Stefan Hartmann & Tobias Ungerer

Chaos Theory, Shmaos Theory

Creativity and Routine in English *shm*-Reduplication

Abstract: This paper investigates an "extravagant" construction at the interface of morphology and syntax: English *shm*-reduplication, a pattern in which a word is immediately repeated, but the initial consonant or consonant cluster is either replaced by /ʃm/, or /ʃm/ is added to the beginning of a word if it begins with a vowel. So far, research on *shm*-reduplication has been limited to small samples of the construction and has mainly focused on its phonological and morphosyntactic properties rather than its semantics. The present study aims at filling this gap with a corpus-based analysis of a larger dataset from the web corpus ENCOW-16AX. Our findings suggest that *shm*-reduplication combines elements of routine and creativity: While the pattern is largely unconstrained in its semantics, its unusual syntactic profile and tight connection with specific communicative contexts mark it as an innovative and extravagant construction.

1 Introduction

Recent years have seen increased interest in so-called expressive morphology, i.e., morphology that "is associated with an expressive, playful, poetic, or simply ostentatious effect of some kind" (Zwicky and Pullum 1987: 335). Examples include pejorative compound patterns (e.g., German *rattenscharf* 'great/attractive', lit. 'rat sharp'; Meibauer 2013), *doubler-upper* compounds (e.g., *stayer-onner-for-nower*; Lensch 2018), libfixes (e.g., *safety-o-cracy*; Norde and Sippach 2019), and pseudo-participles (e.g., German *bebrillt* 'be-glassed', Kempf and Hartmann 2022). Alternatively, such phenomena have been characterized as instances of linguistic "extravagance," a term popularized by Haspelmath (1999) to describe speakers' use of noticeable language to stand out (see also Ungerer and Hartmann 2020). Another particularly interesting pattern in this regard is English *shm*-reduplication, as exemplified in (1).

(1) a. *And I did, and didn't actually say anything, just sort of nn uh yuh un uh.* ***Language, schmanguage****.*
(http://www.guardian.co.uk/theobserver/2000/sep/03/features.magazine37, ENCOW)

https://doi.org/10.1515/9783111321905-011

b. ***Peer review, schmeer review**. Good article!*[1]
(http://www.labspaces.net/blog/profile/624/Odyssey, ENCOW)

In this reduplicative pattern, a word is immediately repeated, but the initial consonant or consonant cluster (if any) is replaced by /ʃm/ (McCarthy and Prince 1986; Nevins and Vaux 2003; Mattiello 2013: 153). In writing, the cluster appears as <shm> or <schm>. So far, most research on English *shm*-reduplication has focused on its phonological and syntactic properties (e.g., Nevins and Vaux 2003; Grohmann and Nevins 2004; Saba Kirchner 2010; Kołłątaj 2016; but see Mattiello 2013 for a discussion of semantic and pragmatic aspects). Previous studies also had to rely on a fairly limited amount of authentic data as the construction does not occur very often in standard language data that make up for the bulk of widely-used corpora. For instance, a quick case-insensitive search for words starting with *shm-* and *schm-* in the 100-million-word British National Corpus (BNC) reveals that the construction is not attested there at all (although there are some *shm*-attestations without reduplication, e.g., *schmendrick* or *he schmiled*); and while Mattiello's (2013) study is based on data from multiple dictionaries and neologism databases, she could only work with 22 attestations of *shm*-reduplication (Mattiello 2013: 8). However, web corpora like ENCOW (Schäfer and Bildhauer 2012; Schäfer 2015) make available large amounts of non-standard language data that allow for investigating low-frequency phenomena like *shm*-reduplication in more detail.

Against this background, the aim of the present paper is twofold. On the one hand, we will present a corpus-based account of the morphosyntactic and semantic properties of *shm*-reduplication. On the other hand, and on a more theoretical level, we will discuss the "extravagant" potential of the pattern against the backdrop of the trade-off between "creativity" and "routine". The remainder of this paper is structured as follows. Section 2 gives an overview of previous research on *shm*-reduplication and discusses how it can be approached from the theoretical perspective of Construction Grammar. Section 3 introduces the data sources and presents the descriptive results of our corpus study. Section 4 links the corpus results to a wider theoretical discussion of how creativity and routine interact in giving rise to the extravagant nature of *shm*-reduplication. Section 5 provides a summary and conclusion.

1 We would like to take this opportunity to thank the anonymous reviewers as well as the editors of the present volume for their helpful feedback. The usual disclaimers apply (in other words: as for any remaining shortcomings – errors, shmerrors).

2 The English *shm*-Reduplication Construction

As Nevins and Vaux (2003) and Finkbeiner, Meibauer, and Wiese (2016: 4) point out, *shm*-reduplication seems to have its origin in Yiddish, although it is a matter of debate whether it originated in Yiddish or whether it is an English-internal development based on "the numerous Yiddish words beginning with this cluster" (OED 2023, "schm-"; also see Southern 2005 for an in-depth study of the pattern's origins). Nevins and Vaux (2003) follow Weinreich (1980: 623–624) in assuming that the construction dates back several centuries in Yiddish; according to the OED, formations with *shm-* are found from the 1920s onwards.

The sound combination /ʃm/ seems to be closely associated with words of Yiddish origin with a slightly negative or even strongly derogatory meaning component. Compare Table 1, which lists the lemmas of all word forms beginning with the grapheme combination <schm> or <shm> (with lowercase <s> to exclude proper names) in the BNC. According to their OED definitions, most of these words have pejorative connotations: *schmaltz*, for instance, is used to describe "[e]xtreme or excessive sentimentality" (OED 2023, "schmaltz"), and *schmuck* refers to a "a stupid or foolish person" (OED 2023, "schmuck"). Given these emotive-evaluative meaning components, the sound combination /ʃm/ qualifies as a phonæstheme (Firth 1930), i.e., a frequently recurring sound-meaning pattern on the submorphemic level (see, e.g., Bergen 2004; Stroebel 2017).

Table 1: Words beginning with <shm-> or <schm-> in the BNC.

