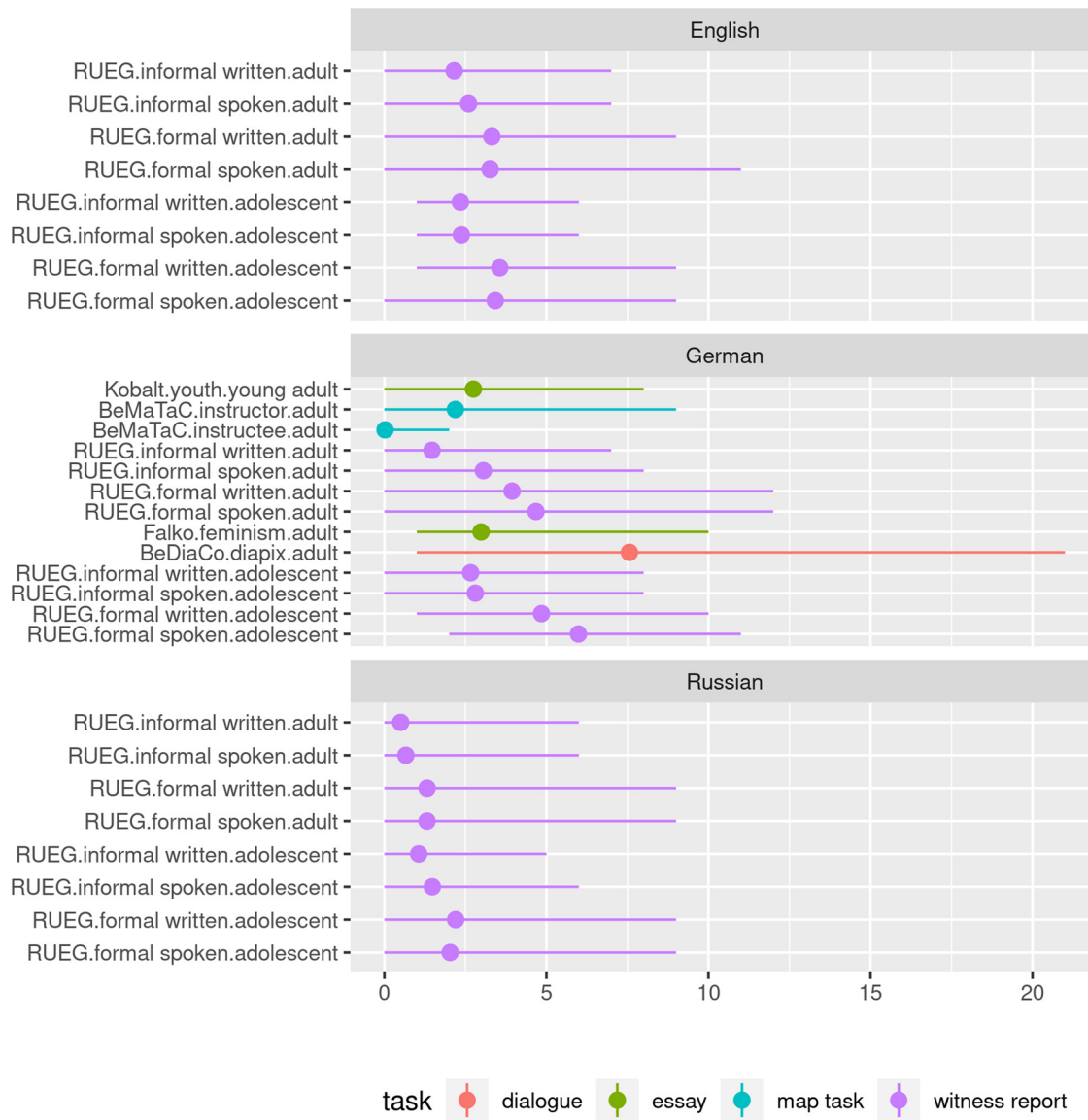


Figure 7: Summary statistics of verb lexeme overlap for sets of four texts from all corpora (mean/min./max., only matching conditions).

especially between the two formal vs the two informal settings, the descriptive content remains largely identical. High within-speaker lexical overlap for lexemes that are not highly register-specific would hence be expected. Figures 10 and 11 show that this is not generally the case. Most lexemes occur as speaker-specific hapaxes. Speakers rarely reuse any verb or noun lexemes in all four settings. As predicted, the data suggest an effect of formality, which is visualized in Figure 12: if speakers use a lexeme in two settings, it is more likely that the two settings match by formality than modality or that they are unmatched.

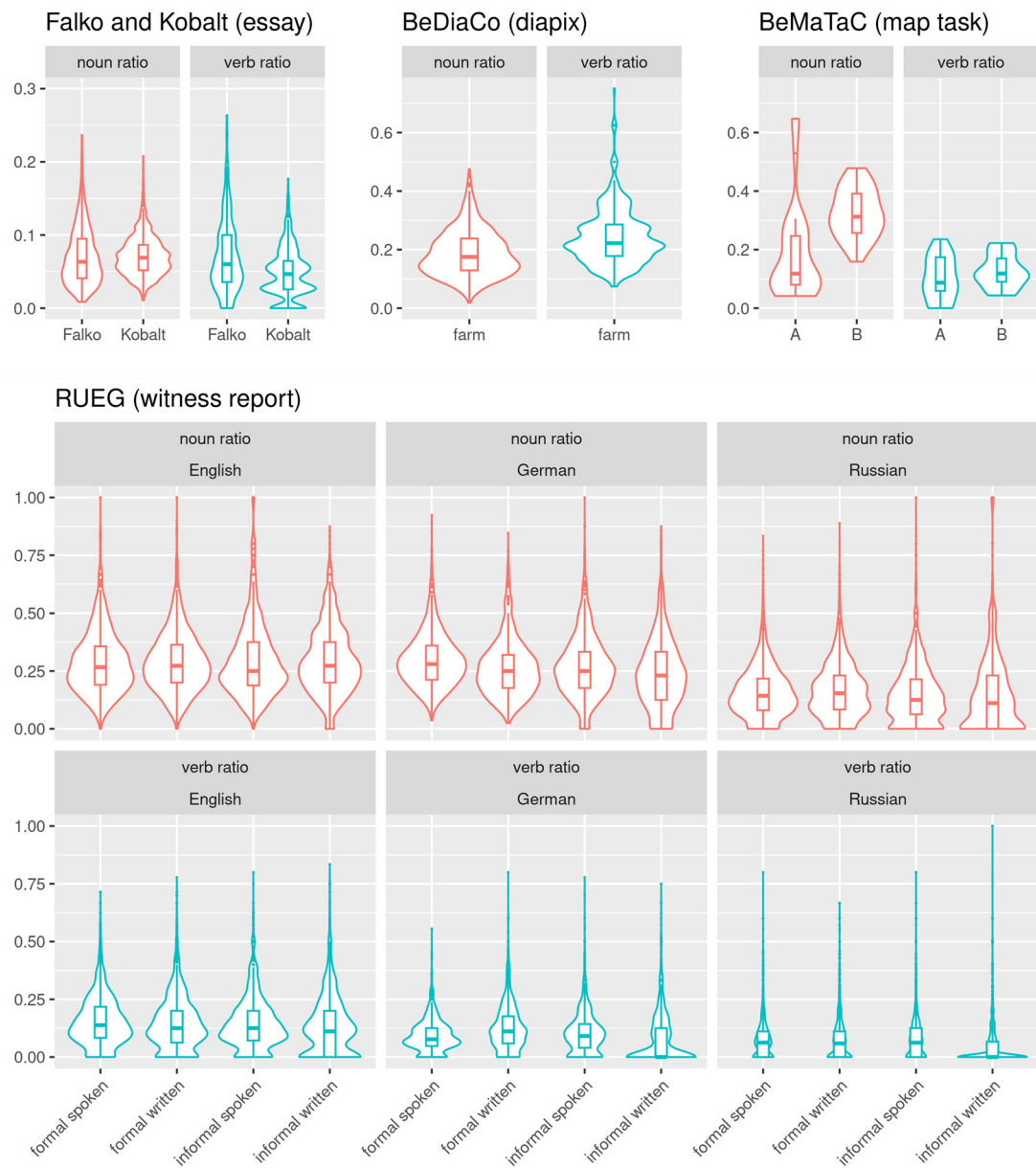


Figure 8: Proportion of inter-individually shared noun and verb lexemes out of all lexemes for verbs and nouns in all analyzed corpora (only sets of three texts, only matching conditions). Please note the different scales in the top row left vs center and right.

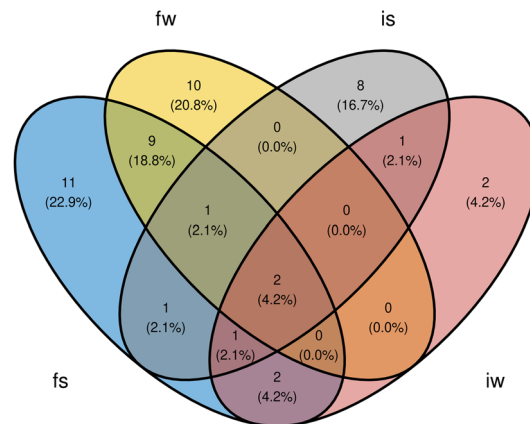


Figure 9: Venn diagram representing overlap between the verb lexemes used in four texts by speaker DEmo50 in the German subcorpus of RUEG (fs = formal spoken, fw = formal written, is = informal spoken, iw = informal written). The speaker uses two of their verb lexemes in all four settings (*kommen* ‘to come’ and *bremsen* ‘to brake’). Eleven verb lexemes only occur in the formal spoken setting, and ten only in the formal written, etc. The speaker has four ways of using the same lexeme in a single setting (the four settings), six ways of using the same lexeme in two settings (same formality, different modality (2); same modality, different formality (2), opposite modality/formality (2)), and four ways of using the same lexeme in three settings (leaving out one setting each).

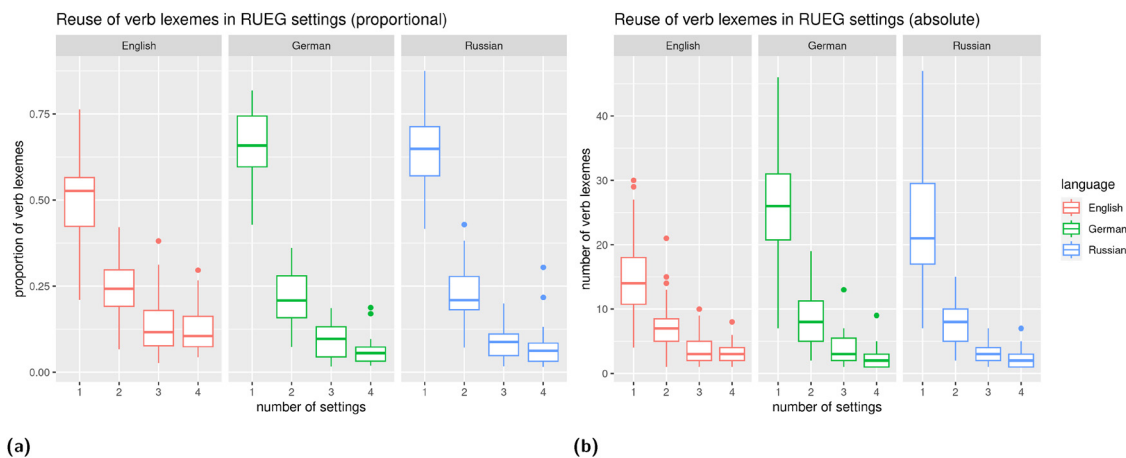


Figure 10: Reuse of verb lexemes by the same speaker in RUEG proportionally (a) and in absolute numbers (b). The x-axis describes the number of settings a lexeme is used in. For example, if the verb *to run* occurs in both the formal written and the formal spoken text of the same speaker, it would count towards the box visualized at 2 on the x-axis. Most verb lexemes occur in only one setting, and very few occur in all four settings for the same speaker (median = 2). The effect is stronger proportionally in English compared to the other two languages mostly because Russian and German are more productive in the verb domain due to the productivity of complex verbs. In absolute numbers, reuse is almost identical between languages.

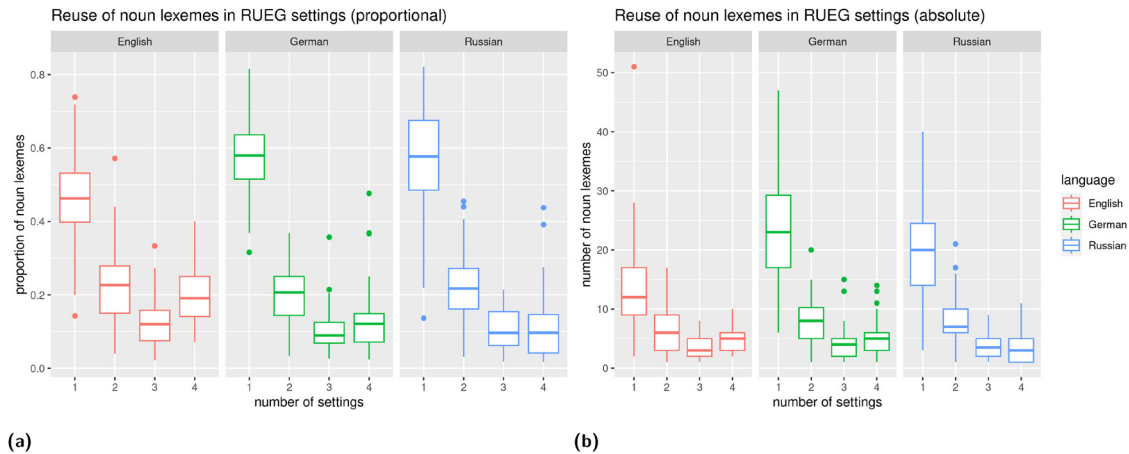


Figure 11: Reuse of noun lexemes by the same speaker in RUEG proportionally (a) and in absolute numbers (b). The x -axis describes the number of settings a lexeme is used in. For example, if the noun *ball* occurs in both the formal written and the formal spoken text of the same speaker, it would count towards the box visualized at 2 on the x -axis. Most noun lexemes occur in only one setting. Proportionally, English speakers reuse as many nouns in four settings as they do in two.

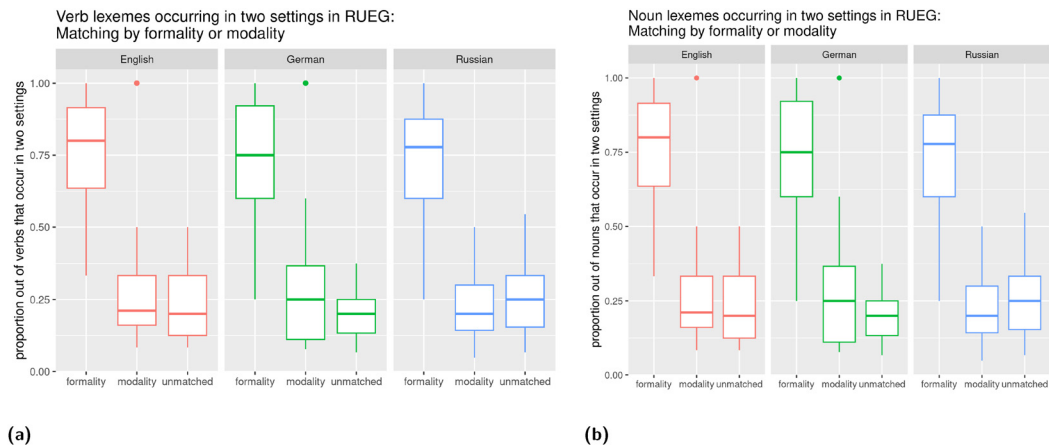


Figure 12: Matching of formality or modality in verb (a) and noun (b) reuse. If speakers use the same lexeme in exactly two settings, those settings are likely to be matched by formality, but not modality. For example, if a speaker uses the verb *to walk* in two settings, it is more likely that these settings are both formal or both informal than it is that they are both written or both spoken, or on opposite ends, such as formal written and informal spoken.

4.2 Qualitative perspectives

This section offers qualitative perspectives on some of the phenomena that create the high degree of variability measured in the data. An exhaustive analysis of the full productivity is impractical to annotate and lies beyond the scope of this study. However, even a rough overview points towards interesting directions for future research. These concern the interface of lexical productivity and discourse granularity, referential diversity, cognitive effects, and task interpretation and interaction with the communicative situation, and will be discussed in that order below.

4.2.1 Granularity of discourse

As briefly mentioned before, speaker productions often cover a wide thematic variety in essay corpora in spite of the intended limitation through elicitation design. In a predictable distribution, this would be an example of

Table 3: Frequency of occurrence of nouns related to education from the five Kobalt texts with the highest lexical overlap and their distribution across all Kobalt texts

Lemma	Translation	001	002	004	005	007	008	009	010	011	012	013	017	018	019	021
Schule	School	1	1	0	2	1	1	1	0	1	1	1	1	0	1	1
Wissen	Knowledge	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Kindergarten	Kindergarten	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0
Bildung	Education	0	0	0	1	2	0	0	2	0	1	0	0	0	3	5
Forschung	Research	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
Schulbesuch	School attendance	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Schulpflicht	Compulsory education	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0
Abitur	High school degree	0	0	0	0	0	4	1	0	0	0	1	0	1	1	0
Gymnasium	<i>Gymnasium</i> school*	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Schülerin	Female student	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Schülerschaft	Student body	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Unterricht	Lesson	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Unterrichtsinhalt	Lesson content	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Studium	College studies	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Instrument	Instrument	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Musik	Music	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Total lexemes	1	2	1	6	3	2	8	1	2	2	2	2	1	4	7

The documents containing the highest number of lexemes from the list, 005, 009, and 021, occur with separate clusters of four, six, and three lexemes, respectively. Those clusters do not occur in this combination in any other text and for most of the lexemes are exclusive to the clustering text. Some form small thematic clusters, such as *Musik* and *Instrument* ('music' and 'instrument'), or *Schulbesuch* and *Schulpflicht* ('school attendance' and 'compulsory education'). The table does not include all education-related nouns in Kobalt, only those extracted from the maximally overlapping sets. *German students are divided into a three-tier system beginning in middle school. *Gymnasium* is designed to be the most challenging of the tiers with students graduating high school after 12 or 13 years with a degree that allows them to move on to higher education.

thematic path dependence of lexical choice. At first, speakers may walk the same lexical path, and certain lexemes are more likely to occur, until they diverge into different paths resulting in different lexical subsets. In the corpora at hand, for instance, in responding to whether young people have a better life today compared to previous generations (Kobalt), some speakers write about historical events, living conditions, modern amenities, or the current sociopolitical situation, while others include personal narratives or list their own interests and activities. Where participants converge on a subtopic, such as *education*, they still diverge in the aspects they explore in some depth (Table 3).

Interestingly, even where subdiscourses are shared, they do not appear to entail the use of identical verb lexemes. For instance, in one of the three-speaker sets from Falko on the *feminism*-topic (DEU_005, DEU_009, and DEU_020), the shared nouns *Frau*, *Gesellschaft*, *Karriere*, *Kind*, *Krieg*, *Leben*, *Mutter*, and *Rolle* ('woman', 'society', 'career', 'child', 'war', 'life', 'mother', and 'role') appear to point to a common semantic field, but the shared verbs are semantically very light and generic: *geben*, *kommen*, and *machen* ('to give', 'to come', and 'to make/do').

The prompt, designed for thematic or topic control with the aim of keeping lexical parameters comparable between speakers has not succeeded at eliciting a prototypical lexical distribution. Although speakers converge on the same general discourse, they diverge in their lexical use even within relatively narrow aspects of discourse. This does not appear to be an artifact from small corpus size (since higher average length of production for essays than map task dialogues does not lead to larger overlapping sets), but rather shows that discourse can fractal out into granularity of any scale, offering abundant amounts of paths on any of those levels.²⁷

²⁷ These results are in line with observations from thematics in literature studies and information/library studies, whereby topics or similar thematic constructs do not allow for comprehensive or final taxonomization, but rather emerge as infinitely variable (Shadrova 2021).

4.2.2 Morphological productivity and referential diversity

Divergent subdiscourses naturally entail different lexemes. The low between-speaker lexical overlap observed in the essay corpora might then simply be a cautionary tale for elicitation design, setting a higher bar for the required specificity of thematic prompts for controlling lexical parameters. However, speakers in the dialogue and witness report corpora were prompted with visual material in extremely narrow communicative settings and still only share a small proportion of their lexemes with any two other speakers. How so?

The data allows for the identification of three intertwined phenomena: (1) seemingly random choice from relatively large sets of near-synonymous lexemes, (2) many degrees of freedom in the interpretation of all details of the prompt material and task, and (3) a high degree of productivity in word formation that reflects on-the-spot situational processing and socio-pragmatic interaction. The following summary of observations is rather dense, which is unfortunately often the case with lexical data. For the purposes of this study, the many examples seem appropriate and supportive, as they highlight interesting pathways for future research as well as empirical contradictions to the theoretical models discussed in Section 2. However, to avoid unnecessary overwhelm, the reader is encouraged to freely skip ahead to the next subsection when they find their curiosity satisfied. Table 4 provides a more concise overview of examples of the different phenomena.

- (1) There is abundant use of near-synonyms. Aside from a small number of diatopically or diastratically marked lexemes, this includes many lexemes that do not appear to follow any particular pattern or mark register differentiation. This is particularly the case for verb semantics in the witness report corpus (RUEG), where speakers even individually employ a large set of lexemes in reference to small and seemingly less relevant aspects of the situation. For instance, one of the scenes in the video shows a woman handling groceries by her car. The German monolingual L1 corpus contains at least eight verbs describing the action of putting/stowing/packing groceries into the car: *räumen*, *packen*, *laden*, *legen*, *einräumen*, *einpacken*, *einladen*, *verladen* ('stow/store', 'pack', 'load', 'lay', 'stow into', 'pack into/wrap', 'load into', 'load into/away'). At least nine more lexemes occur in another subcorpus of RUEG featuring bilingual L1 speakers of German from Germany: *tun*, *sortieren*, *einsortieren*, *reinräumen*, *reinpacken*, *reinlegen*, *reintun*, *hineinlegen*, *befüllen* ('to do (put)', 'to sort', 'to sort away/into', 'to stow/load in', 'to pack into', 'to lay/put into', 'to put into', 'to lay into' (variant), 'to fill up (literally: 'be-fill)'). Most of these verbs are directional particle verbs formed from the base lexemes *to put*, *to sort*, *to lay*, *to load*, *to pack*, *to store*, *to fill*, and none appear to be unquestionably marked for register. Almost all are perfectly idiomatic. The most frequent one, *packen*, 'to pack', occurs in about 10% of the productions.

Russian speakers use at least five lexemes to describe the placement of groceries in the trunk (клать, грузить, загружать, упаковывать, укладывать – *klat'*, *grusit'*, *zagruzhat'*, *upakovyvat'*, *ukladyvat'* – 'to put', 'to load', 'to load up', 'to unload (refl.)', 'to pack/wrap', 'to lay out/put away'), but, interestingly, overall interpret the event more commonly as unloading, for which they use at least nine more lexemes: разгружать, выгружаться, выгружать, распаковывать, убирать, разбирать, разобрать, разложить, выкладывать (*razgruzhat'*, *vygruzhat'sa*, *vygruzhat'*, *raspakovyvat'*, *ubirat'*, *razbirat'*, *razobrat'sa*, *razloshit'*, and *vykladyvat'* – 'to unload', 'to unload/load out of', 'to load out', 'to unpack', 'to pack away', 'to sort out' (2x), 'to lay out', and 'to put out').

The English data does not feature as many near-synonymous verbs, likely due to lower morphological richness and word formation productivity. Still, speakers use at least six different lexemes for the same event: *to load*, *to move*, *to place*, *to sort*, *to pack*, and *to put*.

Importantly, individual speakers frequently offer several alternatives within productions of the same condition (e.g., formal written and formal spoken), and even *within the same text*, as illustrated in Figure 13. Near-synonymous noun lexemes without obvious register or diastratic differentiation occur more rarely, but some examples exist, e.g., in *Katzenklappe* and *Katzentür*, 'cat door' or *Tierhandlung* and *Tierfachgeschäft*, 'pet shop' (BeDiaCo - diapix); or English RUEG *junction* and *intersection* or *carriage* and *stroller*. In the Russian subcorpus, we find competing diminutives, as in *парень*, *paren'* – 'guy' → *парнишка*, *паренёк*, *parnishka*, *parenyok*; or *пёс*, *pyos* – '(male) dog' → *псинка*, *пёсик*, *psinka*, *pyosik*. While the use of diminutives itself can imply register differentiation, the use of two different diminutives does not further distinguish between registers.

Table 4: Examples from BeDiaCo (diapix), BeMaTaC (map task), and RUEG (witness report corpus)

Category	Corpus	Examples	Translation
Regional variation	BeDiaCo	Pantoffel vs Hausschuh	Slipper (regional vs general)
	BeMaTaC	Broiler, Brathähnchen, Hühnchen, Hähnchen Möhre, Karotte	Fried chicken (regional variants) Carrot (regional variants)
Near-synonyms with register difference	RUEG-DE	Fahrzeug, Kleinwagen, Auto	Vehicle, car
Word formation with placeholder	BeDiaCo BeMaTaC	Ovalding, Ziehding Reparaturladendings	Oval thing, pull thing repair shop Thingamajig
High specificity	BeMaTaC	Milchbrötchen, Kaiserbrötchen, Metzgergeselle	Milky bread roll, Kaiser bread roll, butcher assistant
High specificity with amalgamation	BeDiaCo	Holzbretterhütte	Wood board hut (more common: Holzhütte, Bretterhütte)
Conceptual blending within agreed-upon common ground	BeMaTaC	Toasterladen, Motorradladen, Hafenbar	Toaster shop, motor bike shop, port bar
Near-synonymous Word formation	BeDiaCo	Spaßtagbanner, Spaßtagschild Katzentür, Katzenklappe Tierhandlung, Tierfachgeschäft Stadtfassade, Stadtsilhouette, Stadtkulisse	Fun day banner, fun day sign Cat door (both variants common) Pet shop (both variants common) City facade (uncommon), city silhouette (skyline, rare), city backdrop
Rare lexemes	BeDiaCo	Sackhüpfrennen, Stadtsilhouette	Sack race, city skyline (<i>Skyline</i> more common in German)
Referential diversity (differences in interpretation/identification of referents)	RUEG-DE	Typ, Mann, Kerl, Dude, Junge, Herr Couple, couple (diminutive), family, man and woman, sir and madam, mother and two sons VW, Renault, Van, Kleinwagen	Type, man, dude, dude (English), boy, Sir Paar, Pärchen, Familie, Mann und Frau, Herr und Dame, Mutter und zwei Söhne Volkswagen, Renault, van, compact car
Referential diversity, possibly unintended	BeDiaCo	Quadrat, Quadrant, Quader	square, quadrant, cuboid
High specificity for spatial orientation	BeMaTaC	Fünfunddreißiggradwinkel, Fünfundvierziggradwinkel, Stückchen	35° angle, 45° angle, tiny bit
Meronymy, hyperonymy, hyponymy	BeDiaCo	Treppenstufe, Treppe	Step, staircase
Plenoastic (redundant semantics)	BeMaTaC BeDiaCo	Wohnwagenbild, Nagelbild, Kornähre Zeituhr	Caravan picture, nail picture, grain oar Time clock
Slip of the tongue	BeDiaCo	Radsteinfeger	? (Rad = wheel, Schornsteinfeger = chimney sweep)
		Schmettersching	Butterfly (should be Schmetterling)

- (2) Speakers exhibit high referential diversity in describing the events in all three visually prompted corpora. For instance, the grocery bags in the RUEG video are variably referred to as *REWE-Tüte*, *Papiertüte*, and *Einkaufstüte* ('REWE bag' – REWE is a local supermarket chain, 'paper bag', 'plastic bag'). These uses are not synonymous – they reflect differences in the interpretation and specification of events and referents. Referential diversity is abundant with all details of the prompt material. For instance, a man playing with a

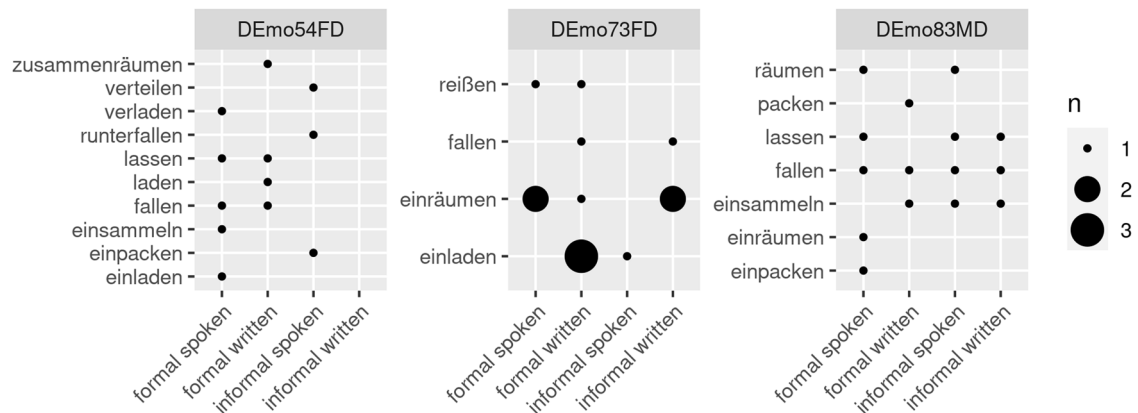


Figure 13: Verbs used with the GROceries subevent by three speakers in RUEG-DE. Verbs sometimes occur as the simplex, i.e., just the stem (*räumen*, ‘to load, to stow’, *packen*, ‘to pack’), sometimes as complex verbs (*einräumen*, *einpacken*, ‘to load/pack into, to stow away’) within the same text; the groceries can be described as being dropped (*fallen lassen*) or as falling (*fallen*) by the same speaker.

ball in the RUEG video is variably referred to as *father*, *husband*, *partner*, *friend*, *teenager*, *son*, *pedestrian*, or *passerby* in the English subcorpus; the ball is identified as a *ball*, *basketball*, *volleyball*, or *baseball* in the English data, while many German and Russian speakers identify it as a *football*. One Russian speaker identifies the woman handling the groceries as a vendor (продавец, *prodavec*). The location of the accident is variably identified as a *street/road*, *intersection/junction* or *parking lot* at a *supermarket*, *grocery store*, or *apartment complex* in the English data.

Some of the lexemes are byproducts of playacting, i.e., occur when speakers invent information that was not at all provided in the material. We will return to this in the next subsection. More commonly, the lexemes describe different interpretations and conceptualizations of the situation. Although objectively/externally, all speakers were shown the same material, their experience of the material and their experience of the interaction with the material differ significantly. This results in a wide range of lexical materials for the everyday referents and events they were presented with.

- (3) Many compounds in the dialogue corpora reflect on-the-spot semantic mapping. This includes lexical oddities as processing artifacts, for example, the incorporation of placeholders (*Reparaturladendings* – ‘repair shop thingamajig’), compounds with pleonastic (redundant) semantics (*Zeituhr* – ‘time clock’, commonly just *Uhr*), or unidiomatic amalgamations (*Holzbretterhütte* – ‘wood board hut’, commonly *Holzhütte* or *Bretterhütte*). Similarly, some referential expressions occur as almost identical word formations, such as *Spaßtagbanner*, ‘fun day banner’, vs *Spaßtagschild*, ‘fun day sign’. These cases appear to reflect linearizations of different associations into a single word (wood + board + hut; fun day + ? banner or sign?). This word formation activity is not communicatively required, as syntactic paraphrase with prepositional or relative clause modifiers would certainly be available. In a phraseological model, associative linearization of this kind with the result of idiosyncratic lexemes should be avoided, but it is abundantly found in the data.²⁸

More evidence for on-the-spot semantic mapping comes from a degree of specificity that seems unnecessarily high relative to the objective requirements of the communicative situation. For example, one speaker in the German subcorpus of RUEG (DEmo51M) describes the location of the accident as *Parkplatz*, *Parkplatzfläche*, *Parkfläche*, *Parkplatzrand*, and *Bürgersteig* (‘parking lot’, ‘parking lot area’, ‘parking area’, ‘end/edge of the parking lot’, ‘sidewalk’), all *within the same production* (formal written). In

²⁸ Many words of this kind occur as types due to the German convention of merging compounds into a single graphematic word. Although this is a language-specific idiosyncrasy, low convergence is not unique to the German data and cannot overall be attributed to graphematic artifacts. In fact, the German subcorpus of RUEG has a higher number of shared lexemes compared to the other two languages.

the map task corpus (BeMaTaC), speakers refer to an icon of a bread roll as *Milchbrötchen* ‘milky bread roll’ or *Kaiserbrötchen* ‘Kaiser bread roll’. In the absence of other types of bread rolls in the prompt material, speakers would not be required to specify the type of roll, and they could use the more general lexeme *Brötchen*, which would also be the more frequently occurring noun.²⁹ Similarly, we find the word *Metzgergeselle* ‘butcher assistant’, which refers to a trained butcher who has completed a 3-year apprenticeship, but has not obtained the professional master’s qualification required to open their own butchery. It is not obvious why the depicted person – a man holding a steak – should be an assistant, and not simply a butcher (*Metzger*). It is unclear why this degree of specification would be required either. In the map task corpus, speakers also use highly specific spatial expressions such as *Fünfundvierziggradwinkel* and *Fünfunddreißiggradwinkel*, ‘45° angle’ and (oddly) ‘35° angle’. Communicatively, the more common *schräg links*, *schräg nach oben* would appear to suffice (‘diagonally to the left, diagonally upward’).

The cases discussed in this section demonstrate that speakers do not generally pick the most frequent, most general, or most phraseologically predictable word in spontaneous reference or opt for syntactic paraphrase. It is possible that the high degree of specificity is found more in the visually prompted corpora because speakers in these elicitations need to (a) first map the visual input to their own semantic space and (b) formulate references that are recognizable, i.e., can be related to the same material by the interlocutor or understood in a way that the same images would emerge in the imagination of the interlocutor case of the witness report. This may invite a higher degree of specificity compared to the types of argumentative texts elicited in Falko and Kobalt, which are fundamentally intertextual, i.e., relate to more general, external, and abstract concepts and discourses. Moreover, word formation can also be used to signify socio-pragmatic meaning. For instance, compounds such as *Ovalding* ‘oval thing’ and *Reparaturladendings* ‘repair shop thingamajig’ may signify social and status information such as a relaxed attitude or a willingness to cooperate and remain within the common ground; or semiotic status, such as referential uncertainty in an ambiguous space (e.g., ‘I am aware and willing to warn you that this is not easy to recognize: the thing I want to refer to is somewhat oval’). The fact that these types of word formations are not abundantly found in the essay corpora suggests that different types of elicitation invite different types of path dependence and cognitive engagement.

4.2.3 Task interpretation and speaker engagement

A third path of divergence between speakers results from playacting and task engagement. This frequently delivers rather unpredictable results, of which two examples will be provided in this section.

In the map task corpus (BeMaTaC), speakers were prompted to describe a path on a map. They were not asked to pretend as though they were walking that path in an actual environment. Some speakers still align on this interpretation as common ground and consequentially refer to the icons as landmarks, even though not all of them easily lend themselves to this interpretation. Speakers then creatively interpret the icons as symbols for shops or other environments, e.g., *Motorradladen*, ‘motor bike shop’ for a motorbike icon, or *Hafenbar*, ‘port bar’ for an icon of a wooden steering wheel. In some cases, this results in semantically obscure lexemes, such as *Toasterladen*, ‘toaster shop’. Although it is contextually interpretable, as speakers blend the concept of shops with the concept represented in the icon, this example is surprising from a phraseological perspective of the lexicon. Designated toaster shops (likely) do not exist – or if they do, they must be exceedingly rare. It seems very unlikely that the speaker chose *Toasterladen* from their available vocabulary or a distributional niche. It was not situationally unavoidable either, as the dyad had already chosen *Hafenbar*, ‘port bar’, for the steering wheel icon. They could have chosen to refer to the toaster icon as a department store, a cafe, or some other conceptual blend or metonymic target that is more likely than a toaster shop. However, speakers seem to

²⁹ Moreover, *Milchbrötchen* refers to a commonly known, but different type of roll than is shown in the icon – meaning the speaker may map for specificity, but not for semantic correctness. This is, in a sense, the opposite of what would be expected in H2 (distributional niches).

easily overwrite both the semantic plausibility and the frequency of occurrence of a lexeme in favor of fostering discourse alignment.

In the witness report corpus (RUEG), many speakers add – i.e., invent – details for contextual embedding, for instance, where to or from they were walking when they witnessed the accident, or their emotional engagement. This points towards the activation of semantic frames of contextual embedding and results in some lexical divergence. In some cases, contextual embedding gains such importance that their narrations leave out most of the information provided in the stimulus video. One example is speaker RUmo27MR from the Russian subcorpus, whose four productions are provided in the appendix in full (Table A1). For the two formal conditions (spoken/written), speakers are prompted to provide a witness report to the police. For the two informal conditions, they are prompted to inform a friend about the events they witnessed. This speaker does not mention the actual events of the accident in three out of four conditions, notably both of the formal conditions. This might to some degree reflect a cultural bias of not readily providing information to the police. However, that alone does not fully explain the degree of variability, as the speaker does not give details on what happened in the informal spoken condition either.³⁰ As a result of their narrational diversity, 43 of the verbs used by the speaker are unique to one setting each, and new referents are introduced in each setting. While most speakers in RUEG supply more specific information across settings, the divergence between this speaker's four elicitations as well as their divergence with the other speakers' productions highlights the potential for creative and variable interpretation and solution of the same task. This also has repercussions on the statistical analysis: if all speakers behaved like this one, adding the factor 'formal' or 'informal' to the analysis would result in the statistically significant outcome of a register marking as informal for lexemes such as 'accident' or 'car'.

Another aspect of variable task interpretation is reflected in the observation that speakers in the essay corpora will sometimes produce texts that are partially argumentative essay, partially political speech, partially personal narrative, and so on (see also Shadrova 2020, Ch. 7). Some of this fluctuation between genres and styles, etc. may reflect a lack of experience or 'incomplete acquisition' of the prompted genre or register, especially since many of the participants were high school students at the time of elicitation.³¹ However, their productions are a genuine expression of these speakers' language activity; and are communicatively real in the sense that participants were asked to communicate a certain meaning and most of them cooperated sincerely, i.e., delivered linguistically plausible and coherent data on the prompted theme. Thus, variability in task interpretation should be considered to reflect genuine response variability in corpus-based elicitations.

This is of particular relevance to the interpretation of the data where task design is intended to limit speech production through visual prompts. If speakers respond with such variability in task interpretation, playacting, engagement with the story, add personal details, and so on even within such narrowly constrained tasks, then we can presume that they would also do this in an communicative situations outside the elicitation context, regardless of how narrow and limiting those situations may appear. This raises some questions for theories of usage-based lexicology that will be discussed in the final section of this article.