Frequency	Lemma
13	*schmaltz*
9	*schmuck*
8	*schmooze*
5	*schmaltzy*
2	*schmall, schmeichal, schmutter, shmuck*
1	*schmaltze, schmecker, schmendrick, schmidt, schmile, schmoe, schmoozer, shmaltz, shmatte, shmeckle, shmoozing*

In line with the semantics of its Yiddish-derived base words, *shm*-reduplication is typically connected with "mockery and ridiculing" (Kołłątaj 2016: 243) and "derogatoriness" (Mattiello 2013: 46), and it has been described as "ironic" (Inkelas and Zoll 2005: 42). According to Mattiello (2013: 153), "this type of construction generally shows a dismissive usage, but can also be employed to downplay a situation or problem that is potentially overwhelming or threatening, or to lighten a situation with humour".

shm-reduplication is a special case of echo reduplication (Grohmann and Nevins 2004), which in turn is a special case of reduplication, a phenomenon that is widespread in the languages of the world (Inkelas 2014: 169) but relatively rare in Indo-European languages (Schwaiger 2015: 478). This makes exceptions such as identical constituent compounding (*salad-salad,* Hohenhaus 2004; Finkbeiner 2012; Frankowsky 2022) or ablaut reduplication (*riff-raff, tip-top*, Minkova 2002) all the more salient. Reduplication shows a wide range of functions that have been characterized in terms of a radial network by Regier (1994), e.g., signaling smallness, affection, contempt, or intensity. As such, many reduplicative patterns can be seen as key examples of evaluative morphology in the sense of Grandi and Körtvelyéssy (2015), who argue that the function of patterns that have been described as "expressive" or "evaluative" morphology can be characterized along two parameters: descriptive or quantitative evaluation on the one hand (diminution and augmentation), and qualitative evaluation on the other (e.g., intensification, endearment, contempt; also see Grandi 2017). Many evaluative patterns can fulfill several of these functions in different contexts – for example, diminutives, in many languages, can express not only smallness but also endearment and contempt, alongside a broad variety of other meanings (Jurafsky 1996). Although echo reduplication tends to be associated with a smaller range of meanings (Inkelas 2014: 184), even the semantically fairly homogeneous pattern of *shm*-reduplication can have a range of different functions, as we will discuss in more detail below.

The functions of *shm*-reduplication, as of other reduplication patterns, can partly be explained in terms of iconicity (see, e.g., Fischer 2011; Kentner 2017, 2022). Bybee et al. (1994: 167) assume that reduplicatives emerged as maximally iconic patterns in which the repetition of a verb signals the repetition of the action described by the verb. Fischer (2011: 64) sees the principle of "more of the same" as the key iconic source that can explain many apparently non-transparent instances of reduplication. This increase in quantity can occur in the vertical dimension (augmentation, intensification) or horizontal dimension (plurality, iteration, distribution; also see Kentner 2023: 105). This can entail, among other possibilities, a meaning of 'smallness': "Observing a multitude of similar items spread out on a horizontal plane makes each individual item appear relatively small and blurry." (Kentner 2023: 105) Moreover, Fischer (2011: 65) assumes that reduplication in child-directed speech "reflects the onomatopoeic imitation of the actual CV-CV syllabic babbling sounds made by children". Kentner (2023: 105) points out that this invites both negative and positive connotations. This general account can also be applied to *shm*-reduplication, which can be interpreted as a diminutivizing strategy ascribing a lack of importance to the reduplicant.

Turning from the pattern's function to its form, the replacement of the onset (or, sometimes, word-internal material) that is characteristic of echo reduplication

(Inkelas 2014: 170) has been described as "melodic overwriting" (McCarthy and Prince 1986: 68; Inkelas and Zoll 2015: 42). A prime example of melodic overwriting is *m*-reduplication, which has been characterized as "virtually pan-Asian" (Inkelas and Zoll 2015: 42) or "pan-Anatolian" (Donabedian and Sitaridou 2021: 409), e.g., Turkish *beyaz-meyaz* 'white and such things/allegedly white' (Donabedian and Sitaridou 2021: 414). From Turkish, *m*-reduplication has been borrowed into many other languages (including Kiezdeutsch, a variety of German used by young speakers in multilingual urban settings, see Wiese and Polat 2016). According to Donabedian and Sitaridou (2021: 414), *m*-reduplication is "mostly used to refer to a whole conceptual domain or to recall any word from the context with scepticism or irony". *m*-reduplication shares the latter function with *shm*-reduplication, while the former is fulfilled by other morphological patterns in English (e.g., *-ish* suffixation).

The status of reduplication in general, and of echo reduplication in particular, has been a matter of debate (Downing and Inkelas 2015: 520–526; Schwaiger 2018). Inkelas and Zoll (2005: 2–4) distinguish two major approaches to reduplication, one focusing on phonology, the other on semantic as well as syntactic properties. From the perspective of usage-based Construction Grammar (see, e.g., Ungerer and Hartmann 2023), we can conceive of *shm*-reduplication as a construction, i.e., a pairing of form and meaning at various levels of generality. By combining the concatenation of words with a quasi-paradigmatic stem alternation, *shm*-reduplication blends features traditionally attributed to syntax with features that are commonly seen as morphological. A constructionist approach, however, does not assume a strict division between morphology and syntax and can therefore be seen as particularly well-suited for investigating phenomena such as *shm*-reduplication, which can be "located at the border of word-formation and syntax" (Finkbeiner, Meibauer and Wiese 2016: 4).[2] Following Nagaya (2020: 267), who analyzes reduplication phenomena in terms of Construction Morphology (Booij 2010), a simplified formalization of the *shm*-reduplication construction could look as follows:

(2) $< [X_i$ *shm*-$X_i]_j \leftrightarrow [\text{DISMISSIVE } [\text{SEM}]_i]_j >$

This simple schema shows the form side of the construction on the left-hand side of the double arrow, and the meaning (or function) side on the right. A construc-

2 In fact, a reviewer points to another aspect in which *shm*-reduplication falls in between morphology and syntax: While identical constituent compounding (ICC) like *salad-salad* shows compound stress, *salad shmalad* shows phrasal stress. Note that this corresponds with the function of *shm*-reduplication: While ICCs refer to a specific entity and can therefore be readily interpreted as compound nouns, *shm*-reduplicatives express a certain evaluation of the reduplicant that is added to the unaltered original word.

tionist perspective would assume that speakers of English have a schema like (2) available as part of their linguistic knowledge, i.e., they know that echo reduplication with *shm-* modifies the semantics (SEM_i) of a given element (X_i) in such a way that it is (at least prototypically) interpreted as dismissive.