4.3 Summary

This section summarizes the results relative to the hypotheses presented in Section 3.2 (- hypothesis not met, + met, ++ met and exceeded). The analysis has yielded no evidence to support high between-speaker convergence in the sense of a limitation to a single or a small set of linguistic forms in context, phraseological or otherwise (–H1). A 1:1 mapping of individual or groupwise preferences or distributional niches for particular

³⁰ The police aspect of the RUEG elicitation has rightfully been criticized for cultural sensitivity.

³¹ It is worthy of reflection that a comprehensive metalinguistic awareness of the precise characteristics of genres is available almost exclusively to language and literature scholars and teachers and a deficiency hypothesis would severely limit the number of 'complete acquirers' in this field.

expressions in context, e.g., for common activities such as loading a car with groceries, is not supported by the data. No evidence of an observable set of qualitatively self-similar distributions was found. Thus, the stronger interpretation of transitivity through distributional niches of the construction cannot be confirmed at least at the level of individual contextualized productions (–H2). Rather, speakers make use of what appears to be the full range of near-synonyms in descriptive tasks, and only overlap in very generic thematic vocabulary in the essay productions, especially for verbs. Maximal variability between speakers and between productions of the same speakers appears to be the baseline across elicitation formats, and there is abundant evidence of idiosyncratic, path-dependent speaker engagement with the situation (++H5). This is partially realized through the use of constructional blends for idiosyncratic and socio-pragmatic meaning. However, constructional activity is not limited to this area, but overall abundant in descriptive tasks (++H4). Intra-individual distributions do not primarily vary by elicitation context, and near-synonyms do not divide into groups by elicitation factors. There are no identifiable subgroups of more convergent speakers, and speakers who form a community outside of the elicitation do not converge more than other speakers in the essay corpora. However, as it appears that different types of elicitation foster different types of path dependence, distributional/variationist perspectives might still explain some of the variability on a meta-linguistic level (+/–H3). Different communicative contexts might in fact work as ecologies for different types of cognitive strategies and occur with predictable diversity in linguistic expression. This might result in the use of more or fewer exemplars of particular types of word formation, without predicting particular identical word forms. Hypothesis 6 regarding language-specificity is partially confirmed (+/–H6). The English subcorpus of RUEG shows overall lower productivity as expected, and English speakers converge more with one another compared to Russian speakers. At the same time, the German subcorpus is the most convergent in RUEG. More research including parallel elicitations of the other formats in other languages is needed to fully understand those effects.

5 Discussion

This section offers a discussion of the results in light of (1) the presented models of lexical production, (2) the methodological tradition of deriving conclusions for individual cognition and system abstractions from population data, and (3) possible explanations of the apparent mismatch between real and perceived lexical variability in language and linguistics.

5.1 Revisiting models of lexical production

Sections 2.2 and 2.3 have pointed out that the principle of no synonymy, the idiom principle, and the stochastic principles of entrenchment predict stable lexical distributions by speaker or situation. Let us now revisit this claim with empirical data. In the RUEG corpus,³² speakers use 17 different lexemes to refer to the activity of loading groceries into the trunk of a car they witness in the same video: *räumen*, *packen*, *laden*, *legen*, *einräumen*, *einpacken*, *einladen*, *verladen*, *tun*, *sortieren*, *einsortieren*, *reinräumen*, *reinpacken*, *reinlegen*, *reintun*, *hineinlegen*, *befüllen* ('to stow', 'to pack', 'to load', 'to lay', 'to stow in', 'to pack into/wrap', 'to load up', 'to load away', 'to do (to put)', 'to sort', 'to sort into', 'to stow into', 'to pack into', 'to lay into', 'to do (put) into', 'to lay into', 'to fill (literally to be-fill)'). This does not say much about synonymy: the verbs are not formally synonymous and would likely occur in complementary distributions in large corpora. Even where the same speakers produce several verbs to refer to the same activity in the same production, we cannot know whether they perceive these words as interchangeable or whether they aim to provide new details with every lexical realization. Then, in a lexicographic sense, the observation may not be striking.

³² This includes L1 German subcorpora that were not used in the previous section.

However, if we consider the procedural and cumulative effects of this variability, logical challenges arise for the model of dynamic entrenchment of finely distributed semantic niches. In the RUEG example, the most frequent lexemes *packen* and *einpacken* ('to pack'; 'to pack/to wrap') are plausible semantic prototypes in the given situation, but only occur in about 10 and 6% of productions, respectively. This means that well over 80% of productions do not rely on prototypes, or that the most frequent lexemes would not occur 8–9 out of 10 times a similar situation is encountered. In the meantime, regardless of whether or not speakers perceive semantic differences between near-synonyms, their minds would be entrenched with associations between the lexemes of the set and with the situational meaning, as the model expects to 'wire together what fires together' (Sprenger et al. 2024). These mutual associations should also make the lexemes more synonymous over time, unless entrenchment is *a priori* semantically selective.

This is all the more powerful since interlocutors, like corpus linguists, have no way of knowing whether a speaker used a particular verb to relate the same meaning that they would have associated with a different verb, or whether the other person was highlighting another, subtly different aspect of the situation. Did they interpret the situation in a way that is better expressed with *einladen* than *reinräumen*, or are the two semantically indistinguishable in their map? Unless speakers engage in meta-communication to reflect on all sensory, perceptual, and interpretational aspects of a situation (which they rarely do), situational ambiguity will always prevail. This may seem trivial, but it has a direct effect on the model by imposing an upper boundary for the potential sharpness of the constructicon: under high lexical variability, neither form, nor meaning would be sharpened by naturalistic interactions. Entropy should then increase with every interaction.

Moreover, to successfully navigate such a space, i.e., to avoid perpetual cognitive dissonance and confusion from slightly off descriptions, speakers must accept and anticipate imprecise mappings between words and situational meaning. Then, even if they started out with an *a priori* ideal 1:1 map, it would quickly blur into an approximate *n:m* mapping – referential diversity in a passive entrenchment environment entails semiotic ambiguity. This means that the data presented is inconsistent with an interpretation of a unique and bidirectional (1:1) mapping of the constructicon as far as language production is concerned. The constructicon must either be of a different composition, not divided into stochastically precise semantic niches, or not be applied in a precise manner in production.³³

Similar conclusions can be drawn with respect to the role of stochastic learning of those mappings. In the presence of many acceptable slot fillers, the precise acquisition of fine-grained lexicographic distributions from stochastics alone is challenged by limitations of time and space. To remain with the previous example: all of the verbs used by the RUEG speakers seem acceptable for the expressed meaning. All pertain to the set of basic German vocabulary and would not generally be considered as marked for formality. Although loading groceries into the trunk of a car is a frequent activity for many speakers, it would still only occur a few times a week. Let us simplify and maximize exposure by assuming that speakers handle groceries twice a week and talk about it everytime. Let us further assume that they do not repeat lexemes until the entire set is instantiated – this is only for simplicity of illustration; of course, speakers are not expected to actualize the distribution in such orderly ways in reality. Even in this unlikely case, it would take 17 car fillings, or 8.5 weeks, to go through each of the verbs once. Most of the verbs in the set are formed from the directional particles *ein*, *rein*, and *hinein* (all meaning 'in'/'into') and various bases denoting variants of 'packing'. Speakers in the RUEG elicitation did not actualize the full productive potential of this pattern: they do not use *reinladen*, for instance, although it is perfectly acceptable and would probably occur in another elicitation.³⁴ Iterating through the complete list of plausible verbs would then require more instantiations, i.e., take more time. At this point, no distributional information at all would have been transmitted yet, as each verb would only have been realized

³³ It is possible that the semantic frames of the verbs mentioned do not correspond to an entity of meaning as is represented in the constructicon. Since the model is underspecified in this respect, more theoretical specification is needed to fully interpret the results in that framework.

³⁴ A morpholexical analysis of the combinatorics of prefixes, particles and verb stems will be provided elsewhere (Shadrova et al. in prep.).

once. How many months would it take to instantiate a stable distribution that accurately reflects the different frequencies of each verb as they are found in large corpora even for this everyday activity? A structural interpretation of the construction that requires constant reinforcement of the distributional properties of all lexemes does not seem plausible in a field as lexically variable as it is observed in the data presented, especially at scale for the entire lexicon.

Functional arguments towards predictability seem to fail under high local variability as well: using a verb should prime the speaker towards its reuse locally, and add to its global frequency. In a frequency-optimized model, it should then be preferred for the course of the production and in the future as well. The same should go for phraseological models that center conventionalization through convergence on form, or limitation of choice in context. By being used in context, a lexeme should gain phraseological weight and be prioritized in the future for easy access. This should be most valid for the most common and recurrent situations, as to provide the greatest cumulative advantage. But that is not the case – speakers barely reuse any lexemes and actively provide many variants even for the most basic referents and events.

The simplest way to coherently interpret that data is this: each subset of near-synonymous lexemes may reflect a lexically interconnected semantic cluster or network from which lexemes are realized more or less randomly. Higher frequency in production reflects (a) prototypicality in context and (b) genericity/low semantic weight, not idiosyncratic probability features. This view not only explains, but expects higher referential diversity in basic situations:³⁵ everytime a type of situation is experienced and expressed, a speaker has a chance to find a new way of perceiving and relating it, which can then be integrated into the long-term association of the frame. We would then find *less* conventionalization, *more diversity*, and *more innovation* in more frequent situations, as well as more lexical diversity over a lifetime.

5.2 Population vs individual and system vs behavior

The degree of variability in the presented corpora suggests that population averages may not scale down to speaker behavior for lexical frequency in a way that optimizes predictability. The most frequent lexemes only occur in a small number of productions, and most lexemes occur at very low rates. This is at odds with the established practice of connecting corpus frequencies to adult speaker behavior via population statistics, highlighting the need for more research into the distributions of individual behavior and their relationship to models of linguistic abstraction. In fact, the high inter- and intra-individual variability and the high degree of local interaction suggests general non-ergodicity and local non-stationarity even within limited semantic fields, which raises more general concerns with respect to the statistical paradigm in lexicology. Even if statistical patterns of pooled averages were stable,³⁶ speakers would only have access to the distributions they encounter, which do not seem to converge. There is no guarantee that speakers (a) could gather enough frequential evidence to reconstruct the underlying distributions, (b) would notice differences between the distributions they encounter, would then (c) average out and internalize those averages in order to (d) generalize from these patterns to their own usage. As words are discrete – a slot can only be occupied with a particular lexeme, not a distribution of lexemes or relative frequency of a lexeme – the loss in resolution for the representation of the distribution grows proportionately to the size of the set of variants.

Let it be noted that these conclusions are not incompatible with lexical association experiments that correlate corpus data with response times or collocational preferences based on corpus frequency. These associations may still reflect prototypes/basic level categories, i.e., semantically salient, connotationally most

³⁵ This is in line with the observation that highly conventionalized phrases, such as idioms, are fleetingly rare even in large corpora (Moon 1999).

³⁶ We do not know that they are, as replications of fine-grained lexical distributions through other than the originally presented corpora are virtually non-existent in the corpus-linguistic tradition. Since words have to be distributed in some way, a skewed distribution itself does not provide evidence of persistent distributional patterns (see also Piantadosi (2014); Aitchison et al. (2016); Shadrova (2021, 2022)).

neutral, or featurewise representative members of a category. Their frequency might be elevated due to their position in the ontological space, not *vice versa*. They may also be neutral association patterns in the absence of contextual alternatives. If the stochastic space changes in context (as would be predicted by the CDS perspective), global associations may only occur as one of many factors in local production. The epistemological problem arises when the two are equated and their presumed equivalence loaded with functional or causal explanations. Similarly, the observation that category learning is influenced by category distributions (e.g., Emberson et al. 2019, Casenhiser and Goldberg 2005) is untouched by this, as a skewed prototype distribution does not require fine-grained stochastic details, but can as well be understood in qualitative terms of feature similarity and abstraction.

More questions arise to the extrapolation from speaker *productions* to speaker *systems*. First, what appears as inter-individual variability in the presented corpora may well reflect intra-individual variability, i.e., participants might behave like each other in other elicitations or longer productions. Corpus-linguistic research, where it aspires to cognitive plausibility, is thus in need of repeated elicitations from the same speakers, i.e., longitudinal L1 data beyond acquisition contexts, to disentangle situational dynamics from stable individual and population effects. A better understanding of the relationship between behavior and system requires a clear demarcation of the factors that influence behavior – qualitatively, quantitatively, and from actual production data.

If there is no simple connection between population statistics and individual behavior, models of the relationship between individual behavior and individual system may also require more specification. Is language production an actualization of the existing system (a direct mapping of what already exists in the mental lexicon), or is it an active *application* of the productive/generative potential of the system?³⁷ The degree to which this matters depends on the proportion of such applications we expect to find in corpus data. How much of a speaker's situated lexical production is a reflection of their stable lexicographic system, and how much is a creative application under local uncertainty, and what does that say about the organization of words? The data presented suggests that productivity in context is actually abundant. This would be relevant to psychologically or cognitively oriented subfields that connect lexicographic evidence with theories of behavior, e.g., language acquisition and language dynamics research; but also to the descriptive adequacy of lexicographic models. For instance, when a speaker produces both 'parking lot *area*' and 'edge of the parking lot' (*Parkplatzfläche*, *Parkplatzrand*) in the same production, this is likely not a simple actualization of an existing lexicographic mapping to referents in space. In fact, the speaker may not have had a concept or word to speak of 'parking lot edges' before this interaction – while not impossible, it seems at least unusual. It may be spontaneously mapped because a) the speaker may be socio-pragmatically expressing their willingness to cooperate and be helpful by providing high resolution details, which may exceed the usual specificity of the parking lot semantics in their system; or b) the two realizations may be traces of a preliminary, and ongoing, categorization of an *a priori* undelineated amount of sensory input, and not of primarily communicative purpose. Such an explication of inner monologue, an otherwise internally and silently ongoing meaning-making process, may not be representative of the speaker's lexical system in a lexicographic sense, but an active application thereof.³⁸

Obviously, lexical productivity is a well-known phenomenon. However, in corpus linguistics, it has been treated as the exception, an addition to a largely complete lexicographic map. This is partially due to the fact that it is often impossible to know whether a word is productively formed or lexicalized for the speaker from corpus data alone. However, this problem is a correlate of the global lexicographic approach with lexical productivity: viewing all unique words in a large corpus as a bag of productively formed lexemes does not

³⁷ This debate has a long tradition for morphological de-/re-composition and structural vs 'holistic' processing in general (see for instance, Siyanova-Chanturia (2015)), but does not appear to include the conceptual mappings of words.

³⁸ Vygotsky and Piaget claim the same for observations of child speech, according to which younger, but not older children primarily use language to explain situations to themselves rather than as a communicative device (Vygotsky 1962). They interpret this as a slow submergence of observations into internal monologue at later stages. However, as internal monologue is not universal (see Nedergaard and Lupyan 2023, Fernyhough and Borghi 2023), Vygotsky's and Piaget's analysis might require more differentiation.

allow for behavioral analysis in context. The data presented in this article shows that productivity is not only omnipresent in description tasks, but also easy to spot due to many idiosyncrasies and divergences between speakers. In finding words like ‘35° angle’ (*Fünfunddreißiggradwinkel*) and ‘parking lot edge’, we can observe a process of situational meaning making and deduce, with some caution, that these exemplars are not plain realizations of generally used vocabulary (even though they may not be ‘truly novel’), but local applications – especially if none of the other speakers use it. The contrastive analysis of the commonalities and idiosyncrasies in lexical realizations between many different speakers in the same situation and multiple elicitations from the same speakers can thus illuminate aspects of the relationship between speaker system and speaker behavior. Even the analysis of the verbalization of minor aspects of a visual prompt, like the same micro-subevent of loading a car with groceries, has shown to be surprisingly rich in this regard (see also Shadrova et al. in prep.).

5.3 Perceptions of identity

One fascinating aspect to working with lexical corpus data is that readers of the material tend to vastly underestimate the degree of variability. For some reason, text seems to appear more lexically similar than it is – even to linguistic experts: in an informal survey of researchers working closely with the witness report corpus (RUEG), 14 out of 17 participants estimated an intra-individual lexical overlap of 40% between the four elicitation conditions (median estimation = 65%). In reality, speakers on average reuse about 12% of their lexemes in all four conditions in the German and Russian data, and less than 20% in the English data. Survey participants named 13 different verbs as candidates for the most frequent to occur with groceries, four of which do not occur at all. Fourteen participants estimated the prevalence of those verb lexemes at 40% and above (median = 189 productions or 71%). In reality, the most frequent verb occurs in only 10% of texts and was named only by one participant (*packen*, ‘to pack’).³⁹ Not only were participants closely familiar with the entire corpus – *every production* – but they were also informed on previous occasions that lexical overlap is lower than they might expect. Even so, their estimations exceed observations by half a magnitude.

A similar phenomenon on a meta-scale can be observed in the high estimations of the phraseological extent of everyday speech presented by lexicologists in the past (50% to over 80%, see Section 2.2). Recurrent linguistic structures *per se* are not evidence of situationally conventional ways of speaking, as situational conventionality can only be measured against chance of occurrence: to find many instances of a combination is only surprising if those numbers are higher than expected relative to the total number of realizations of the meaning in this or other ways. However, the fact that this view seems so natural and has gained so much traction might speak to an as-of-yet undescribed linguistic phenomenon: perhaps *in perception*, near-synonyms are actually summarized into a common semantic space, i.e., a mapping of *n* words to one fuzzy meaning complex (e.g., a semantic density field). Like an ‘inner phraseologist’, there might be a function that parses meaning from objectively diverse lexemes into one convenient semantic handle. Since recent research shows that conceptualizations even of common nouns do not align almost at all between speakers (Marti et al. 2023, Wulff et al. 2022), a tendency towards generally approximating, rather than well-defined and precise, semantic interaction seems to be plausible.

The constraints to verbatim production seem to further be defined by the limits of perception and memory. In fact, adult speakers are generally not particularly good at memorizing language verbatim. While humans possess the general ability to memorize long text, for instance, in oral culture or theater, for

³⁹ Verbs suggested as candidates for the most frequent and number of texts out of 267 they occur in: *aufheben* ‘to pick up’ (5), *einpacken* ‘to pack’ (17), *einräumen* ‘to load into’ (14), *fallen* ‘to fall’ (24), *fallen lassen* ‘to drop’ (10), *herunterfallen* ‘to fall down’ (3), *hineinräumen/-tun* ‘to load into, to put into’ (0), *laden* ‘to load’ (12), *machen* ‘to put, to do (to go shopping)’ (0), *packen* ‘to pack’ (27), *runterfallen* ‘to fall down’ (3), *tragen* ‘to carry’ (0). Numbers diverge from the analysis in Section 4.1 as the survey refers to all German subcorpora of RUEG, some of which were not used here.

most, this is also a laborious process that tends to require much focused attention and support through memorization techniques. This view is further supported by Roberts et al. (2023), who found that symbols are remembered better than words of the same meaning, e.g., \$ vs *dollar*.⁴⁰

The debate continues on whether stochastic learning is indeed a key process in child lexical acquisition (Unger et al. 2023, Kray et al. 2024). In either case, the verbatim precision of this learning may change over lifetime as a physiological effect of nervous system maturation or a structural effect of lexical system connectivity. Distributional learning may be helpful for initialization, but not for the maintenance of the associative meaning structure of many exemplars. Indeed, a recent study by Benitez and Li (2023) shows that children prefer unique labeling and avoid variable multiplex ($n:m$) mappings, while adults do not. However, even children will reformulate sentences in sentence repetition tasks to only include structures and lexemes that they have acquired (passive to active voice, Polišenská et al. 2015; related lexemes like ‘aunt’ to ‘uncle’, Torregrossa et al. 2023), i.e., repeat from language parsing, not from sound mimicking. Along with the well-known fact that the phonological processing of unfamiliar languages is limited to the degree that phonemic differentiation is acquired, these findings challenge the assumption that, even if we had access to a speaker’s entire input corpus, they would generally *process* the same content that we can *observe* with corpus-linguistic methods.

Perceived lexical identity of what is objectively a variable lexical space may then be indexical of a parsing activity in which meaning is not triggered (read from a dictionary of constructions in a discrete and sharp 1:1 map), but (inter-)actively reconstructed from all available contextual information, resulting in a complex and multidimensional, but approximate semantic object that is related to many lexemes with partially overlapping semantics. Any of those lexemes might then support the recreation of that object. If stochastics plays a role in this process, it would seem unlikely for it to take scope over the entire lexicon for each localized interaction.

6 Limitations

This study has two limitations. The first is that what appears as inter-individual variability in fact might constitute intra-individual variance, i.e., that the different speakers would behave like each other in other elicitations. If this was the case, it would not limit the implications for lexicological theory, but rather increase their weight. Since communicative situations are limited in time and amount of interactional exchange, speakers would be unable to realize all of their lexical potential for every situation they encounter. Since most communities of speakers do not exist in isolated social bubbles, it does not matter whether they speak to the same people who actualize different lexical potentials each time, or to different speakers who actualize the same lexical potential each time. Unless they were to keep track with individual statistics, their semantico-pragmatic maps or phraseological entrenchment would still become fuzzy-edged.

The other limitation concerns elicitation design. Perhaps none of the tasks in the presented corpora have succeeded at eliciting naturalistic lexical productions – it is possible that speakers behaved in unusually innovative or exceptionally productive ways due to being faced with a creativity-inspiring oddness of tasks. For instance, explaining a path on a map that shows a toaster may cause such cognitive dissonance in some speakers that they are thrown onto highly unnatural lexical paths due to their lack of a phraseological map for this situation. Similarly, speakers may have attempted to impress experimenters with morpholexical productivity as they were aware of being linguistically assessed by experts, especially since students are often incentivized to avoid lexical repetition in writing assignments. This problem unfortunately cannot be truly avoided in controlled task-based elicitations. As the data was elicited by different research groups at different institutions and over the course of almost a decade, it does not seem likely that those factors would be the same for all elicitations, but a cautious view is always warranted. It is reassuring that high variability was found for

⁴⁰ This may explain the smaller number of conventionalized visual symbols in use compared to the high productivity of the lexical domain.

even the most basic everyday vocabulary in all elicitations, i.e., words that are commonly used and do not stand out as attempts to impress. Overall, speakers in all elicitations have provided believable productions that appear cooperative. The data does not seem to reflect any obvious attempts at cutting the elicitation procedures short, an unwillingness to participate, or implausible solutions relative to the communicative prompts.

Unfortunately, more natural, but still contextually comparable productions are difficult to elicit in ethical ways. Collecting corpus data from existing text does not solve the issue, as most existing text is planned and edited, i.e., not spontaneous; only covers certain communicative purposes; and writing impulses or purposes ($\approx$ tasks) are not usually comparable between speakers – unless they are homework or language test assignments, in which case all of the aforementioned limitations still apply. More insight on types of elicitations and the productions they generate would be desirable, but is not currently available to the field. Without access to such a typology of elicitation methods for more naturalistic productions, this limitation cannot be overcome within the current state of corpus linguistics.

7 Summary and conclusion

This study has shown that speakers find a surprising number of ways to express similar meaning and essentially never converge in their lexical use, even for basic referents in narrowly constrained tasks. Based on the data presented, lexical divergence between speakers and productions appears to be the baseline in contextualized language use, systematically introduced by various path-dependent phenomena on cognitive, discourse, and socio-pragmatic levels. Although these phenomena are not new to linguistics, the usage-based tradition treats them as statistically negligible by empirically relying on metrics that are only well-defined in a stationary stochastic space. However, in light of the data presented, assumptions of self-similarity across levels of linguistic observation and abstraction cannot be considered valid for lexical distributions. Thus, a specification of usage-based hypotheses for production in context is needed. This includes in particular transitivity-based models of lexical acquisition and change – whereby speakers pick up very fine statistical patterns from other speakers' usage and project those back into the system – which need to explain how this could arise stochastically at scale for the entire lexicon under the local variability observed.

For descriptive tasks, much of the variability observed appears to be directly related to cognitive engagement within an *a priori* undercategorized and underspecified communicative and sensory space. In the corpus elicitations presented here, we ask speakers to describe what they see or what they think *before they know* what they see or what they think. Of course the need to describe and discuss things without full *a priori* knowledge of the details is common in everyday communication – speakers are constantly under pressure to construct or reconstruct understanding in context.⁴¹ For this, they first have to recognize and map meaning to objects, activities, and the social space they interact in, each of which allows for many degrees of freedom. Sometimes, they use many complementary or partially overlapping expressions, while other times, they refer with very high specificity to an element, for instance, the *edge* of the parking lot vs the parking lot *area*. The resulting referential diversity means that interlocutors are always confronted with many expressions of diverse interpretation of the same situation. If they learn to associate this entire potential of expressions with the meaning potential of a situation, their mapping must be of a many-to-many rather than a one-to-one layout. Analyzed in semantic detail, this complex space may be divisible into many fine semantic aspects and forms mapped to them, i.e., lexicographic distributional niches. But the situational applications may not be defined by such order, as speakers do not seem to mind mapping various near-synonyms to the same event in context, even if they are likely aware that *load*, *pack*, *store*, *fill*, and *put* do not all carry precisely the same semantics. In active language use, a many-to-one-to-many mapping then appears more plausible: in spite of the

⁴¹ This extends to past and previously described events, as episodic memories change and adapt over time, and reconstruction through replay appears to be a general feature of memory recall (Scully et al. 2017, Faber and Gennari 2015; Ólafsdóttir et al. 2018, Chen and Wilson 2023).

lexicographically intricate ties between form and function, in practice, a number of words can be used as handles to activate an approximate meaning complex that is ‘good enough’ for everyday interaction. This may then explain why speakers – including linguists – believe they have seen repetitions of the same words, even when they objectively have not.

Theories of the lexicon then need to specify and adjust their views on the relationship between corpus-based lexicography and usage-based lexicology of production, as neither the stochastic primacy of the lexicographic view of corpus linguistics nor the idiom principle of phraseology seems to serve well as guiding principles of production. As production can only be studied from production data, corpus linguistics is ideally suited to study those questions. This requires the development of methodologies to model and assess situational dynamics and cognitive engagement, e.g., quantitative techniques that account for path dependence; and taxonomies of elicitation that shed light on the ecologies of variability in context, which can only be derived from careful qualitative analysis of multiple elicitations of the same speakers (longitudinal data). For a better understanding of how the structure of the mental lexicon mediates process, and how the process of production creates structure, the field would profit from the triangulation of production data with experimental and neuroimaging data of the same cohorts. A CAS perspective can provide a good starting point, but does not currently provide a theoretical or methodological framework to study lexicology *per se*.

Unlike edited or prepared text from newspapers, academic text, or political speech – the home of many corpus-linguistic investigations past and present – every spontaneous production is influenced by situational dynamics. This dynamic space is the home that defines most of language interaction for speakers.⁴² Usage-based linguistics needs to integrate this fact in theory and methodological practice.

Acknowledgements: I am grateful to the corpus linguistics team at the Department of German Studies and Linguistics at Humboldt-Universität zu Berlin for their close collaboration and critical feedback. I owe special gratitude to Thomas Krause and Felix Golcher for their valuable input on the mathematical questions raised; to all colleagues who elicited, compiled, and annotated the data and shared their work in the various corpus projects referenced; and to Anke Lüdeling, who has designed and lead those data elicitations in ways that have made an analysis of situational variability possible. I would further like to thank four anonymous reviewers at Open Linguistics for their thorough reviews and insightful comments on critical issues, which have greatly helped to clarify and sharpen the central arguments of this manuscript. All remaining flaws and errors are my own.

Funding information: This article results from work performed within the project ‘Pc - corpus-linguistic methods’ funded by the German Research Association (2021-2024, DFG 313607803, GZ LU 856/16-1) as part of the research group ‘RUEG: Research Unit Emerging Grammars in Language Contact Situations’ (2018–2024, DFG 2537).

Author contributions: The author has accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The author states no conflict of interest.

Data availability statement: All datasets are publicly available under the respective references (see Section 3). Scripts for analysis of lexical overlap are available from the author upon reasonable request.

⁴² In the words of the Japanese Edo period poet Matsuo Bashō:
– “Every day is a journey
and the journey itself is home.”

References

- Ädel, Annelie. 2015. "Variability in Learner Corpora." In *The Cambridge Handbook of Learner Corpus Research*, edited by Sylviane Granger, Gaëtanelle Gilquin, and Fanny Meunier, 401–21. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.018>.
- Ägel, Vilmos. 2004. Phraseologismus als (valenz)syntaktischer normalfall. *Wortverbindungen – mehr oder weniger fest*, edited by Kathrin Steyer, 65–86. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110622768-005>.
- Agustín-Llach, M. and J. Rubio. 2024. "Navigating the Mental Lexicon: Network Structures, Lexical Search and Lexical Retrieval." *Journal of Psycholinguistic Research* 53 (2): 21. <https://doi.org/10.1007/s10936-024-10059-8>.
- Aitchison, Laurence, Nicola Corradi, and Peter E. Latham. 2016. "Zipf's Law Arises Naturally When There are Underlying, Unobserved Variables." *PLoS Computational Biology* 12 (12): e1005110. <https://doi.org/10.1371/journal.pcbi.1005110>.
- Al-Hoorie, Ali H., Phil Hiver, Diane Larsen-Freeman, and Wander Lowie. 2023. "From Replication to Substantiation: A Complexity Theory Perspective." *Language Teaching* 56 (2): 276–91. <https://doi.org/10.1017/S0261444821000409>.
- Alexiadou, Artemis and Vicky Rizou. 2023. "The Use of Periphrasis for the Expression of Aspect by Greek Heritage Speakers: A Case Study of Register Variation Narrowing." *Register Studies* 5 (1): 82–110. <https://doi.org/10.1075/rs.20022.ale>.
- Alexiadou, Artemis, Vasiliki Rizou, and Foteini Karkaletsou. 2022. "A Plural Indefinite Article in Heritage Greek: The Role of Register." *Languages* 7 (2): 115. <https://doi.org/10.3390/languages7020115>.
- Alexiadou, Artemis, Vasiliki Rizou, and Foteini Karkaletsou. 2023. "Agreement Asymmetries with Adjectives in Heritage Greek." *Languages*, 8 (2): 139. <https://doi.org/10.3390/languages8020139>.
- Alexopoulou, Theodora, Marije Michel, Akira Murakami, and Detmar Meurers. 2017. "Task Effects on Linguistic Complexity and Accuracy: A Large-scale Learner Corpus Analysis Employing Natural Language Processing Techniques." *Language Learning* 67 (S1): 180–208. <https://doi.org/10.1111/lang.12232>.
- Anders, Royce, Stéphanie Riès, L. van Maanen, and F.-Xavier Alario. 2015. "Evidence Accumulation as a Model for Lexical Selection." *Cognitive Psychology* 82: 57–73. <https://doi.org/10.1016/j.cogpsych.2015.07.002>.
- Anthonissen, Lynn. 2020. "Cognition in Construction Grammar: Connecting Individual and Community Grammars." *Cognitive Linguistics* 31 (2): 309–37. <https://doi.org/10.1515/cog-2019-0023>.
- Arias-Trejo, Natalia and Kim Plunkett. 2013. "What's in a Link: Associative and Taxonomic Priming Effects in the Infant Lexicon." *Cognition* 128 (2): 214–27.
- Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert, and Arne Zeschel. 2010. "Cognitive Corpus Linguistics: Five Points of Debate on Current Theory and Methodology." *Corpora* 5 (1): 1–27.
- Baayen, Harald, Lisa Feldman, and Robert Schreuder. 2006. "Morphological Influences on the Recognition of Monosyllabic Monomorphemic Words." *Journal of Memory and Language* 55 (2): 290–313. <https://doi.org/10.1016/j.jml.2006.03.008>.
- Baayen, Harald and Rochelle Lieber. 1991. "Productivity and English Derivation: A Corpus-based Study." *Linguistics* 29 (5): 801–44. <https://doi.org/10.1515/ling.1991.29.5.801>.
- Baayen, Harald and Eva Smolka. 2020. "Modeling Morphological Priming in German with Naive Discriminative Learning." *Frontiers in Communication* 5: 17.
- Baker, Rachel and Valerie Hazan. 2011. "Diapixuk: Task Materials for the Elicitation of Multiple Spontaneous Speech Dialogs." *Behavior Research Methods* 43: 761–70. <https://doi.org/10.3758/s13428-011-0075-y>.
- Balog, Evelin. 2023. *Entrenchment Revisited: Some Old and New Concepts and Their Empirical Validation*. Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). <https://open.fau.de/handle/openfau/22651>.
- Barth, Danielle and Vsevolod Kapatsinski. 2018. "Evaluating Logistic Mixed-Effects Models of Corpus-Linguistic Data in Light of Lexical Diffusion." In *Mixed-Effects Regression Models in Linguistics*, 99–116. Springer.
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015. *Parsimonious Mixed Models*. arXiv: <http://arXiv.org/abs/arXiv:1506.04967>.
- Bates, Elizabeth, Inge Bretherton, and Lynn Snyder. 1988. *From First Words to Grammar*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/S0305000900013908>.
- Bates, Elizabeth and Judith C. Goodman. 1999. "On the Emergence of Grammar from the Lexicon." In *The Emergence of Language*, edited by Brian MacWhinney, 29–79. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, et al. 2009. "Language is a Complex Adaptive System: Position Paper." *Language Learning* 59: 1–26. <https://doi.org/10.1111/j.1467-9922.2009.00533.x>.
- Belz, Malte, Christine Mooshammer. 2023. *Berlin Dialogue Corpus (BeDiaCo): Version 3. Medien-Repositorium, Humboldt-Universität zu Berlin*. <https://rs.cms.hu-berlin.de/phon>. <https://doi.org/10.5281/zenodo.8142681>.
- Benigni, Valentina, Paola M Cotta Ramusino, Fabio Mollica, Elmar Schafroth, et al. 2015. "How to Apply CXG to Phraseology: A Multilingual Research Project." *Journal of Social Sciences* 11 (3): 275–88.
- Benitez, Viridiana L and Ye Li. 2023. "Cross-situational Word Learning in Children and Adults: The Case of Lexical Overlap." *Language Learning and Development* 20 (3): 195–218. <https://doi.org/10.1080/15475441.2023.2256713>.
- Berry-Rogghe, Godelieve LMGA. 1970. *Collocations: Their Computation and Semantic Significance*. Ph.D. thesis, United Kingdom: The University of Manchester. <https://doi.org/10.1007/s10936-024-10059-8>.

- Biber, Douglas. 2012. "Register as a Predictor of Linguistic Variation." *Corpus Linguistics and Linguistic Theory* 8 (1): 9–37. <https://doi.org/10.1515/clt-2012-0002>.
- Biber, Douglas. 2019. "Multi-Dimensional Analysis: A Historical Synopsis." *Multi-Dimensional Analysis: Research Methods and Current Issues*, edited by Marcia Veirano Pinto and Tony Berber-Sardinha, 11–26. Bloomsbury.
- Biber, Douglas, Jesse Egbert, Bethany Gray, Rahel Opplinger, Benedikt Szmrecsanyi, Merja Kytö, and Päivi Pahta. 2016. "Variationist Versus Text-linguistic Approaches to Grammatical Change in English: Nominal Modifiers of Head Nouns." *Cambridge Handbooks in Language and Linguistics*, 351–75. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139600231.022>.
- Biber, Douglas and James K. Jones. 2009. "Quantitative Methods in Corpus Linguistics." In *Corpus Linguistics. An International Handbook*, edited by Anke Lädelling and Merja Kytö, Vol. 2, 1286–304. Berlin: De Gruyter Mouton. <https://korpling.german.hu-berlin.de/svn/bibliographie/2009>.
- Blomberg, Johan and Jordan Zlatev. 2021. "Metalinguistic Relativity: Does One's Ontology Determine One's View on Linguistic Relativity?." *Language and Communication* 76, 35–46. <https://doi.org/10.1016/j.langcom.2020.09.007>.
- Blumenthal-Dramé, Alice. 2013. *Entrenchment in Usage-based Theories: What Corpus Data Do and Do Not Reveal About the Mind*, vol. 83, Berlin, Boston: Walter de Gruyter. <https://doi.org/10.1515/9783110294002>.
- Boas, Hans Christian. 2010. "The Syntax-lexicon Continuum in Construction Grammar: A Case Study of English Communication Verbs." *Belgian Journal of Linguistics* 24 (1): 54–82.
- Boas, Hans Christian. 2013. Cognitive Construction Grammar. In *The Oxford Handbook of Construction Grammar*, edited by Thomas Hoffmann, Graeme Trousdale. Oxford University Press.
- Boas, Hans Christian and Ivan A Sag. 2012. *Sign-Based Construction Grammar*. Stanford: CSLI Publications.
- Bolinger, Dwight. 1979. Meaning and memory. In *Experience Forms: Their Cultural and Individual Place and Function*, edited by George G. Haydu, 95–111. The Hague: Mouton.
- Borensztajn, Gideon, Willem Zuidema, and Rens Bod. 2009. "Children's Grammars Grow More Abstract with Age-evidence From an Automatic Procedure for Identifying the Productive Units of Language." *Topics in Cognitive Science* 1 (1): 175–88. <https://doi.org/10.1111/j.1756-8765.2008.01009.x>.
- Branigan, Holly P, Martin J. Pickering, and Alexandra A. Cleland. 2000. "Syntactic Co-ordination in Dialogue." *Cognition* 75 (2): B13–25. [https://doi.org/10.1016/S0010-0277\(99\)00081-5](https://doi.org/10.1016/S0010-0277(99)00081-5).
- Branigan, Holly P., Martin J. Pickering, Janet F. McLean, and Alexandra A. Cleland. 2007. "Syntactic Alignment and Participant Role in Dialogue." *Cognition* 104 (2): 163–97. <https://doi.org/10.1016/j.cognition.2006.05.006>.
- Brezina, Vaclav. 2018. *Collocation Graphs and Networks: Selected Applications*, 59–83. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-92582-0_4.
- Bruni, Elia, Nam-Khanh Tran, and Marco Baroni. 2014. "Multimodal Distributional Semantics." *Journal of Artificial Intelligence Research* 49, 1–47.
- Brysbaert, Marc, Michaël Stevens, Paweł Mandera, and Emmanuel Keuleers. 2016. "How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age." *Frontiers in Psychology* 7: 1116. <https://doi.org/10.3389/fpsyg.2016.01116>.
- Bunk, Oliver. 2024. "What Does Linguistic Structure Tell us About Language Ideologies? the Case of Majority Language Anxiety in Germany." *European Journal of Applied Linguistics* 12 (1): 91–116.
- Burch, Brent, Jesse Egbert, and Douglas Biber. 2017. "Measuring and Interpreting Lexical Dispersion in Corpus Linguistics." *Journal of Research Design and Statistics in Linguistics and Communication Science* 3 (2): 189–216. <https://doi.org/10.1558/jrds.33066>.
- Bybee, Joan. 2002. "Sequentiality as the Basis of Constituent Structure." *Typological Studies in Language* 53: 109–34. <https://doi.org/http://dx.doi.org/10.1075/tsl.53.07byb>.
- Bybee, Joan. 2013. "Usage-based Theory and Exemplar Representations of Constructions." In *The Oxford Handbook of Construction Grammar*, 49–68. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0004>.
- Bybee, Joan and Paul Hopper. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins Publishing.
- Bybee, Joan and Rena Torres Cacoullos. 2009. "The Role of Prefabs in Grammaticization: How the Particular and the General Interact." *Formulaic Language* 1: 187–217.
- Cacoullos, Rena Torres and Catherine E. Travis. 2019. "Variationist Typology: Shared Probabilistic Constraints Across (Non-)null Subject Languages." *Linguistics* 57 (3): 653–92. <https://doi.org/10.1515/ling-2019-0011>.
- Calude, Andreea S. 2008. "Demonstrative Clefts and Double Cleft Constructions in Spontaneous Spoken English*." *Studia Linguistica* 62 (1): 78–118. <https://doi.org/10.1111/j.1467-9582.2007.00140.x>.
- Carrol, Gareth and Kathy Conklin. 2020. "Is All Formulaic Language Created Equal? Unpacking the Processing Advantage for Different Types of Formulaic Sequences." *Language and Speech* 63 (1): 95–122.
- Casenhiser, Devin and Adele E Goldberg. 2005. "Fast Mapping between a Phrasal Form and Meaning." *Developmental Science* 8 (6): 500–8. <https://doi.org/10.1111/j.1467-7687.2005.00441.x>.
- Chafe, Wallace L. 1980. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.
- Chen, Zhe Sage and Matthew Wilson. 2023. "How Our Understanding of Memory Replay Evolves." *Journal of Neurophysiology* 129(3): 552–80. 36752404. <https://doi.org/10.1152/jn.00454.2022>.
- Citraro, Salvatore, Michael S Vitevitch, Massimo Stella, and Giulio Rossetti. 2023. "Feature-rich Multiplex Lexical Networks Reveal Mental Strategies of Early Language Learning." *Scientific Reports* 13(1): 1474. <https://doi.org/10.1038/s41598-022-27029-6>.