As this brief overview has shown, previous studies have highlighted several key characteristics of *shm*-reduplication, but they are based on small samples of the construction and do not provide in-depth quantitative analyses of its formal and semantic properties. In the following sections, we will explore what quantitative methods, when applied to a larger corpus sample, can tell us about the morphosyntactic and semantic profile of English *shm*-reduplication.

3 Corpus Study

3.1 Data and Methods

As *shm*-reduplication is expected to be largely a phenomenon of informal or conceptually oral language in the sense of Koch and Oesterreicher (1985), we opted for a corpus that documents computer-mediated communication, namely the webcorpus ENCOW16AX (Schäfer and Bildhauer 2012; Schäfer 2015).[3] ENCOW is part of the COW (= "Corpora from the Web") family of large web corpora in different languages. It contains almost 17 billion tokens from more than 9 billion documents.

We searched the corpus for words starting with <shm-> or <schm->. In the next step, the normalized Levenshtein distance (Levenshtein 1966) between the target word and the immediately preceding word was calculated, as well as between the target word and the second word preceding it. Levenshtein distance is a widely-used measure of edit distance between two strings: For instance, it takes one edit to get from *house* to *mouse*; taking the number of characters into account, this yields a normalized Levenshtein distance of 0.2. Similarly, it takes five edits to get from *house* to *castle*; divided by the length of the longer item, this yields a normalized Levenshtein distance of 0.83. As such, this measure allows for detecting graphemically similar words. In this way, we were able to detect *shm*-reduplicants based on

3 It should be noted, however, that ENCOW does not exclusively consist of computer-mediated communication – it contains a variety of text types, including some texts that are older than the Internet but happen to be available online. Still, the chances of finding instances of *shm*-reduplication are much higher in such a corpus than in other corpora. Compared to other widely-used web corpora, ENCOW has the advantage that it is freely accessible, which entails significant advantages for reproducibility and replicability (see Hartmann 2024).

individual lexemes, such as *Brexit, shmexit*, and based on compounds, such as *Chaos theory, shmaos theory*, or phrases, such as *ecological problems, shmecological problems*. A few cases involving more complex compounds or phrases may have been overlooked by taking this approach, and cases where the *shm-* is inserted word-internally, as in *obscene obshmene* (Nevins and Vaux 2003), are also neglected, but overall, this operationalization seems to offer a good balance between precision and recall. The items with a normalized Levenshtein distance of 0.4 and below were then checked manually. To analyze the morphosyntactic characteristics of the dataset, we manually annotated each instance for part of speech, syntactic position (e.g., separate sentence, sentence-initial but syntactically non-integrated, etc.), and morphological complexity (simplex or complex).

Overall, we ended up with 1,642 tokens of *shm*-reduplication, 1,557 of which had a simplex base (879 different types), while 85 instances had a compound or phrasal base (78 types).

In terms of semantic characteristics, we relied on several complementary approaches to assess the semantic spectrum of the pattern and its expressive potential. On the one hand, we used semantic vector-space modeling (see, e.g., Perek 2016) to map out the overall semantic spectrum of the construction and examine whether it clusters around particular areas of the possible semantic space. The basic assumption of this approach is that similar words tend to occur in similar contexts (Boleda 2020). More specifically, we employed a word2vec model originally created for Hartmann and Ungerer (2024), which was trained with Schmidt and Li's (2022) *wordVectors* package for R (R Core Team 2023) on the basis of the first of the 17 downloadable sets of ENCOW sentence shuffles (comprising around 600 million sentences). The training used a five-word window (i.e., five words before and after the target word were taken into account) and a skip-gram approach with negative sampling (i.e., the algorithm randomly samples other words in the text that are not neighbors; see Mikolov et al. 2013 for details). Note that the vectors resulting from this process are based on the entire dataset, i.e., not just on the occurrences of the target words in the *shm*-reduplication construction. For the visual representation of the results, we also follow Hartmann and Ungerer (2024) in using t-distributed Stochastic Neighborhood Embedding (t-SNE, van der Maaten and Hinton 2008), which allows for representing *n*-dimensional data in two-dimensional space.

On the other hand, we used collostructional analysis (Stefanowitsch and Gries 2003) to determine the lexemes that are strongly attracted to, or repelled by, the construction. This allowed us to gauge, from a more qualitative perspective, whether there are any semantic regularities among the items that most typically combine with *shm*-reduplication.

Finally, we drew on Warriner, Kuperman, and Brysbaert's (2013) affective meaning norms to check whether the construction tends to attract lexemes that

can be considered particularly expressive. We focused on two dimensions of affectedness: valence (defined as the pleasantness of the emotions invoked by a word) and arousal (capturing the intensity of emotion provoked by a word). Using crowdsourcing methods, Warriner, Kuperman, and Brysbaert had participants rate a total of 13,915 lexemes on 9-point scales (1 = lowest, 9 = highest) ranging from *unhappy* to *happy* (for valence) and from *calm* to *excited* (for arousal). Our rationale was that, if the valence and arousal values of *shm*-reduplicants differed from the rest of the norming set, this would suggest that the construction tends to be used to convey speakers' emotional involvement in the subject matter. Naturally, Warriner, Kuperman, and Brysbaert's norms have potential limitations: For instance, as one of our reviewers suggests, participants may have found it difficult to judge the affective properties of abstract words such as *title* or *easy*. Nevertheless, given that norming ratings (especially for valence) were quite consistent across participants (Warriner, Kuperman, and Brysbaert 2013: 1194) and that affective norms have informed a large body of research on emotions and word processing (Kuperman et al. 2014), we used them here as a (tentative) way of probing the expressive potential of our construction.