- Conklin, Kathy and Gareth Carrol. 2021. "Words Go Together like 'Bread and Butter': The Rapid, Automatic Acquisition of Lexical Patterns." *Applied Linguistics* 42 (3): 492–513. <https://doi.org/10.1093/applin/amaa034>.
- Conklin, Kathy and Norbert Schmitt. 2012. "The Processing of Formulaic Language." *Annual Review of Applied Linguistics* 32: 45–61.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198299554.001.0001>.
- Dąbrowska, Ewa. 2014. "Words that Go Together: Measuring Individual Differences in Native Speakers Knowledge of Collocations." *The Mental Lexicon* 9 (3): 401–18. <https://doi.org/10.1075/ml.9.3.02dab>.
- Dąbrowska, Ewa. 2018. "Experience, Aptitude and Individual Differences in Native Language Ultimate Attainment." *Cognition* 178: 222–35.
- Dąbrowska, Ewa. 2019. "Experience, Aptitude, and Individual Differences in Linguistic Attainment: A Comparison of Native and Nonnative Speakers." *Language Learning* 69: 72–100. <https://doi.org/10.1111/lang.12323>.
- Dąbrowska, Ewa. 2020. "Language as a Phenomenon of the Third Kind." *Cognitive Linguistics* 31 (2): 213–29. <https://doi.org/10.1515/cog-2019-0029>.
- Dębowski, Łukasz. 2018. "Is Natural Language a Perigraphic Process? The Theorem about Facts and Words Revisited." *Entropy* 20: 85–111. <https://doi.org/10.3390/e20020085>.
- Dąbrowska, Ewa and Elena Lieven. 2005. "Towards a Lexically Specific Grammar of Childrens Question Constructions." *Cognitive Linguistics* 16 (3): 437–74. <https://doi.org/10.1515/cogl.2005.16.3.437>.
- De Smet, Hendrik. 2020. "What Predicts Productivity? Theory Meets Individuals." *Cognitive Linguistics* 31 (2): 251–78. <https://doi.org/10.1515/cog-2019-0026>.
- Dębowski, Łukasz and Christian Bentz. 2020. "Information Theory and Language." *Entropy* 22(2): 435.
- Demo, Š. 2022. "Lexical Productivity in Early Modern Latin According to the Neulateinsche Wortliste: A Quantitative Study." *Nordic Journal of Renaissance Studies* 19: 27–52.
- Dewaele, Jean-Marc and Aneta Pavlenko. 2003. *Chapter 7. Productivity and Lexical Diversity in Native and Non-Native Speech: A Study of Cross-cultural Effects*, 120–41. Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781853596346-009>.
- Diesendruck, Gil and Lori Markson. 2001. "Children's Avoidance of Lexical Overlap: A Pragmatic Account." *Developmental Psychology* 37 (5): 630. <https://doi.org/10.1037/0012-1649.37.5.630>.
- Diessel, Holger. 2023. *The Constructicon: Taxonomies and Networks*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009327848>.
- Dijkstra, Ton. 2003. "Lexical Processing in Bilinguals and Multilinguals: The Word Selection Problem." *The Multilingual Lexicon*. 11–26.
- Ding, Jianxin. 2018. *Introduction*. Singapore: Springer. https://doi.org/10.1007/978-981-10-7010-5_1.
- Dirdal, Hildegunn. 2022. "Development of L2 Writing Complexity: Clause Types, L1 Influence and Individual Differences." *Complexity, Accuracy and Fluency in Learner Corpus Research*, 81–114. Amsterdam, Netherlands: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.104.04dir>.
- Dobnik, Simon, Christine Howes, and John D. Kelleher. 2015. "Changing Perspective: Local Alignment of Reference Frames in Dialogue." *Proceedings of goDIAL-SemDial*, 24–32.
- Dux, Ryan Joseph. 2016. *A Usage-based Approach to Verb Classes in English and German*. Ph.D. thesis. Austin: University of Texas. <http://hdl.handle.net/2152/41229>.
- Egbert, Jesse, Brent Burch, and Douglas Biber. 2020. "Lexical Dispersion and Corpus Design." *International Journal of Corpus Linguistics* 25 (1): 89–115.
- Ellis, Nick C. 2002. "Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition." *Studies in Second Language Acquisition* 24 (2): 143–88. <https://doi.org/10.1017/S0272263102002024>.
- Ellis, Nick C. 2006. "Language Acquisition as Rational Contingency Learning." *Applied Linguistics*, 27 (1): 1–24. <https://doi.org/10.1093/applin/ami038>.
- Ellis, Nick C. 2008. "The Periphery and the Heart of Language." In *Phraseology: An Interdisciplinary Perspective* 1–13. Amsterdam: John Benjamins. <https://doi.org/https://hdl.handle.net/2027.42/139793>.
- Ellis, Nick C., Rita Simpson-Vlach, and Carson Maynard. 2008. "Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and Tesol." *Tesol Quarterly* 42 (3): 375–96. <https://doi.org/10.1002/j.1545-7249.2008.tb00137.x>.
- Emberson, Lauren L, Nicole Loncar, Carolyn Mazzei, Isaac Treves, and Adele E. Goldberg. 2019. "The Blowfish Effect: Children and Adults Use Atypical Exemplars to Infer More Narrow Categories During Word Learning." *Journal of Child Language* 46 (5): 938–54. <https://doi.org/10.1017/S0305000919000266>.
- Erman, Britt and Beatrice Warren. 2000. "The Idiom Principle and the Open Choice Principle." *Text-Interdisciplinary Journal for the Study of Discourse* 20 (1): 29–62. <https://doi.org/10.1515/text.1.2000.20.1.29>.
- Evert, Stefan, Peter Uhrig, Sabine Bartsch, and Thomas Proisl. 2017. "E-View-Affiliation - A Large-scale Evaluation Study of Association Measures for Collocation Identification." *Proceedings of eLex 2017-Electronic lexicography in the 21st century: Lexicography from Scratch*, 531–49.
- Faber, Myrthe and Silvia P Gennari. 2015. "In Search of Lost Time: Reconstructing the Unfolding of Events from Memory." *Cognition* 143: 193–202. <https://doi.org/10.1016/j.cognition.2015.06.014>.
- Fargier, Raphaël and Marina Laganaro. 2023. "Referential and Inferential Production Across the Lifespan: Different Patterns and Different Predictive Cognitive Factors." *Frontiers in Psychology* 14: 1237523.
- Faulhaber, Susen. 2011. *Verb Valency Patterns: A Challenge for Semantics-based Accounts*. Vol. 71. Mouton: De Gruyter. <https://doi.org/10.1515/9783110240788>.

- Fergadiotis, Gerasimos. 2011. *Modeling Lexical Diversity Across Language Sampling and Estimation Techniques*. Arizona State University. PhD thesis. <https://core.ac.uk/reader/79563501>.
- Fernández-Domínguez, Jesús. 2010. "Productivity vs Lexicalization: Frequency-based Hypotheses on Word-formation." *Poznań Studies in Contemporary Linguistics* 46 (2): 193–219.
- Fernyhough, Charles and Anna M Borghi. 2023. "Inner Speech as Language Process and Cognitive Tool." *Trends in Cognitive Sciences* 27 (12): 1180–93. <https://doi.org/10.1016/j.tics.2023.08.014>.
- Fonteyn, Lauren and Andrea Nini. 2020. "Individuality in Syntactic Variation: An Investigation of the Seventeenth-century Gerund Alternation." *Cognitive Linguistics* 31 (2): 279–308. <https://doi.org/10.1515/cog-2019-0040>.
- Fossa, Pablo, Nicolás Gonzalez, and Francesca Cordero Di Montezemolo. 2019. "From Inner Speech To Mind-wandering: Developing a Comprehensive Model of Inner Mental Activity Trajectories." *Integrative Psychological and Behavioral Science* 53: 298–322. <https://doi.org/10.1007/s12124-018-9462-6>.
- Frath, Pierre and Christopher Gledhill. 2005. "Free-range Clusters or Frozen Chunks? Reference As a Defining Criterion for Linguistic Units." *Recherches Anglaises et Nord Americaines* 38: 25–44. <https://doi.org/https://u-paris.hal.science/hal-01220302>.
- Garrod, Simon and Martin J Pickering. 2007. "Alignment in Dialogue." *The Oxford Handbook of Psycholinguistics*, 443–51. Oxford (UK): Oxford Academic. <https://doi.org/10.1093/oxfordhb/9780198568971.013.0026>.
- Gebril, Atta and Lia Plakans. 2016. "Source-based Tasks in Academic Writing Assessment: Lexical Diversity, Textual Borrowing and Proficiency." *Journal of English for Academic Purposes* 24: 78–88. <https://doi.org/10.1016/j.jeap.2016.10.001>.
- Gilquin, Gaëtan. 2021. "Learner Corpora." In *A Practical Handbook of Corpus Linguistics*, edited by Paquot Magali and Stefan Th Gries, 283–303. Cham: Springer.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>.
- Goldberg, Adele E. 2013. "Constructionist Approaches." In *The Oxford Handbook of Construction Grammar*, edited by Thomas Hoffmann and Graeme Trousdale, 15–31. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0002>.
- Granger, Sylviane. 2005. "Pushing Back the Limits of Phraseology: How Far can We Go." In *Proceedings of the Phraseology 2005 Conference*, 1–4.
- Granger, Sylviane and Fanny Meunier. 2008. *Phraseology: An Interdisciplinary Perspective*. Amsterdam, Netherlands: John Benjamins Publishing.
- Granger, Sylviane and Magali Paquot. 2008. Disentangling the Phraseological Web. In *Phraseology: An interdisciplinary perspective*, edited by Sylviane Granger and Fanny Meunier, 27–49. Amsterdam: John Benjamins. <https://doi.org/http://dx.doi.org/10.1075/z.139.07gra>.
- Gries, Stefan Th. 2005. "Syntactic Priming: A Corpus-based Approach." *Journal of Psycholinguistic Research* 34 (4): 365–99. <https://doi.org/10.1007/s10936-005-6139-3>.
- Gries, Stefan Th. 2010. "Behavioral Profiles: A Fine-grained and Quantitative Approach in Corpus-based Lexical Semantics." *The Mental Lexicon*, 5 (3): 323–46. <https://doi.org/http://dx.doi.org/10.1075/bct.47.04gri>.
- Gries, Stefan Th. 2013. "50-Something Years of Work on Collocations." *International Journal of Corpus Linguistics* 18 (1): 137–66. <https://doi.org/10.1075/ijcl.18.1.09gri>.
- Gries, Stefan Th. 2015. "The Most Under-used Statistical Method in Corpus Linguistics: Multi-level (and Mixed-effects) Models." *Corpora*, 10 (1): 95–125. <https://doi.org/10.3366/cor.2015.0068>.
- Gries, Stefan Th. 2019. "15 Years of Collostructions. Some Long Overdue Additions/Corrections (to/of actually All Sorts of Corpus-linguistics Measures)." *International Journal of Corpus Linguistics* 24 (3): 385–412. <https://doi.org/10.1075/ijcl.00011.gri>.
- Gries, Stefan Th. 2021. ("Generalized Linear) Mixed-effects Modeling: A Learner Corpus Example." *Language Learning* 71 (3): 757–98. <https://doi.org/10.1111/lang.12448>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/lang.12448>.
- Gries, Stefan Th. 2022. "What Do (most of) Our Dispersion Measures Measure (most)? Dispersion?." *Journal of Second Language Studies* 5 (2): 171–205.
- Gries, Stefan Th. 2023. "Overhauling Collostructional Analysis: Towards More Descriptive Simplicity and more Explanatory Adequacy." *Cognitive Semantics* 9 (3): 351–86. <https://doi.org/10.1163/23526416-bja10056>.
- Gries, Stefan Th and Gerrit Jan Kootstra. 2017. "Structural Priming within and Across Languages: A Corpus-based Perspective." *Language and Cognition* 20 (2): 235–50. <https://doi.org/10.1017/S1366728916001085>.
- Gries, Stefan Th and Naoki Otani. 2010. "Behavioral Profiles: A Corpus-based Perspective on Synonymy and Antonymy." *ICAME Journal* 34 (1): 121–50. <https://doi.org/http://dx.doi.org/10.1075/bct.47.04gri>.
- Guliyeva, Kamala Vasif. 2016. "Different Approaches to the Objects of Phraseology in Linguistics." *International Journal of English Linguistics* 6: 104. <https://doi.org/https://api.semanticscholar.org/CorpusID:54680673>.
- Gyllstad, Henrik and Brent Wolter. 2016. "Collocational Processing in Light of the Phraseological Continuum Model: Does Semantic Transparency Matter?." *Language Learning* 66 (2): 296–323. <https://doi.org/10.1111/lang.12143>.
- Heitmeier, Maria, Yu-Ying Chuang, and Harald Baayen. 2023. "How Trial-to-Trial Learning Shapes Mappings in the Mental Lexicon: Modelling Lexical Decision with Linear Discriminative Learning." *Cognitive Psychology* 146: 101598. <https://doi.org/10.1016/j.cogpsych.2023.101598>.
- Hernández-Campoy, Juan Manuel. 2021. "7 Corpus-based Lifespan Change in Late Middle English." *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies*, vol. 164. Oxfordshire: Routledge. <https://doi.org/http://dx.doi.org/10.4324/9780429030314-7>.

- Hilpert, Martin. 2017. "Frequencies in Diachronic Corpora and Knowledge of Language." In *The Changing English Language - Psycholinguistic Perspectives*, Marianne Hundt, Sandra Mollin, and Simone E. Pfenniger, 49–68. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316091746>.
- Hilpert, Martin and Holger Diessel. 2017. "Entrenchment in Construction Grammar." In *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*, edited by Hans-Jörg Schmid, 57–74. Mouton: American Psychological Association; De Gruyter. <https://doi.org/10.1037/15969-004>.
- Hirschmann, Hagen, Anke Lüdeling, Anna Shadrova, Dominique Bobeck, Martin Klotz, Roodabeh Akbari, et al. 2022. "Falko. Eine Familie vielseitig annotierter Lernerkorpora des Deutschen als Fremdsprache." *Zeitschrift Korpora Deutsch als Fremdsprache (KorDaF)*, 139–48. <https://doi.org/10.48694/kordaf.3552>.
- Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge. <https://doi.org/10.4324/9780203327630>.
- Holdaway, Cameron and Steven T Piantadosi. 2022. "Stochastic Time-series Analyses Highlight the Day-to-day Dynamics of Lexical Frequencies." *Cognitive Science* 46 (12): e13215. <https://doi.org/10.1111/cogs.13215>.
- Horton, William S, and Daniel H Spieler. 2007. "Age-related Differences in Communication and Audience Design." *Psychology and Aging* 22 (2): 281. <https://doi.org/10.1037/0882-7974.22.2.281>.
- Hu, Renfen, Jifeng Wu, and Xiaofei Lu. 2022. "Word-combination-based Measures Of Phraseological Diversity, Sophistication, and Complexity and their Relationship to Second Language Chinese Proficiency and Writing Quality." *Language Learning* 72 (4): 1128–69. <https://doi.org/10.1111/lang.12511>.
- Hulstijn, Jan. 2024. "Predictions of Individual Differences in the Acquisition of Native and Non-native Languages: An Update of BLC Theory." *Languages* 9 (5): 173. <https://doi.org/10.3390/languages9050173>.
- Hutchison, Keith A, David A Balota, Michael J Cortese, and Jason M Watson. 2008. "Predicting Semantic Priming at the Item Level." *Quarterly Journal of Experimental Psychology* 61 (7): 1036–66. <https://doi.org/10.1080/17470210701438111>.
- Ibbotson, Paul. 2013. "The Scope of Usage-based Theory." *Frontiers in Psychology* 4: 1–15. <https://doi.org/10.3389/fpsyg.2013.00255>.
- Jacobs, Cassandra L, Gary S Dell, and Colin Bannard. 2017. "Phrase Frequency Effects in Free Recall: Evidence for Redintegration." *Journal of Memory and Language* 97: 1–16. <https://doi.org/10.1016/j.jml.2017.07.003>.
- Jarvis, Scott and M. Daller. 2013. "Defining and Measuring Lexical Diversity." *Vocabulary Knowledge: Human Ratings and Automated Measures*, edited by Scott Jarvis and Michael Daller. Amsterdam, The Netherlands: John Benjamins. <https://doi.org/http://digital.casalini.it/9789027271679>.
- Jiang, Jingyang, Peng Bi, Nana Xie, and Haitao Liu. 2023. "Phraseological Complexity and Low- and Intermediate-level L2 Learners Writing Quality." *International Review of Applied Linguistics in Language Teaching* 61 (3): 765–90. <https://doi.org/10.1515/iral-2019-0147>.
- Jiang, Nan AN and Tatiana M Nekrasova. 2007. "The Processing of Formulaic Sequences by Second Language Speakers." *The Modern Language Journal* 91 (3): 433–45.
- Jolsavi, Hajnal, Stewart M McCauley, and Morten H Christiansen. 2013. "Meaning Overrides Frequency in Idiomatic and Compositional Multiword Chunks." In Markus Knauff, Michael Pauen, Nathalie Sebanz, and Ipke Wachsmuth, *Proceedings of the 35th Annual Meeting of the Cognitive Science Society, Austin, Texas*, 692–7. Cognitive Science Society. <https://escholarship.org/uc/item/5cv7b5xs>.
- Kachkovskaia, Tatiana, Alla Menshikova, Daniil Kocharov, Pavel Kholiavin, Anna Mamushina. 2022. "Social and Situational Factors of Speaker Variability in Collaborative Dialogues." In *Speech Prosody*, vol. 11, 455–9. Lisbon, Portugal. <https://doi.org/http://dx.doi.org/10.21437/SpeechProsody.2022-93>.
- Kalantari, Reza and Javad Gholami. 2017. "Lexical Complexity Development from Dynamic Systems Theory Perspective: Lexical Density, Diversity, and Sophistication." *International Journal of Instruction* 10 (4): 1–18. <https://doi.org/http://dx.doi.org/10.12973/iji.2017.1041a>.
- Kalashnikova, Marina, Aimee Oliveri, Karen Mattock. 2019. "Acceptance of Lexical Overlap by Monolingual and Bilingual Toddlers." *International Journal of Bilingualism* 23 (6): 1517–30. <https://doi.org/10.1177/1367006918808041>.
- Kambanaros, Maria, Kleanthes K Grohmann, and Michalis Michaelides. 2013. "Lexical Retrieval for Nouns and Verbs in Typically Developing Bilectal Children." *First Language* 33 (2): 182–99. <https://doi.org/10.1177/0142723713479435>.
- Karjus, Andres, Richard A Blythe, Simon Kirby, and Kenny Smith. 2020. "Quantifying the Dynamics of Topical Fluctuations in Language." *Language Dynamics and Change* 10 (1): 86–125. <https://doi.org/10.1163/22105832-01001200>.
- Kaschak, Michael P, Timothy J Kutta, and Jacqueline M Coyle. 2014. "Long and Short Term Cumulative Structural Priming Effects." *Language, Cognition and Neuroscience* 29 (6): 728–43. <https://doi.org/10.1080/01690965.2011.641387>.
- Keller, Rudi. 2014. *Sprachwandel: Von der unsichtbaren Hand in der Sprache*. Stuttgart: utb GmbH.
- Kemmerer, David. 2023. "Grounded Cognition Entails Linguistic Relativity: A Neglected Implication of a Major Semantic Theory." *Topics in Cognitive Science* 15 (4): 615–47. <https://doi.org/10.1111/tops.12628>.
- Kerz, Elma and Florian Haas. 2009. "The Aim is to Analyze np. the Function of Prefabricated Chunks." *Formulaic Language Volume 1. Distribution and Historical Change* 82: 97–116. <https://doi.org/10.1075/tsl.82.05the>.
- Kerz, Elma and Daniel Wiechmann. 2020. "Individual Differences." In *The Routledge Handbook of Second Language Acquisition and Corpora*, 394–406. Oxfordshire: Routledge. <https://doi.org/http://dx.doi.org/10.4324/9781351137904-35>.
- Kidd, Evan. 2012. "Individual Differences in Syntactic Priming in Language Acquisition." *Applied Psycholinguistics* 33 (2): 393–418. <https://doi.org/10.1017/S0142716411000415>.
- Kilgarriff, Adam. 2005. "Language is Never, Ever, Ever, Random." *Corpus Linguistics and Linguistic Theory* 1 (2): 263–76. <https://doi.org/10.1515/cllt.2005.1.2.263>.

- Kim, Minkyung, Scott Crossley, and Kristopher Kyle. 2018. "Lexical Sophistication as a Multidimensional Phenomenon: Relations to Second Language Lexical Proficiency, Development, and Writing Quality." *The Modern Language Journal* 102 (1): 120–41. <https://doi.org/10.1111/modl.12447>.
- Koplenig, Alexander. 2017. "Against Statistical Significance Testing in Corpus Linguistics." *Corpus Linguistics and Linguistic Theory* 15(2), 321–46. <https://doi.org/10.1515/cllt-2016-0036>.
- Kray, Jutta, Linda Sommerfeld, Arielle Borovsky, and Katja Häuser. 2024. "The Role of Prediction Error in the Development of Language Learning and Memory." *Child Development Perspectives* 18 (4): 190–203. <https://doi.org/10.1111/cdep.12515>.
- Kroll, Judith F., Jason W Gullifer, and Eleonora Rossi. 2013. "The Multilingual Lexicon: The Cognitive and Neural Basis of Lexical Comprehension and Production in Two or More Languages." *Annual Review of Applied Linguistics* 33: 102–27.
- Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kveim, Vilde A., Laurenz Salm, Talia Ulmer, Maria Lahr, Steffen Kandler, Fabia Imhof, and Flavio Donato. 2024. "Divergent Recruitment of Developmentally Defined Neuronal Ensembles Supports Memory Dynamics." *Science* 3856710, eadk0997. <https://doi.org/10.1126/science.adk0997>.
- Kyle, Kristopher and Scott Crossley. 2015. "Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application." *Tesol Quarterly* 49 (4): 757–86. <https://doi.org/10.1002/tesq.194>.
- Kyle, Kristopher and Scott Crossley. 2016. "The Relationship between Lexical Sophistication and Independent and Source-based Writing." *Journal of Second Language Writing* 34: 12–24. <https://doi.org/10.1016/j.jslw.2016.10.003>.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago.
- Lakoff, George. 1999. "Cognitive Models and Prototype Theory." In *Concepts: Core Readings*, edited by Margolis, Eric & Laurence, and Stephen, 391–421. Cambridge: MIT Press.
- Lamy, Dominique, Andrew B Leber, and Howard E Egeth. 2012. "Selective Attention." *Handbook of Psychology*, Second Edition 4. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9781118133880.hop204010>.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Vol. 1. Stanford: Stanford University Press.
- Larsen-Freeman, Diane. 2019. "On Language Learner Agency: A Complex Dynamic Systems Theory Perspective." *The Modern Language Journal* 103: 61–79. <https://doi.org/10.1111/modl.12536>.
- Larsson, Tove and Douglas Biber. 2024. "On the Perils of Linguistically Opaque Measures and Methods: Toward Increased Transparency and Linguistic Interpretability." In *Corpora for Language Learning: Bridging the Research-Practice Divide*, edited by Peter Crosthwaite, 131–41. London: Routledge. <https://doi.org/10.4324/9781003413301-10>.
- Larsson, Tove, Jesse Egbert, and Douglas Biber. 2022. "On the Status of Statistical Reporting versus Linguistic Description in Corpus Linguistics: A Ten-year Perspective." *Corpora* 17 (1): 137–57.
- Laufer, Batia and Tina Waldman. 2011. "Verb-noun Collocations in Second Language Writing: A Corpus Analysis of Learners English." *Language Learning* 61 (2): 647–72. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>.
- Leclercq, Benoit and Cameron Morin. 2023. "No Equivalence: A New Principle of No Synonymy." *Constructions* 15 (1): 1–16. <https://doi.org/10.24338/cons-535>.
- Lei, Siyu and Ruiying Yang. 2020. "Lexical Richness in Research Articles: Corpus-based Comparative Study among Advanced Chinese Learners of English, English Native Beginner Students and Experts." *Journal of English for Academic Purposes* 47, p. 100894. <https://doi.org/10.1016/j.jeap.2020.100894>.
- Li, Luan, Tingting Hu, and Shuting Liu. 2024. "Graded Phonological Neighborhood Effects on Lexical Retrieval: Evidence from Mandarin Chinese." *Journal of Memory and Language* 137, 104526. <https://doi.org/10.1016/j.jml.2024.104526>.
- Lowie, Wander. 2017. "Lost in State Space?." *Complexity theory and language development: In Celebration of Diane Larsen-Freeman* 48: 123.
- Lowie, Wander and Marjolijn Verspoor. 2015. "Variability and Variation in Second Language Acquisition Orders: A Dynamic Reevaluation." *Language Learning* 65 (1): 63–88.
- Lowie, Wander and Marjolijn Verspoor. 2019. "Individual Differences and the Ergodicity Problem." *Language Learning* 69 (S1): 184–206. <https://doi.org/10.1111/lang.12324>.
- Lucas, Margery. 2000. "Semantic Priming without Association: A Meta-analytic Review." *Psychonomic Bulletin and Review* 7: 618–30. <https://doi.org/10.3758/BF03212999>.
- Luce, Paul A, Stephen D Goldinger, Edward T Auer, and Michael S. Vitevitch. 2000. "Phonetic Priming, Neighborhood Activation, and PARSYN." *Perception and Psychophysics* 62 (3): 615–25.
- Luckman, Courtney, Stacy A Wagovich, Christine Weber, Barbara Brown, Soo-Eun Chang, Nancy E Hallet al. 2020. "Lexical Diversity and Lexical Skills in Children Who Stutter." *Journal of fluency disorders* 63: 105747. <https://doi.org/10.1016/j.jfludis.2020.105747>.
- Lüdeling, Anke. 2017. *Variationistische Korpusstudien*, 129–44. Berlin: De Gruyter. <https://doi.org/10.1515/9783110518214-009>.
- Lüdeling, Anke, Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, et al. 2024, May. *RUEG Corpus*. Version 1.0. Zenodo. <https://doi.org/10.5281/zenodo.11234583>.
- Lüdeling, Anke, Hagen Hirschmann, and Anna Shadrova. 2017. "Linguistic Models, Acquisition Theories, and Learner Corpora: Morphological Productivity in SLA Research Exemplified by Complex Verbs in German." *Language Learning* 67 (S1): 96–129. <https://doi.org/10.1111/lang.12231>.
- Nickels, Lyndsey, Leonie F. Lampe, Catherine Mason, Solène Hameau. 2022. "Investigating the Influence of Semantic Factors on Word Retrieval: Reservations, Results and Recommendations." *Cognitive Neuropsychology* 39 (3–4): 113–54 PMID: 35972430. <https://doi.org/10.1080/02643294.2022.2109958>.

- MacWhinney, Brian. 2014. "Item-based Patterns in Early Syntactic Development." *Constructions, Collocations, Patterns* 2562: 33–69. <https://doi.org/10.1515/9783110356854.33>.
- Mark Knobel, Matthew Finkbeiner and Alfonso Caramazza. 2008. "The Many Places of Frequency: Evidence for a Novel Locus of the Lexical Frequency Effect in Word Production." *Cognitive Neuropsychology* 25 (2): 256–86, PMID: 18568814. <https://doi.org/10.1080/02643290701502425>.
- Marti, Louis, Shengyi Wu, Steven T. Piantadosi, and Celeste Kidd. 2023. "Latent Diversity in Human Concepts." *Open Mind* 7: 79–92.
- Martín Villena, Fernando. 2023. *L1 Morphosyntactic Attrition at the Early Stages: Evidence from Production, Interpretation, and Processing of Subject Referring Expressions in L1 Spanish-L2 English Instructed and Immersed Bilinguals*. PhD thesis. University of Granada. <https://doi.org/https://hdl.handle.net/10481/81920>.
- Matos, Joana and Cristina Flores. 2024. "More Insights into the Interaction between Age, Exposure, and Attitudes in Language Attrition and Retention from the Perspective of Bilingual Returnees." *International Journal of Bilingualism* 28 (1): 24–42. <https://doi.org/10.1177/13670069221136941>.
- McCarthy, Philip M and Scott Jarvis. 2010. "MTLD, VOC-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment." *Behavior Research Methods* 42 (2): 381–92. <https://doi.org/10.3758/BRM.42.2.381>.
- Meara, Paul. 2005, March. "Lexical Frequency Profiles: A Monte Carlo Analysis." *Applied Linguistics* 26 (1): 32–47. <https://doi.org/10.1093/applin/amh037>.
- Medaglia, John D, Deepa M Ramanathan, Umesh M Venkatesan, and Frank G Hillary. 2011. "The Challenge of Non-ergodicity in Network Neuroscience." *Network: Computation in Neural Systems* 22 (1–4): 148–53.
- Melinger, Alissa and Christian Dobel. 2005. "Lexically-driven Syntactic Priming." *Cognition* 98 (1): B11–20. <https://doi.org/10.1016/j.cognition.2005.02.001>.
- Mellado Blanco, Carmen, Fabio Mollica, and Elmar Schafroth. 2022. *Konstruktionen zwischen Lexikon und Grammatik: Phrasem-Konstruktionen Monolingual, Bilingual und Multilingual*. Berlin, Germany: De Gruyter.
- Meylan, Stephan and Susanne Gahl. 2014. "The Divergent Lexicon: Lexical Overlap Decreases with Age in a Large Corpus of Conversational Speech." In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 36. <https://escholarship.org/uc/item/4s3757m>.
- Michaelis, Laura. 2024. "Staying Terminologically Rigid, Conceptually Open and Socially Cohesive: How to Make Room for the Next Generation of Construction Grammarians." *Constructions and Frames* 16 (2): 278–310.
- Minnaard, Liesbeth and Till Dembeck. 2014. "Introduction: How to Challenge the Myth of Monolingualism?." In *Challenging the Myth of Monolingualism*, 9–14. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789401210980_002.
- Moon, Rosamund. 1999. "Needles and Haystacks, Idioms and Corpora: Gaining Insights into Idioms, Using Corpus Analysis." *The Perfect Learners Dictionary*, 265–81. <https://doi.org/http://dx.doi.org/10.1515/9783110947021.265>.
- Nedergaard, Johanne and Gary Lupyan. 2023. "Not Everyone has an Inner Voice: Behavioral Consequences of Anendophasia." In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 45. <https://escholarship.org/uc/item/93p4r8td>.
- Neels, Jakob. 2020. "Lifespan Change in Grammaticalisation as Frequency-sensitive Automation: William Faulkner and the Let Alone Construction." *Cognitive Linguistics* 31 (2): 339–65. <https://doi.org/10.1515/cog-2019-0020>.
- Nelson, Robert. 2018. "How 'chunky' is Language? Some Estimates Based on Sinclair's Idiom Principle." *Corpora* 13 (3): 431–60. <https://doi.org/10.3366/cor.2018.0156>.
- Neth, Hansjörg and Gerd Gigerenzer. 2015. "Heuristics: Tools for an Uncertain World." In *Emerging Trends in the Social and Behavioral Sciences*, 1–18. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9781118900772.etrds0394>.
- Nurmi, Arja, Tanja Rütten, and Päivi Pahta. 2018. *Challenging the Myth of Monolingual Corpora*, Vol. 80, Leiden, The Netherlands: Brill. <https://doi.org/10.1163/9789004276697>.
- Ólafsdóttir, H Freyja, Daniel Bush, and Caswell Barry. 2018. "The Role of Hippocampal Replay in Memory and Planning." *Current Biology* 28 (1): R37–50. <https://doi.org/10.1016/j.cub.2017.10.073>.
- Ordines, Pedro Ivorra. 2023. "Spanish Comparative Constructional Idioms and Their English and French Counterparts. A Corpus-based Study." In *Technological Innovation Put to the Service of Language Learning, Translation and Interpreting: Insights from Academic and Professional Contexts*. 157–79. Berlin, Germany: Peter Lang Verlag.
- Özsoy, Onur and Frederic Blum. 2023. "Exploring Individual Variation in Turkish Heritage Speakers Complex Linguistic Productions: Evidence from Discourse Markers." *Applied Psycholinguistics* 44: 1–31. <https://doi.org/10.1017/S0142716423000267>.
- Paquot, Magali. 2019. "The Phraseological Dimension in Interlanguage Complexity Research." *Second Language Research* 35 (1): 121–45. <https://doi.org/10.1177/0267658317694221>.
- Paquot, Magali, Dana Gablasova, Vaclav Brezina, Hubert Naets, A. Lénko-Szymańska, and S. Götz. 2022. "Phraseological Complexity in EFL Learners Spoken Production Across Proficiency Levels." *Complexity, Accuracy and Fluency in Learner Corpus Research*, edited by Agnieszka Leńko-Szymańska and Sandra Götz. 115–36. Amsterdam, Netherlands: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.104.05paq>.
- Partington, Alan. 2017. "Varieties of Non-obvious Meaning in CL and CADs: From 'Hindsight Post-dictability' to Sweet Serendipity." *Corpora* 12 (3): 339–67. <https://doi.org/10.3366/cor.2017.0124>.
- Pavlova, Anna. 2020. "Und ob es phraseologie ist!." In *Russland übersetzen/Russia in Translation*, edited by Engel Christine, Irina Pohlan and Stephan Walter, 117–132. Festschrift für Birgit Menzel. Berlin: Frank and Timme.
- Pawley, Andrew and Frances Hodgetts Syder. 1983. "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency." *Language and Communication* 191: 225.

- Pecina, Pavel. 2010. "Lexical Association Measures and Collocation Extraction." *Language Resources and Evaluation* 44 (1–2): 137–58.
- Pérez-Paredes, Pascual and M Camino Bueno-Alastuey. 2019. "A Corpus-driven Analysis of Certainty Stance Adverbs: Obviously, Really and Actually in Spoken Native and Learner English." *Journal of Pragmatics* 140, 22–32. <https://doi.org/10.1016/j.pragma.2018.11.016>.
- Pescuma, Valentina N, Dina Serova, Julia Lukassek, Antje Sauermaun, Roland Schäfer, Aria Adli, et al. 2023. "Situating Language Register Across the Ages, Languages, Modalities, and Cultural Aspects: Evidence from Complementary Methods." *Frontiers in Psychology* 13: 964658. <https://doi.org/10.3389/fpsyg.2022.964658>.
- Petré, Peter and Lynn Anthonissen. 2020. "Individuality in Complex Systems: A Constructionist Approach." *Cognitive Linguistics* 31 (2): 185–212. <https://doi.org/10.1515/cog-2019-0033>.
- Piantadosi, Steven T. 2014. "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions." *Psychonomic Bulletin and Review* 21: 1112–30. <https://doi.org/10.3758/s13423-014-0585-6>.
- Pickering, Martin J and Holly P. Branigan. 1998. "The Representation of Verbs: Evidence from Syntactic Priming in Language Production." *Journal of Memory and Language* 39 (4): 633–51. <https://doi.org/10.1006/jmla.1998.2592>.
- Polišenská, Kamila, Shula Chiat, and Penny Roy. 2015. "Sentence Repetition: What Does the Task Measure?." *International Journal of Language and Communication Disorders* 50 (1): 106–18. <https://doi.org/10.1111/1460-6984.12126>.
- Poulsen, Sonja. 2022. *Collocations as a Language Resource: A Functional and Cognitive Study in English Phraseology*. Amsterdam, Netherlands: John Benjamins. <https://doi.org/10.1075/hcp.71>.
- Reitter, David. 2008. *Context Effects in Language Production: Models of Syntactic Priming in Dialogue Corpora*. Ph.D. thesis, University of Edinburgh. <https://doi.org/http://hdl.handle.net/1842/3059>.
- Reitter, David and Johanna D. Moore. 2014. "Alignment and Task Success in Spoken Dialogue." *Journal of Memory and Language* 76: 29–46. <https://doi.org/10.1016/j.jml.2014.05.008>.
- Reznicek, Marc, Maik Walter, Karin Schmidt, Anke Lüdeling, Hagen Hirschmann, Cedric Krummes, Torsten Andreas. 2012. *Das Falko-Handbuch: Korpusaufbau und Annotationen*. Technical report, Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin. https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/at_download/file.
- Roberts, Brady RT, Colin M MacLeod, and Myra A Fernandes. 2023. "Symbol Superiority: Why \$ is Better Remembered than 'dollar'." *Cognition* 238: 105435. <https://doi.org/10.1016/j.cognition.2023.105435>.
- Roelofs, Ardi. 2008. "Dynamics of the Attentional Control of Word Retrieval: Analyses of Response Time Distributions." *Journal of Experimental Psychology: General* 137 (2): 303–23. <https://doi.org/10.1037/0096-3445.137.2.303>.
- Rosch, Eleanor and Carolyn B Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7 (4): 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9).
- Rubin, Rachel, Alex Housen, and Magali Paquot. 2021. "5 Phraseological Complexity as an Index of L2 Dutch Writing Proficiency: A Partial Replication Study." In *Perspectives on the L2 Phrasicon: The View from Learner Corpora*, edited by Sylviane Granger, 101–25. Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781788924863-006>.
- Sag, Ivan A. 2012. "Sign-based Construction Grammar: An Informal Synopsis." In *Sign-based Construction Grammar*, Hans C Boas and Ivan A Sag, 69–202. Stanford: CSLI Publications.
- Sauer, Simon and Anke Lüdeling. 2016. "Flexible Multi-Layer Spoken Dialogue Corpora." *International Journal of Corpus Linguistics* 21 (3): 419–38. <https://doi.org/10.1075/ijcl.21.3.06sau>.
- Schmid, Hans-Jörg. 2015. "A Blueprint of the Entrenchment-and-conventionalization Model." *Yearbook of the German Cognitive Linguistics Association* 3 (1): 3–26.
- Schmid, Hans-Jörg. 2018. "Unifying Entrenched Tokens and Schematized Types as Routinized Commonalities of Linguistic Experience." *Yearbook of the German Cognitive Linguistics Association* 6 (1): 167–82.
- Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. USA: Oxford University Press.
- Schmid, Hans-Jörg and Scott Jarvis. 2014. "Lexical Access and Lexical Diversity in First Language Attrition." *Bilingualism: Language and Cognition* 17 (4): 729–48. <https://doi.org/10.1017/S1366728913000771>.
- Schmidt, Jürgen Erich and Joachim Herrgen. 2011. *Sprachdynamik. Eine Einführung in die moderne Regionalsprachenforschung*. Berlin: Erich Schmidt.
- Schmitt, Norbert, ed. 2004. *Formulaic Sequences: Acquisition, Processing and Use*, Volume 9 of Language Learning and Language Teaching, Amsterdam: John Benjamins.
- Scully, Iiona D, Lucy E Napper, and Almut Hubach. 2017. "Does Reactivation Trigger Episodic Memory Change? A Meta-analysis." *Neurobiology of Learning and Memory* 142: 99–107. <https://doi.org/10.1016/j.nlm.2016.12.012>.
- Sedelow, Sally Yeates. 1985. "Computational Lexicography." *Computers and the Humanities* 19 (2): 97–101.
- Shadrova, Anna. 2020. *Measuring Coselectional Constraint in Learner Corpora: A Graph-based Approach*, Ph.D. thesis, Humboldt-Universität zu Berlin, Sprach- und literaturwissenschaftliche Fakultät. <https://doi.org/http://dx.doi.org/10.18452/21606>.
- Shadrova, Anna. 2021. "Topic Models Do not Model Topics: Epistemological Remarks and Steps Towards Best Practices." *Journal of Data Mining and Digital Humanities* 2021: 1–28. <https://doi.org/10.46298/jdmdh.7595>.
- Shadrova, Anna. 2022. "It may be in the Structure, not the Combinations: Graph Metrics as an Alternative to Statistical Measures in Corpus-linguistic research." In A. Kuczera and F. Diehr, *Proceedings of Graph Technologies in the Humanities 2020*, 245–278. <https://doi.org/CEUR-WS.org/Vol-3110/paper12.pdf>.
- Shadrova, Anna. 2024. "Ein graphbasierter Ansatz zur Untersuchung usueller Wortverbindungen in der L2 Deutsch." *Deutsch als Fremdsprache* 2024 (3): 131–41. <https://doi.org/10.37307/j.2198-2430.2024.03.02>.