3.2 Results

We first present the results of our morphosyntactic analysis (Section 3.2.1), before examining the semantic characteristics of the construction (Section 3.2.2).

3.2.1 Morphosyntactic Analysis

Figure 1 illustrates how different parts of speech are distributed among the instances in our dataset. Overall, *shm*-reduplication is clearly biased towards nominal uses, which account for almost 70% of all instances. Adjectives form the second most common group (29.1%), while verbs and adverbs are very infrequent (together less than 2% of the data). An example of each category is provided in (3).

(3) a. ***Maps, schmaps..*** *They work just fine* (http://www.zdnet.com/sinofskys-departure-from-microsoft-politics-or-products-to-blame-7000007297)
b. ***Trustworthy Shmustworthy****. Who's the judge anyway.* (http://www.mattcutts.com/blog/snarky-or-not)
c. *In the end I built – oh,* ***built schmilt****.* (http://www.gold.ac.uk/glits-e/glits-earchive/2010-2011/contentspage/intertextualtransformationandtextualerasure)

d. ***Maybe, schmaybe****, but I hold UK standards in education in high regard* (http://www.chinasmack.com/2013/stories/dual-track-pension-system-civil-servants-6000-farmers-55-rmb.html)

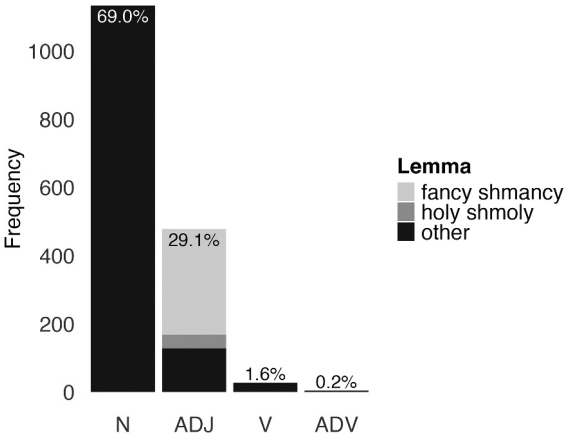

Figure 1: Distribution of parts of speech in the data.

Moreover, Figure 1 also shows that most adjectival uses of the construction are, in fact, restricted to two highly frequent types. First, *fancy shmancy* (310 attestations) accounts for 64.9% of adjectival uses, and for 18.9% of all instances of the construction, thus constituting by far the most frequently attested example of *shm*-reduplication. Second, *holy shmoly* (40 attestations) accounts for 8.4% of adjectival uses, while making up 2.4% of the entire dataset. This means that the rest of the construction is even more heavily biased towards nominal uses: While a range of other adjectives are attested in the construction (e.g., *indie schmindie*, *legal schmegal*), they together make up less than 10% of our data.

As a second step, we examined the syntactic positions in which *shm*-reduplication occurs. Our findings are summarized in Figure 2. The results indicate that most instances of *shm*-reduplication fill syntactically isolated positions.[4] Around half the time (48.9%), they occur in a separate sentence (or occasionally, split into two sentences), as illustrated in (4a). Moreover, when they form part of a sen-

4 As a reviewer points out, this tendency can be observed for other reduplicative patterns as well, e.g., reduplicative phrases such as German *Argumente hin, Argumente her* (lit. 'arguments to, arguments fro'; Finkbeiner 2017), which may be connected to the pragmatic function of these patterns: The base words have previously been mentioned or are otherwise salient in the discourse and are taken up as a quasi-quotation.

tence, they frequently occur in syntactically non-integrated positions, i.e., not as an argument or adjunct of the verb. Most of these non-integrated positions are sentence-initial (15.4%), as in (4b); while sentence-medial (1.4%) and sentence-final positions (3.4%) are also attested but less frequent, see (4c) and (4d). Syntactic non-integration is often signaled by a punctuation mark (e.g., comma, colon, dash), although this is not always the case given the inconsistency in the use of punctuation in informal (especially internet) language.

(4) a. ***Everest schmeverest****. Try climbing an active volcano!* (http://ngadventure.typepad.com/blog/the-adventure-top-10)
b. ***Beatles, schmeatles*** *– who else is missing from iTunes?* (http://blog.washingtonpost.com/clicktrack/2010/11/us_royalty_bridging_indie_rock.html)
c. *Who cares about all that self analysis as artists,* ***ego schmeego****, it's just painting!!!* (http://clicks.robertgenn.com/nom-de-brush.php)
d. *The issue is scarcely ever raised in public by US officials –* ***commitment, shmemitment****.* (http://badattitudes.com/MT/archives/2007/09/mr_olmert_tear.html)

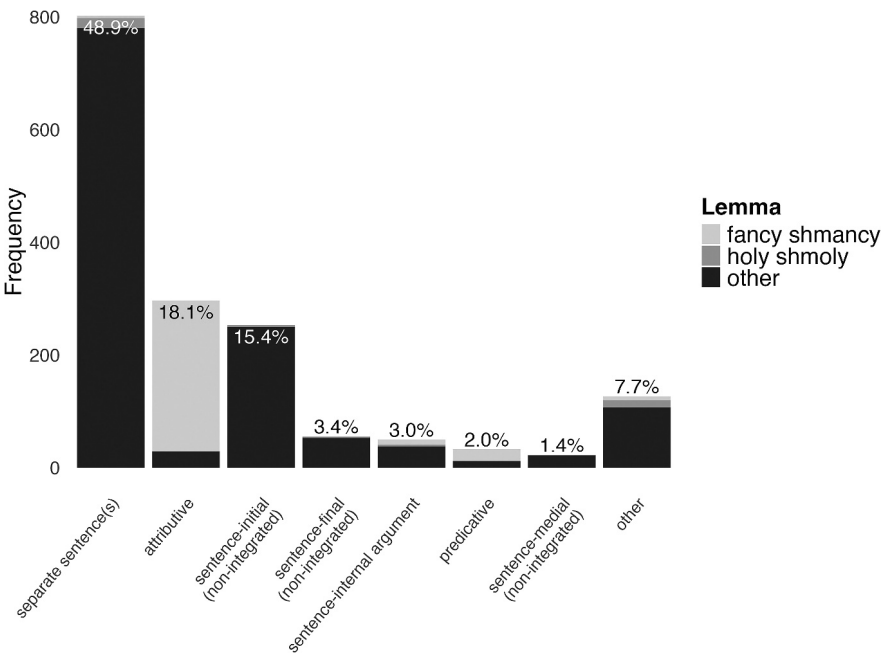

Figure 2: Distribution of syntactic positions.

In contrast, syntactically integrated positions within the sentence are less common. In particular, while attributive and predicative uses make up a considerable proportion (together 20.1%) of the dataset, most of these instances (around 90%) consist of *fancy shmancy*, as illustrated in (5a). Interestingly, this means that many other adjectives, when they occur in the construction, are not integrated into sentence-internal syntactic structures but rather appear in separate sentences, as in (5b). Beyond this, only relatively few instances of *shm*-reduplication (3%) occur as sentence-internal arguments, for example as a subject in (5c). Finally, there were a number of cases that we classified as "other," for example because they appeared sentence-internally but as quotations, or because their syntactic status was unclear.

(5) a. *i'd buy myself a* ***fancy schmancy*** *dress for Christmas* (http://www.shef fieldforum.co.uk/archive/index.php/t-166255.html)
b. ***Fair, shmair****. That's life in Big City, USA, Planet Real World.* (http://www.j-bradford-delong.net/movable_type/2003_archives/001699.html)
c. ***Details schmetails*** *will be forthcoming.* (http://derspatchel.livejournal.com/810339.html?page=3%26cut_expand=1)

Turning to the morphosyntactic complexity of the reduplicants, 64 tokens in our dataset (58 types) are based on compounds, usually N+N compounds as in *chaos theory, shmaos theory*, while 21 tokens (20 types) are based on phrases. The distinction between compounds and phrases is, of course, notoriously difficult (see, e.g., Schlechtweg 2018: 39–135), and there are a few doubtful cases in our data as well (e.g., *civil liberties*, annotated as a phrase here, although it could also be argued to be lexicalized as a holistic unit). Overall, however, compounds seem to outnumber phrases, which indicates that *shm*-reduplication tends to target single units. Phrasal reduplicants, as in *early morning, shmearly shmorning*, are less typical and hence arguably more salient or "extravagant" (see Section 4 for a more detailed discussion of this term). This is also supported by the observation that the proportion of phrases is higher among the 11 attestations in which both constituents are subject to melodic overwriting than among the remaining attestations in which *shm-* is only inserted into one (always the first) constituent: Five out of 11 instances with double melodic overwriting are phrases (45.5%), e.g., *optical density, schmoptical schmensity* or *enigmatic smile, schmnegmatic schmile*. Among the 74 complex instances in which only the first constituent is subject to melodic overwriting, only 16 are phrases (21.6%). Assuming that double melodic overwriting entails a higher degree of salience than single melodic overwriting, it seems plausible that speakers exploit the full extravagant potential of the pattern

by combining an unusual type of reduplicant (a phrase) with a highly salient doubling of melodic overwriting.

3.2.2 Semantic Analysis

For the semantic analyses, only *shm*-reduplications with a simplex base were taken into account; see below for a separate analysis of the (few) compound and phrasal attestations.

Figure 3 displays the results of the semantic vector-space analysis. The size of the words corresponds to their (logged) frequency in the construction. To read the many small-print (i.e., low-frequency) words, readers may refer to the online version of the diagram (see Section 6 "Data availability"). The ellipses in Figure 3 show clusters that were identified using partitioning around medoids (*pam*), a non-hierarchical clustering technique in which observations are clustered around the exemplar from which the average semantic distance to all other members of the cluster is minimal (see, e.g., Levshina 2015: 317).[5] Although some of the resulting clusters are quite heterogeneous, some interesting tendencies can be detected. Going clockwise through the graph, the cluster in the upper-left corner mostly contains lexemes from the domains of media and technology, e.g., *internet, cloud, graphics, usability*. Next to it, the top-center cluster revolves around financial and administrative terms, e.g., *capital, price, agency, logistics*. The cluster on the upper right comprises terms related to science and legal matters, e.g., *statistics, ethics, rule, privacy*. The cluster below that contains concepts from politics and society (*democracy, tradition*), abstract values (*honor, dignity*), and discourse (*contradiction, fact*). At the bottom center, we see a variety of terms related to technology (*helicopter, gear*), food and drinks (*pizza, booze*), and animals (*goose, turtle*). The bottom-left cluster is also rather heterogeneous, comprising concepts from the domains of religion (*holy, martyr*), people and celebrities (*princess, beckham*), sports (*basketball*), and time (*saturday, morning*). Finally, the center of the diagram contains terms from medicine and science (*cancer, photon, multicomponent*) as well as some more generic adjectives (*easy, moderate*).

Interestingly, there are some tendencies that apply across the different clusters. For example, the construction seems to show a preference for more or less technical terms from domains like science and technology, as well as from social and political domains. At the same time, the semantics of the lexemes that un-

5 The plot was created using the packages *ggplot2* (Wickham 2016) and *ggforce* (Pedersen 2024). For detecting the *pam* clusters, the package *cluster* (Maechler et al. 2023) was used.