- Shadrova, Anna, Anke Lüdeling, and Rahel Gajaneh Hartz (in prep.). *Bilingualism Modulates Complex Verb Use in a Task-based Corpus of 11 German*.
- Shadrova, Anna, Pia Linscheid, Julia Lukassek, Anke Lüdeling, and Sarah Schneider. 2021. "A Challenge for Contrastive L1/L2 Corpus Studies: Large Inter- and Intra-individual Variation Across Morphological, but not Global Syntactic Categories in Task-based Corpus Data of A Homogeneous L1 German Group." *Frontiers in Psychology* 12: 5267. <https://doi.org/10.3389/fpsyg.2021.716485>.
- Siepmann, Dirk. 2012. "Sinclair Revisited: Beyond Idiom and Open Choice." *The Phraseological View of Language: A Tribute to John Sinclair*, edited by Thomas Herbst, Susen Faulhaber and Peter Uhrig, 59–86. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110257014.59>.
- Silvennoinen, Olli O. 2023. "Is Construction Grammar Cognitive?." *Constructions* 15 (1): 1–17.
- Sinclair, Arabella and Raquel Fernández. 2023. "Alignment of Code Switching Varies with Proficiency in Second Language Learning Dialogue." *System* 113, 102952. <https://doi.org/10.1016/j.system.2022.102952>.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford, United Kingdom: Oxford University Press.
- Singleton, David and Justyna Leśniewska. 2021. "Phraseology: Where Lexicon and Syntax Conjoin." *Research in Language and Education: An International Journal [RILE]* 1 (1): 46–58. https://rile.yyu.edu.tr/article_389.html.
- Siyanova-Chanturia, Anna. 2015. "On the "Holistic" Nature of Formulaic Language." *Corpus Linguistics and Linguistic Theory* 11 (2): 285–301.
- Siyanova-Chanturia, Anna and Stefania Spina. 2020. "Multi-word Expressions in Second Language Writing: a Large-scale Longitudinal Learner Corpus Study." *Language Learning* 70 (2): 420–63.
- Smolka, Eva, Gary Libben, and Wolfgang U Dressler. 2019. "When Morphological Structure Overrides Meaning: Evidence From German Prefix and Particle Verbs." *Language, Cognition and Neuroscience* 34 (5): 599–614. <https://doi.org/10.1080/23273798.2018.1552006>.
- Smolka, Eva, Katrin H Preller, and Carsten Eulitz. 2014. "Verstehen (understand) Primes 'Stehen'('Stand'): Morphological Structure Overrides Semantic Compositionality in the Lexical Representation of German Complex Verbs." *Journal of Memory and Language* 72, 16–36. <https://doi.org/10.1016/j.jml.2013.12.002>.
- Song, Jongmin. 2008. "Lexical Productivity of Korean EFL Learners in Spoken and Written English." *English Teaching* 63 (2): 167–91.
- Speelman, Dirk, Kris Heylen, and Dirk Geeraerts. 2018. *Mixed-effects Regression Models in Linguistics*. Cham: Springer.
- Sprenger, Simone A, Sara D Beck, and Andrea Weber. 2024. "What Fires Together, Wires Together: The Effect of Idiomatic Co-occurrence on Lexical Networks." *Languages* 9 (3): 105.
- Steels, Luc. 2000. "Language as a Complex Adaptive System." In *International Conference on Parallel Problem Solving from Nature*, 17–26. Berlin, Heidelberg: Springer.
- Steels, Luc. 2013. "Fluid Construction Grammar." In *The Oxford Handbook of Construction Grammar*. Oxford (UK): Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0009>.
- Steels, Luc and Martin Loetzsch. 2006. *Perspective Alignment in Spatial Language*. arXiv: <http://arXiv.org/abs/cs/0605012>.
- Stefanowitsch, Anatol and Susanne Flach. 2017. *The Corpus-Based Perspective on Entrenchment*, 101–28. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110341423-006>.
- Stefanowitsch, Anatol and Stefan Th Gries. 2003. "Collostructions: Investigating the Interaction of Words and Constructions." *International Journal of Corpus Linguistics* 8 (2): 209–43.
- Stella, Massimo, Nicole M Beckage, and Markus Brede. 2017. "Multiplex Lexical Networks Reveal Patterns in Early Word Acquisition in Children." *Scientific Reports* 7 (1): 46730. <https://doi.org/10.1038/srep46730>.
- Stumpf, Sören. 2016. "Modifikation oder Modellbildung? Das ist hier die Frage – Abgrenzungsschwierigkeiten zwischen modifizierten und modellartigen Phrasemen am Beispiel formelhafter (Ir-)Regularitäten." *Linguistische Berichte* 2016 (247): 317–42.
- Sturdy, Christopher B and Elena Nicoladis. 2017. "How Much of Language Acquisition Does Operant Conditioning Explain?." *Frontiers in Psychology* 8: 01918. <https://doi.org/10.3389/fpsyg.2017.01918>.
- Sullivan, Margot D, Gregory J. Poarch, and Ellen Bialystok. 2018. "Why is Lexical Retrieval Slower for Bilinguals? Evidence From Picture Naming." *Bilingualism: Language and Cognition* 21 (3): 479–88. <https://doi.org/10.1017/S1366728917000694>.
- Szmrecsanyi, Benedikt. 2006. "Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis." Vol. 177 of *Trends in Linguistics. Studies and Monographs*. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110197808>.
- Szmrecsanyi, Benedikt. 2017. "Variationist Sociolinguistics and Corpus-based Variationist Linguistics: Overlap and Cross-pollination Potential." *Canadian Journal of Linguistics/Revue canadienne de linguistique* 62 (4): 685–701. <https://doi.org/10.1017/cnj.2017.34>.
- Szmrecsanyi, Benedikt. 2019. "Register in Variationist Linguistics." *Register Studies* 1 (1): 76–99. <https://doi.org/10.1075/rs.18006.szm>.
- Tavakoli, Parvaneh and Takumi Uchiyara. 2019. "To What Extent are Multiword Sequences Associated with Oral Fluency?." *Language Learning* 70 (2): 506–47. <https://doi.org/10.1111/lang.12384>.
- Tomasello, Michael. 1995. "Language is not an Instinct." *Cognitive Development* (10): 131–56.
- Tomasello, Michael. 2000. "The Item-based Nature of Children's Early Syntactic Development." *Trends in Cognitive Sciences* 4 (4): 156–63. [https://doi.org/10.1016/S1364-6613\(00\)01462-5](https://doi.org/10.1016/S1364-6613(00)01462-5).
- Tomasello, Michael. 2009. *Constructing a Language. A Usage-based Theory of Language Acquisition*. Cambridge, USA: Harvard University Press.
- Torregrossa, Jacopo, Cristina Flores, and Esther Rinke. 2023. "What Modulates the Acquisition of Difficult Structures in a Heritage Language? A Study on Portuguese in Contact with French, German and Italian." *Bilingualism: Language and Cognition* 26 (1): 179–92. <https://doi.org/10.1017/S1366728922000438>.

- Tsehay, Wintai, Tatiana Pashkova, Rosemarie Tracy, and Shanley EM Allen. 2021. "Deconstructing the Native Speaker: Further Evidence from Heritage Speakers for Why This Horse Should Be Dead!" *Frontiers in Psychology* 12: 717352. <https://doi.org/10.3389/fpsyg.2021.717352>.
- Tversky, Amos and Daniel Kahneman. 1974. "Judgement Under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking Under Uncertainty." *Science* 185 (4157): 1124–31. <https://doi.org/10.1126/science.185.4157.1124>.
- Uhrig, Peter. 2015. "Why the Principle of No Synonymy is Overrated." *Zeitschrift für Anglistik und Amerikanistik* 63 (3): 323–37. <https://doi.org/10.1515/zaa-2015-0030>.
- Underwood, Geoffrey, Norbert Schmitt, and Adam Galpin. 2004. "The Eyes have It." *Formulaic Sequences: Acquisition, Processing, and Use* 9, 153.
- Unger, Layla, Hyungwook Yim, Olivera Savic, Simon Dennis, and Vladimir M Sloutsky. 2023. "No Frills: Simple Regularities in Language can Go a Long Way in the Development of Word Knowledge." *Developmental Science* 26 (4): e13373.
- Ungerer, Tobias. 2024. "Vertical and Horizontal Links in Constructional Networks: Two Sides of the Same Coin?." *Constructions and Frames* 16 (1): 30–63. <https://doi.org/10.1075/cf.22011.ung>.
- van Trijp, Remi. 2024. "Nostalgia for the Future of Construction Grammar." *Constructions and Frames* 16 (2): 311–45. <https://doi.org/10.1075/cf.23013.van>.
- Vandeweerd, Nathan, Alex Housen, and Magali Paquot. 2022. "Comparing the Longitudinal Development of Phraseological Complexity Across Oral and Written Tasks." In *Studies in Second Language Acquisition* 1–25. <https://doi.org/10.1017/S0272263122000389>.
- Verspoor, Marjolijn, Wander Lowie, and Kees de Bot. 2021. "Variability as Normal as Apple Pie." *Linguistics Vanguard* 7 (s2), 20200034. <https://doi.org/10.1515/lingvan-2020-0034>.
- Vitevitch, Michael S. 2022. "What can Network Science Tell Us About Phonology and Language Processing?." *Topics in Cognitive Science* 14 (1): 127–42. <https://doi.org/10.1111/tops.12532>.
- Vitevitch, Michael S., Kit Ying Chan, and Rutherford Goldstein. 2014. "Insights into Failed Lexical Retrieval from Network Science." *Cognitive Psychology* 68: 1–32. <https://doi.org/10.1016/j.cogpsych.2013.10.002>.
- Vogt, Anne, Barbara Kaup, and Rasha Abdel Rahman. 2023. "Experience-driven Meaning Affects Lexical Choices During Language Production." *Quarterly Journal of Experimental Psychology* 76 (7): 1561–84, PMID: 36062350. <https://doi.org/10.1177/17470218221125425>.
- Vyacheslavovna, Pyzhak Julia and Vinogradova Olga Il'ichna. 2022. "Word-formation Complexity: A Learner Corpus-based Study." *Russian Journal of Linguistics* 26 (2): 471–92. <https://doi.org/10.22363/2687-0088-31187>.
- Vyatkina, Nina, Hagen Hirschmann, and Felix Golcher. 2015. "Syntactic Modification at Early Stages of L2 German Writing Development: A Longitudinal Learner Corpus Study." *Journal of Second Language Writing* 29: 28–50. <https://doi.org/10.1016/j.jslw.2015.06.006>.
- Vygotsky, Lev Semenovich. 1962. *Thought and Language*. Massachusetts: MIT Press.
- Watson, Matthew E, Martin J Pickering, and Holly P Branigan. 2004. "Alignment of Reference Frames in Dialogue." In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 26. <https://doi.org/https://escholarship.org/uc/item/9t95b48d>.
- Weigel, Aubrey V, Blair Simon, Michael M Tamkun, and Krapf Krapf. 2011. "Ergodic and Nonergodic Processes Coexist in the Plasma Membrane as Observed by Single-molecule Tracking." *Proceedings of the National Academy of Sciences* 108 (16): 6438–43.
- Weinert, Regina. 1995. "The Role of Formulaic Language in Second Language Acquisition: A Review." *Applied Linguistics* 16 (2): 180–205.
- Wiese, Heike. 2020. "Language Situations: A Method for Capturing Variation within Speakers Repertoires." In *Methods in Dialectology XVI*, Yoshiyuki Asahi, Volume 59 of Bamberg Studies in English Linguistics, Frankfurt a.M., 105–17. Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko et al. 2022. "Heritage Speakers as Part of the Native Language Continuum." *Frontiers in Psychology* 12: 717973. <https://doi.org/10.3389/fpsyg.2021.717973>.
- Williams, Jake Ryland, Paul R Lessard, Suma Desu, Eric M Clark, James P Bagrow, Christopher M Danforth, and Peter Sheridan Dodds. 2015. "Zipf's Law Holds for Phrases, Not Words." *Scientific Reports* 5: 12209. <https://doi.org/10.1038/srep12209>.
- Wills, Peter and François G Meyer. 2020. "Metrics for Graph Comparison: A Practitioner's Guide." *Plos One* 15 (2): e0228728.
- Wittgenstein, Ludwig. 1953. *Ludwig Wittgenstein Werkausgabe, Volume "Band 1"*. Suhrkamp, Frankfurt am Main.
- Wood, David. 2015. *Fundamentals of Formulaic Language: An Introduction*. London, United Kingdom: Bloomsbury Publishing.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511519772>.
- Wray, Alison. 2012. "What Do We (think We) Know About Formulaic Language? An Evaluation of the Current State of Play." *Annual Review of Applied Linguistics* 32: 231–54. <https://doi.org/10.1017/S026719051200013X>.
- Wray, Alison. 2013. "Formulaic Language." *Language Teaching* 46 (3): 316–34. <https://doi.org/10.1017/S0261444813000013>.
- Wray, Alison and Michael R Perkins. 2000. "The Functions of Formulaic Language: An Integrated Model." *Language and Communication* 20 (1): 1–28.
- Wulff, Dirk U, Simon De Deyne, Samuel Aeschbach, and Rui Mata. 2022. "Using Network Science to Understand the Aging Lexicon: Linking Individuals' Experience, Semantic Networks, and Cognitive Performance." *Topics in Cognitive Science* 14 (1): 93–110.
- Wulff, Dirk U, Thomas T Hills, Margie E Lachman, and Rui Mata. 2016. "The Aging Lexicon: Differences in the Semantic Networks of Younger and Older Adults." In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 38, <https://escholarship.org/uc/item/0jz483dm>.
- Wulff, Stefanie. 2008. *Rethinking Idiomaticity: A Usage-based Approach*. London: A&C Black.
- Wulff, Stefanie. 2020. "Usage-based Approaches." In *The Routledge Handbook of Second Language Acquisition and Corpora*, 175–88. Oxfordshire: Routledge.

- Wulff, Stefanie and Stefan Th Gries. 2021. *Exploring Individual Variation in Learner Corpus Research: Methodological Suggestions*, 191–213. Cambridge Applied Linguistics. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108674577.010>.
- Zeldes, Amir. 2012. *Productivity in Argument Selection: From Morphology to Syntax*, Volume 260 of Trends in Linguistics. Studies and Monographs. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110303919>.
- Zeldes, Amir. 2013. “Komposition als Konstruktionsnetzwerk im fortgeschrittenen L2-Deutsch.” *Zeitschrift für germanistische Linguistik* 41 (2): 240–76. <https://doi.org/10.1515/zgl-2013-0014>.
- Zerbian, Sabine, Marlene Böttcher, and Yulia Zuban. 2022. “Prosody of Contrastive Adjectives in Mono-and Bilingual Speakers of English and Russian: A Corpus Study.” In *Proceedings 11th International Conference on Speech Prosody. International Speech Communication Association (ISCA) Online Archive, accepted*. <https://doi.org/10.21437/SpeechProsody.2022-165>.
- Ziem, Alexander. 2018. “Construction Grammar Meets Phraseology: Eine Standortbestimmung.” *Linguistik online* 90 (3): 3–19.
- Zinsmeister, Heike, Marc Reznicek, Julia Ricart Brede, Christina Rosén, and Dirk Skiba. 2012. “Das wissenschaftliche Netzwerk Kobalt-DaF.” *Zeitschrift für germanistische Linguistik* 40 (3): 457–8. <https://doi.org/10.1515/zgl-2012-0030>.
- Zykova, Irina V. 2016. “Linguo-cultural Studies of Phraseologisms in Russia: Past and Present.” *Yearbook of Phraseology* 7 (1): 127–48. <https://doi.org/10.1515/phras-2016-0007>.