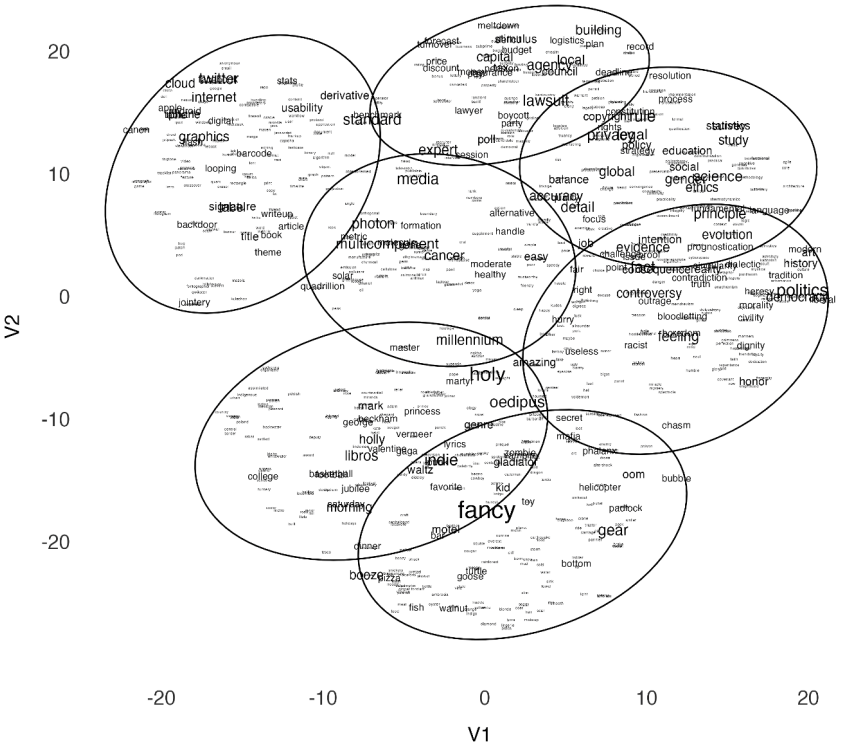

Figure 3: Results of a semantic vector-space analysis using a word2vec model trained on ENCOW data.

dergo *shm*-reduplication is extremely broad, which indicates that constraints on the construction's productivity may pertain more to formal and phonological features than to the semantics of the base words.

This also becomes clear when taking a closer look at the results of the collostructional analysis. There have been some debates regarding the adequate statistical methods to use when performing collexeme analyses, and the general consensus that is currently emerging is that different measures can be seen as complementary as they cover different relevant dimensions (see, e.g., Schneider 2020; Hartmann and Ungerer 2024). In Table 2, we draw on three measures, the first being the log likelihood G^2, a bidirectional measure that quantifies the *mutual* attraction between lexemes and constructions. G^2 is strongly correlated with other popular collostructional measures such as the *p*-value of the Fisher-Yates Exact Test or chi-squared residuals (Gries 2023), which is why only one of these three measures is reported here. These measures have in common that they are

sensitive to frequency. The odds ratio, by contrast, is independent of sample size and can therefore be used as a measure of "pure" association that is not influenced by sample size. Like G^2, it is a bidirectional measure. Delta P (ΔP), in turn, is unidirectional and assesses the degree to which a construction "attracts" a specific lexeme, or vice versa (Schneider 2020). For example, *fancy* has a very high construction-to-word ΔP, as *fancy shmancy* makes up a considerable proportion of all attestations of *shm*-reduplication in our data. By contrast, its word-to-construction ΔP is fairly low as *fancy* is attested much more often outside of than within the construction.

Table 2 shows the top 30 attracted collexemes, sorted by G^2. Perhaps most importantly, the results clearly show a skew in the distribution of the collexemes, especially when looking at the G^2 value, which takes the frequency of the lexemes into account. *fancy shmancy* is by far the most frequent exemplar, and consequently, *fancy* is the most strongly attracted lexical item, followed by *holy*, as in *holy shmoly*. This confirms the observation made in Section 3.2.1 that these two types are particularly prominent examples of *shm*-reduplication, which make up the lion's share of its adjectival instances. Apart from that, the semantic groups identified in the vector-space analysis above can also be seen in the results of the collostructional analysis, with terms from (pop) culture and media as well as technical terms and socio-political vocabulary ranking high.

Table 2: Top 30 attracted collexemes.

Collexeme	Corpus frequency	Frequency in construction	Odds ratio	ΔP word-to-construction	ΔP construction-to-word	G^2
fancy	194089	310	3.68	0.0016	0.2	4564.42
holy	202491	40	2.69	0.0002	0.026	414.13
oedipus	89	10	5.5	0.11	0.0066	231.33
indie	49388	11	2.75	0.00022	0.0072	116.31
multicomponent	509	5	4.42	0.0098	0.0033	90.74
politics	548713	12	1.74	0.00002	0.0077	71.6
twitter	81869	8	2.4	0.0001	0.0052	71.43
monkeon	35	3	5.41	0.086	0.002	67.67
ethics	937	4	4.06	0.0043	0.0026	65.9
lawsuit	126192	8	2.21	0.00006	0.0052	64.54
photon	32776	6	2.68	0.00018	0.0039	61.07
internet	10245	5	3.11	0.00049	0.0033	60.68
libros	134	3	4.81	0.022	0.002	59.42
rights	2	2	7.08	1	0.0013	58.77
statistics	375	3	4.35	0.008	0.002	53.2
millennium	64147	6	2.39	0.00009	0.0039	53.04

Table 2 (continued)

Collexeme	Corpus frequency	Frequency in construction	Odds ratio	ΔP word-to-construction	ΔP construction-to-word	G^2
privacy	267201	8	1.88	0.00003	0.0052	52.66
vermeer	8	2	5.97	0.25	0.0013	49.77
dilepton	57	2	5.03	0.035	0.0013	41.44
jointery	73	2	4.92	0.027	0.0013	40.43
rule	1820515	11	1.18	0.00001	0.0067	38.42
oom	228	2	4.42	0.0088	0.0013	35.84
beckham	296	2	4.31	0.0068	0.0013	34.79
accuracy	179452	5	1.87	0.00003	0.0032	32.19
graphics	202436	5	1.82	0.00002	0.0032	31
build	1	1	6.86	1	0.00066	29.38
hillphones	1	1	6.86	1	0.00066	29.38
mexican	1	1	6.86	1	0.00066	29.38
windows	1	1	6.86	1	0.00066	29.38
iphone	22409	3	2.57	0.00013	0.002	28.66

Given that the overall number of attestations is comparatively low and, more importantly, that very few instances of the pattern are attested more than a handful of times, the results of the collexeme analysis have to be taken with caution, regardless of which measure we use. But even some of the facts that potentially contribute to reducing the validity of the results reported in Table 2 can be quite informative about the pattern itself: For example, some of the collexemes have a high rank because they are attested very infrequently in the corpus as a whole, which entails that small differences in either corpus or construction frequency can make a large difference regarding the rank of the lexical item in question. For example, both corpus and construction frequencies are low in the case of some proper names like *Beckham, Monkeon* (apparently the name of a forum user) and *Vermeer*, which tend to occur in very specific contexts only. Others are highly technical, such as *dilepton* or *jointery*. The case of *oom* illustrates a further problem of collostructional analysis, namely that it is blind to polysemy (Dekalo and Hampe 2017). In ENCOW, *oom* typically occurs as a discourse marker (as in *uhm*) or in Dutch material (as a preposition meaning 'around'). In *oom-shmoom*, however, *oom* represents the Hebrew pronunciation of *U.N.*, as (6) shows.

(6) *United Nations is 'oom', and as David Ben Gurion said about the philanderings of the U.N., "**oom shmoom**!"* (http://www.michaeltotten.com/archives/2009/05/did-hezbollah-k.php, ENCOW)

These limitations, however, are informative about the pattern in at least three ways: First, the high type-token ratio indicates that the pattern is very productive and that it allows for creating many ad-hoc coinages tailored to specific situations. Second, the fact that some very infrequent lexemes occur in the construction also supports the impression that its uses are often tailored to highly specific contexts. Third, the fact that the data contain material from other languages such as Hebrew or Yiddish testifies to the inherently multilingual (and multicultural) character of the construction. *shm*-reduplication seems to be connected to the Yiddish language, and to Jewish culture, at least to some extent. The third-ranked collexeme, *Oedipus*, also bears witness to this as it refers to a classic Jewish joke (Paley 2019), the punchline of which is quoted in (7).[6]

(7) *Remember the wise words of Jewish mothers everywhere. "**Oedipus, shmoedipus** – who cares, so long as he loves his mother."*
(http://newhumanist.org.uk/595/hail-mary-, ENCOW)

The next step of our analysis was to assess the emotional valence of the words that occur in *shm*-reduplication, by drawing on Warriner, Kuperman, and Brysbaert's (2013) widely used affective meaning norms. As described in Section 3.1, these norms provide valence and arousal ratings for almost 14,000 English lexemes, measured on a scale from 1 (lowest) to 9 (highest). Figure 4 shows the distribution of valence and arousal ratings for all lemmas from our dataset that also occur in the norming data. The font size of the lexemes corresponds to their (logged) frequency in the *shm*-reduplication construction. For better readability of the high-frequency lexemes, lexemes attested at least three times are shown in black while the less frequent ones are displayed in gray. The graph is also available online (see Section 6 "Data availability"), allowing readers to zoom into its details.

If Kołłątaj's (2016: 243) assumption that "[t]he first component of a reduplicative compound in itself is normally endowed with a positive meaning" was correct, then we would expect the data to be skewed towards high valence ratings. Indeed, the median of the valence value is slightly higher for lexemes that occur in the *shm*-reduplication construction than for those lexemes in Warriner, Kuperman, and Brysbaert's (2013) norms that are not attested in our dataset. According to a two-sample t-test, the difference is highly significant (t=5.49, df=636, p<0.001). This suggests that the construction shows a certain affinity to lexical items with a positive semantic prosody, although there is also a considerable number of items on the left-hand (i.e., low-valency) side of the scale, some of which occur quite fre-

6 Thanks to Barbara Wehr for pointing this out to us.

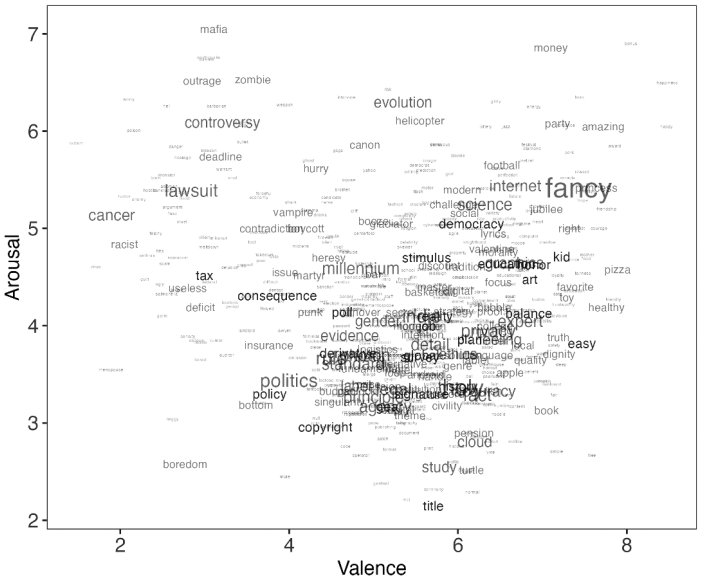

Figure 4: Distribution of the lexemes occurring in the *shm*-reduplication construction across the continuum of valence and arousal according to Warriner, Kuperman, and Brysbaert's (2013) affective words list.

quently in the construction. As for the arousal values, we did not find any significant difference between the lexemes occurring in the construction and the lexemes not attested in the construction (two-sample *t*-test, t=1.13, df=636, p=0.26). Overall, then, the construction seems to show a slight preference towards items with a positive semantic prosody, but at the same time, it combines with a broad variety of different items across the entire spectrum of valence and arousal values.

Finally, let us take a look at the instances of *shm*-reduplication with a compound or phrasal base that we have neglected in the semantic analysis so far. Only five compounds are attested more than once in our data (*black mail, climate change, earth day, tea party, peer review*). Still, they are quite representative of the semantic spectrum of *shm*-reduplication as discussed above, as they contain environmental and socio-political terms as well as one term from the academic domain. The 11 compounds or phrases in which both constituents are subject to melodic overwriting also show the full range of the semantic spectrum discussed above, from socio-political and legal terms (*individual mandate, civil liberties*) to pop culture (*kindergarten cop*) and technical terms from science and technology (*optical density*).