Appendix

Table A1: Speaker RUmo27MR's four contributions to the RUEG corpus

Setting	Text	Translation
Formal spoken	здрасте код дела ф шестнадцать. сегодня, двадцать четвёртого ноября, я был свидетелем дорожно-транспортного происшествия. выйдя из метро я надрывался по тротуар... по тротуару в сторону работы, как вдруг услышав лай собаки, на что я обратил внимание, и после этого последовал звук удара машины. звук был громкий, а я был немного напуган, после чего оценив ситуацию я подошёл к месту ДТП. осмотрев всё я так так скажем убедился в том, что все все были целы, все были живы, всё ограничилось с... с... ограничилось поломкой машины, после чего два водителя набрали ноль два для... для продолжения выяснения дальнейших обстоятельств. ну, кто был причиной ДТП назвать не могу, так как в первое время я ничего не видел. всё происходило очень и... очень быстро. всё, до свидания	Hello, code of incident F16. Today, on the 24th of November I was the witness of a traffic incident. As I was leaving the metro, I was moving on the sidewalk towards my workplace, when I suddenly heard a dog's bark which drew my attention, followed by the sound of a car crash. The sound was loud and I was a little bit scared, after which upon assessing the situation I stepped up to the place of the accident. Having assessed everything you could say I made sure that everyone was in one piece, everyone was alive, there was limited damage to the car, after which two drivers called 02 for the clarification of the further circumstances. About who was the cause for the accident I cannot tell since I did not see at first, everything happened very and very fast. That's it, goodbye.
Formal written	Я направлялся на работу по пешеходному тротуару, как вдруг слева от меня я услышал громкий лай пса. После этого последовал звук трения шин машины и характерный громкий удар металла. всё произошло очень быстро и неожиданно, немного постояв на месте, оценивая ситуацию, подошёл к месту ДТП. Я боялся, что могли быть пострадавшие и был готов оказать необходимую помощь. к счастью всё обошлось	I was on my way to work on the sidewalk, when to my left I heard the loud bark of a dog. After this followed the sound of screeching car wheels and the characteristic sound of a metal crash. Everything happened very quickly and unexpectedly, after waiting for a bit to assess the situation, I went to the place of the accident. I was worried that there might be casualties and was ready to offer the necessary help. Luckily that was not necessary
Informal spoken	салам алейкум брат как у тебя дела слушай тут такое произошло прям буквально минут десять назад шёл я на работу вышел с метро уже почти дохожу и буквально мимо меня проходил парень и девушка с коляской завернули и там такой перекрёсток не очень хороший и такой звук был прям но тут-то получается столкновение. я испугался что там кто-то пострадал или может сбили коляску благо обошлось там получается собака начала лаять и за счёт этого водитель дал по тормозам но вот больше меня испугал вот этот звук столкновения и прям я уже ожидал худшего так как ещё сам врач я уже был наготове наготове помогать поэтому прям руки дрожат поэтому нужно быть как-то поосторожнее	Selam Aleikum brother, how are you doing? Listen, something happened here just ten minutes ago. I was going to work, exited the subway, and I was almost there and right next to me a guy and a young woman with a stroller walked by. They turned and there was a not-so-good crossroads, and there was such a sound, like an actual crash. I was worried that somebody would be hurt or maybe they hit the stroller, but thank god it didn't happen. So apparently a dog started barking and because of that a driver hit the brakes, but what startled me more was the sound of the crash and I was right there expecting the worst. Since I am a doctor myself, I was getting ready to help, that's why my hands are shaking and people should somehow be more careful
Informal written	Салам Алейкум Как дела? Не забыл, что в Воскресенье у меня на обед собираемся?) Прикинь, что сегодня произошло... По дороге на работу, на моих глазах, чуть было человека машина не сбила. Благо всё обошлось, вовремя успел дать по тормозам. Сзади в бампер прилетела другая машина, не сильно. Я сразу подбежал, испугался, что кто-то пострадал... Благо всё хорошо, все живы здоровы. Так что брат, нужно быть осторожным...	Selam Aleikum, how are you? Do you remember that we are getting together for lunch at my place on Sunday? :) can you imagine what happened today... On my way to work, right before my eyes, somebody almost got hit by a car. Thank god it didn't happen, (the driver) hit the brakes right on time. Another car flew into his bumper, not strongly. I ran up right away, was worried that someone had suffered (an injury)... Thank God everything is good, everybody is alive and well. So, brother, you have to be careful...

The speaker uses unique referents in each setting while leaving out almost the entire scene in three out of four conditions. Forty-three verbs occur in only one of the settings. Punctuation was added to the spoken contributions for better legibility.

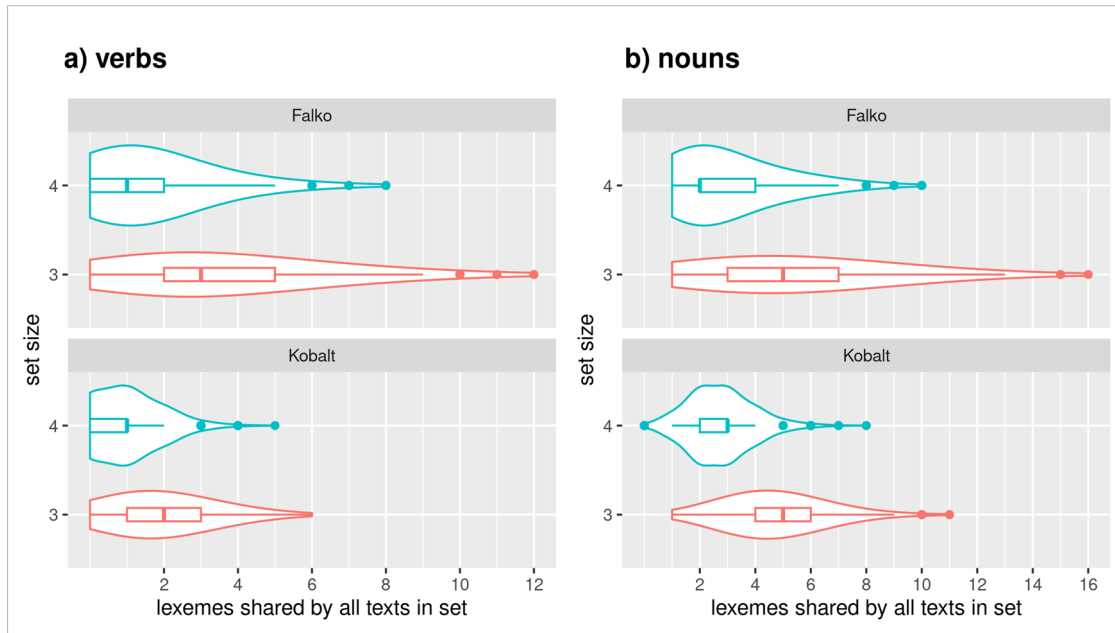


Figure A1: Distribution of shared verb (a) and noun (b) lexemes in sets of three and four texts from Falko and Kobalt (essay). Nouns are shared at a higher rate in both corpora and both set sizes. Verbs are rarely shared in either set size or corpus (median of one in sets of four texts, median of two and three respectively in sets of three). More nouns are shared in Kobalt, while more verbs are shared in Falko.

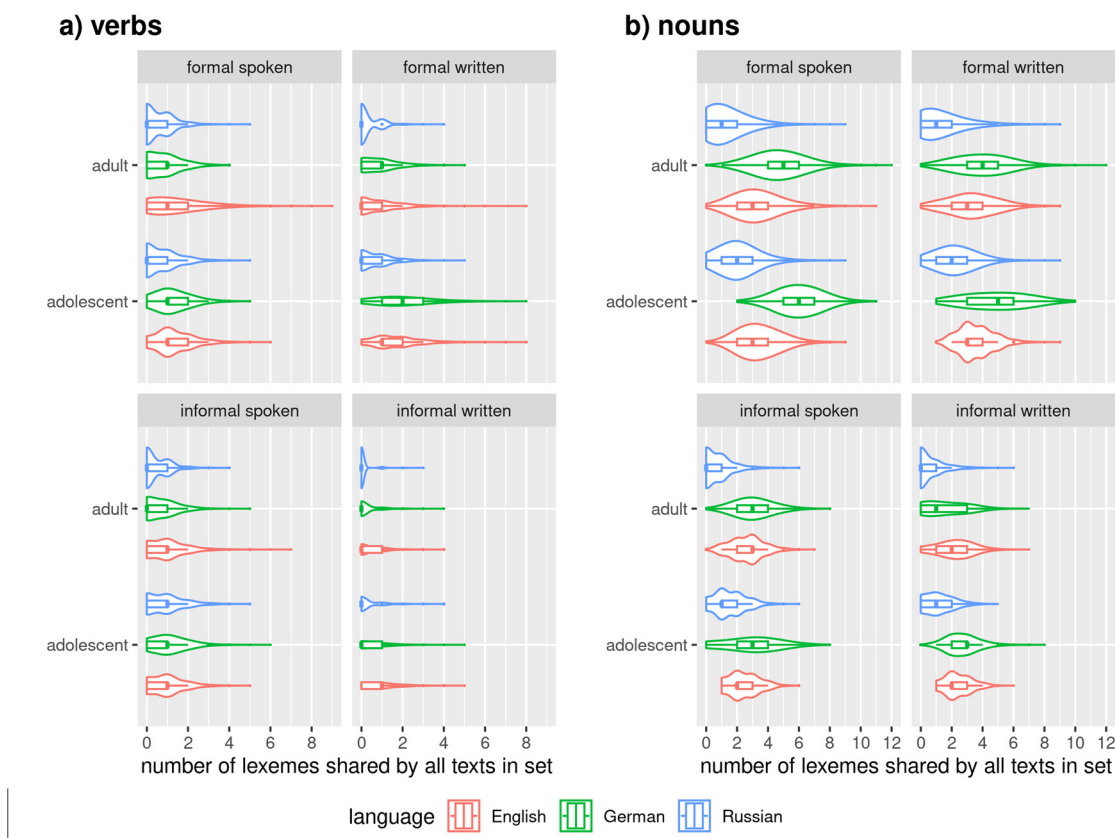


Figure A2: Distribution of shared verb (a) and noun (b) lexemes in sets of four texts from RUEG (witness report). Very few sets share more than 3 verbs across conditions and languages, with many sharing none. Nouns are shared at a higher rate than verbs. German and English speakers share more nouns in the formal condition.

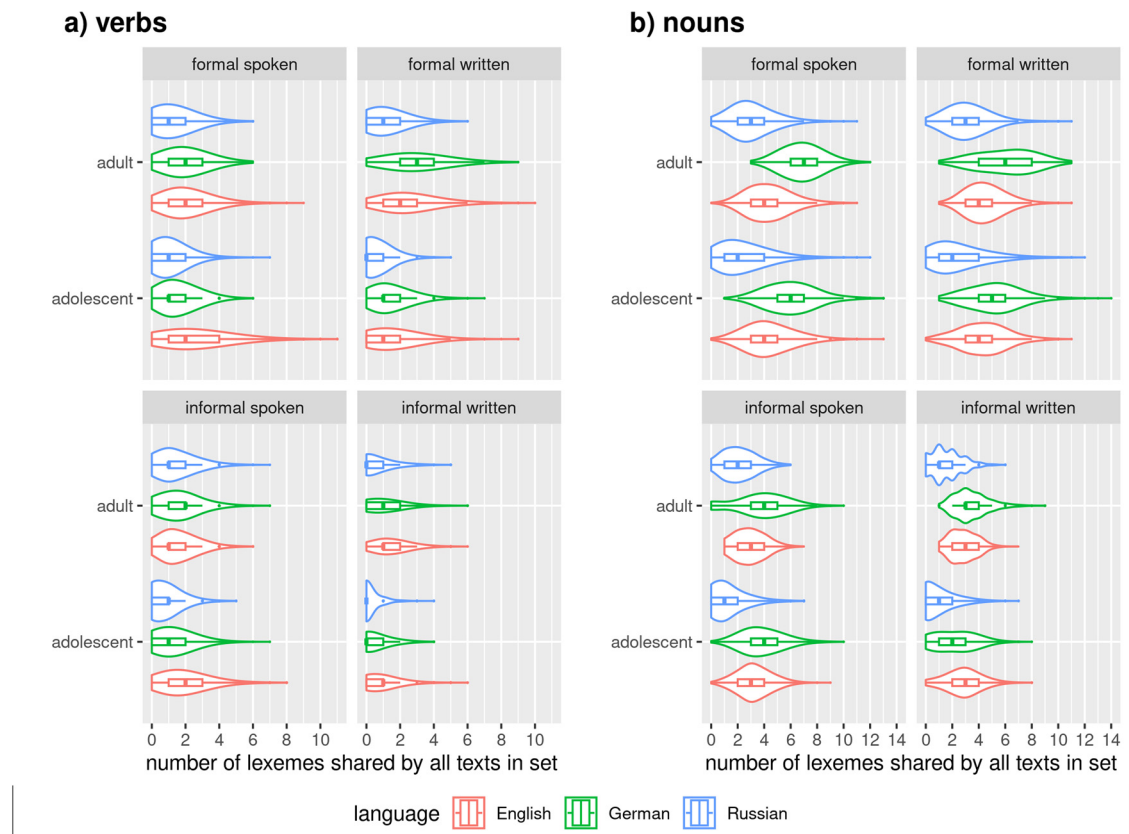


Figure A3: Distribution of shared verb (a) and noun (b) lexemes in sets of **three** texts from RUEG (witness report). Few sets share more than 4 verbs across conditions and languages, with many sharing none. Nouns are shared at a higher rate than verbs. German speakers share more nouns in the formal conditions. German speakers also share more verbs in the formal written condition.

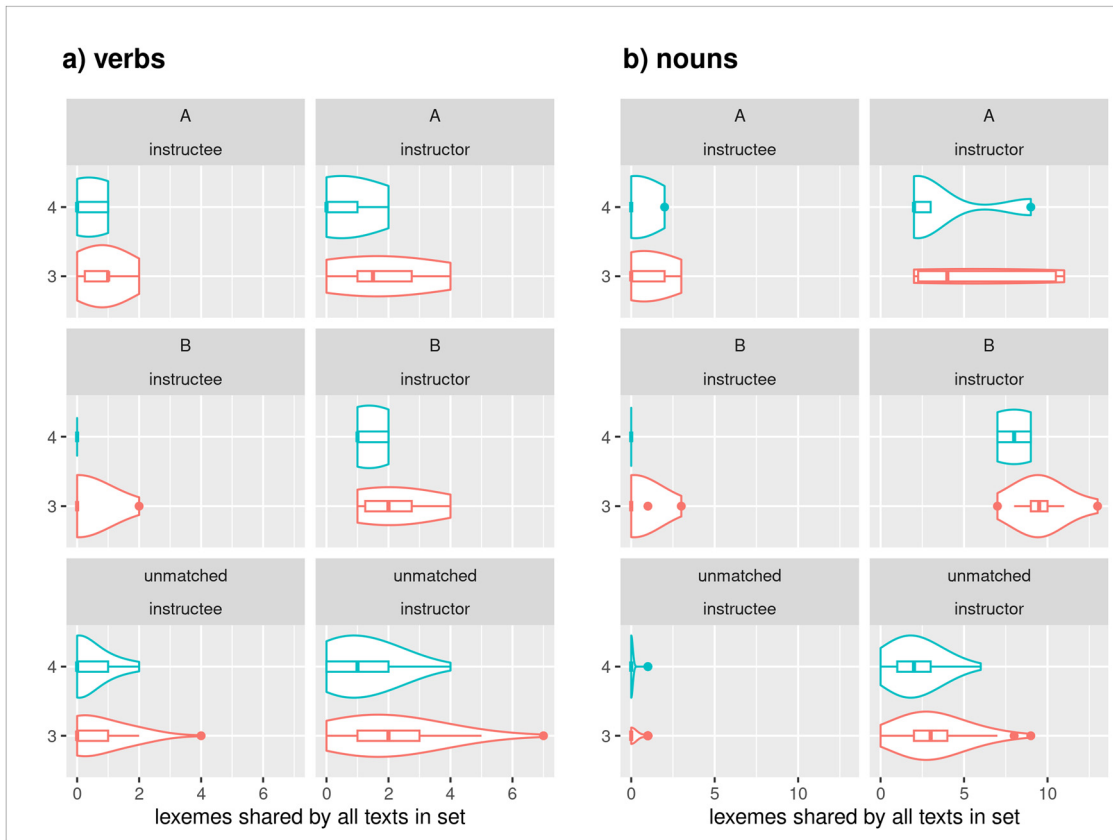


Figure A4: Distribution of shared verb (a) and noun (b) lexemes in sets of three and four texts from BeMaTaC (map task). Instructors produce most of the language and thus also share more lexemes. Few sets share more than two verb lexemes. In condition A, many sets only share few noun lexemes. Some sets share almost as many noun lexemes in the unmatched condition (different maps) as in condition A.

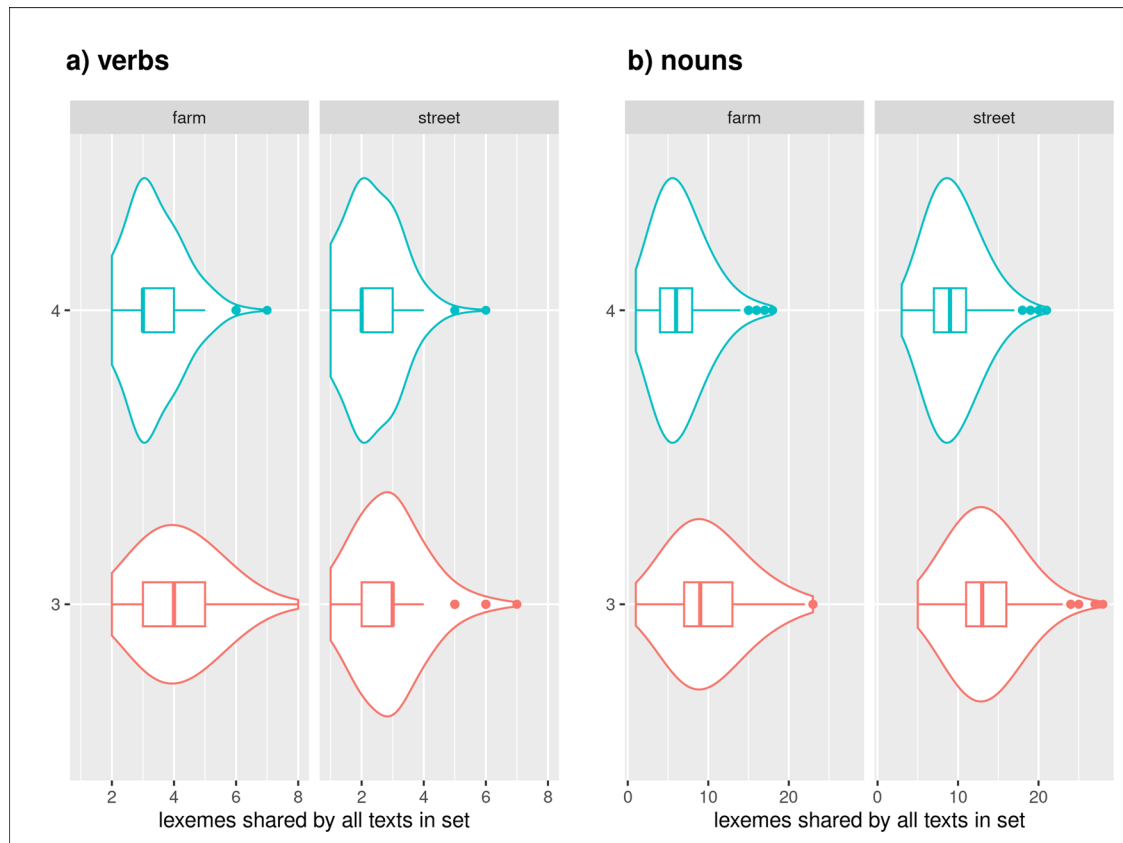


Figure A5: Distribution of shared verb (a) and noun (b) lexemes in sets of three and four texts from BeDiaCo (diapix dialogues). Lexemes are shared unequally between the two pictures (diapix *farm* and diapix *street*).