3.3 Discussion

Overall, our corpus results illustrate that *shm*-reduplication has a relatively "specialized" syntactic and pragmatic profile, while at the same time being flexibly used across a variety of semantic domains. On the one hand, the syntactic analysis suggests that the construction is prototypically applied to nouns, and that most of its instances are syntactically non-integrated, occurring either as separate sentences or at the sentence periphery. This also provides indirect evidence for the pragmatic function of *shm*-reduplication, which is typically used to express dismissive comments, often in the form of asides or afterthoughts. A related finding comes from our semantic valency analysis, which provides at least some tentative evidence that *shm*-reduplication tends to combine with lexemes that are slightly above average in terms of their emotional valency. This hints again at the fact that the construction is pragmatically biased towards addressing concepts that are emotionally relevant to the speaker and/or hearer, and which the speaker wants to express a potentially emotion-laden stance towards. Note, however, that the skew towards positive semantic prosody only concerns the base words that enter into the construction, not the valency of the entire *shm*-instances. Considering the constructional semantics outlined in Section 2, this means that speakers tend to use *shm*-reduplication to express a dismissive stance towards terms that may otherwise invoke positive associations.[7]

On the other hand, our semantic-vector space analysis indicates that the lexemes that undergo *shm*-reduplication belong to a diverse range of semantic fields, including for example food, music, science, technology, and society. Nevertheless, as was also reflected in the collostructional analysis, the construction seems especially productive within the domains of science and technology, the latter also including media- and internet-related terms. This may be partly an artifact of the data we used, since these topics are naturally quite well-represented in web corpora. Still, the high type frequency of the construction in these domains can also be seen as evidence that *shm*-reduplication is a particularly prevalent construction within online communities, where it is applied to concepts that are typically of concern to internet users.

Our corpus analysis also yielded several other interesting findings. One is that *shm*-reduplication can be applied to compound words and phrases, and that there

7 Importantly, as a reviewer correctly points out, the pragmatic function of the construction also hinges on its phonological properties. As such, phonological properties constrain the pattern's domain of application – for instance, monosyllabic words emerge as strongly repelled items in the collostructional analysis. One reason for this may be that the rhyming pattern works better with polysyllabic words.

are even some salient cases in which *shm-* is doubly attached to the constituents (e.g., *civil liberties, schmivil schmiberties*). We will discuss this together with other particularly creative examples of the construction in Section 4. Moreover, we identified two collexemes that are particularly strongly associated with *shm*-reduplication: the (highly frequent) *fancy shmancy* and the (somewhat less frequent) *holy shmoly*. Interestingly, both of these display characteristics that are atypical of the rest of the construction. Not only is *fancy shmancy* an adjective, while other adjectival uses of *shm*-reduplication are rare; but it also almost exclusively occurs attributively and predicatively, i.e., in prototypical sentence-internal positions, rather than in separate sentences or at the sentence periphery. Meanwhile, *holy shmoly* lacks the typical dismissive pragmatics of the construction, rather being used as an intensified exclamation that signals the speaker's surprise. It is actually unclear if *holy shmoly* is an instance of the *shm*-reduplication construction as we have defined it here, also given that it is merely one among several playful phonological variations, such as *holy moly*, *holy camoly*, and *holy guacamole*. *fancy shmancy*, on the other hand, typically expresses a depreciative element, characterizing something as "extremely fancy [. . .] in a pretentious or ostentatious way" (OED 2023, "fancy-schmancy"). While it therefore qualifies as an instance of *shm*-reduplication, it may be exactly in virtue of its high type frequency that *fancy shmancy* has developed a distinct syntactic profile from the rest of the construction. As a result, *fancy shmancy* can potentially be regarded as a subconstruction in its own right, which might serve as a salient analogical model for other adjectival *shm*-reduplicants – a point that we will return to in Section 4.

4 The Interplay of Creativity and Routine

The results of our corpus study illustrate the close interplay between aspects of linguistic *creativity* and *routine* in sanctioning speakers' use of *shm*-reduplication. We will explore these implications now in the broader context of concepts and ideas that have been developed in the recently growing constructionist literature on linguistic creativity (e.g., Bergs 2018, 2019; Hoffmann 2018, 2020; Trousdale 2020; Uhrig 2020).

A useful starting point is Sampson's (2016) distinction between *F-creativity* (for "fixed" creativity) and *E-creativity* (for "enlarging" or "extending" creativity), which has also been put to fruitful use in other constructionist work (e.g., Bergs 2018; Hoffmann 2018). According to Sampson (2016: 19), F-creative activities "characteristically produce examples drawn from a fixed and known (even if infinitely large) range," while E-creative activities "characteristically produce exam-

ples that enlarge our understanding of the range of possible products of the activity". Applied to language, F-creativity is usually understood in terms of the productivity of a conventionalized rule or pattern, thus resembling the traditional Chomskyan use of the term "creativity" (Chomsky 1965; see Bergs 2018: 277–279). E-creativity, on the other hand, involves some notable deviation from the established linguistic norms, thus extending existing patterns in previously non-licensed ways.

The phenomenon of *shm*-reduplication displays several features that can be regarded as E-creative, i.e., deviating from typical linguistic norms. First, word-initial /ʃm/ is an atypical onset in English, occurring only in relatively few, usually loaned words (see Section 2). Second, reduplication as a morphological process is very rare in English, and in Indo-European languages in general (Schwaiger 2015: 478).[8] Third, as our corpus results show, most instances of *shm*-reduplication are not syntactically integrated with the sentence, but they rather form separate (syntactically incomplete) sentences or fill peripheral positions that are structurally marked off from the sentence core. These non-prototypical syntactic positions lend themselves to the specific discourse function of *shm*-reduplication, which is typically used to express dismissive comments, often in the form of asides or afterthoughts.

Together, these features differentiate *shm*-reduplication from the rest of the linguistic system, thus arguably contributing to its E-creative nature. As a result, speakers can use the construction to demonstrate their linguistic skills and thereby attract attention. In this sense, *shm*-reduplication can be seen as an "extravagant" construction, thus illustrating another concept that has gained prominence in (diachronic) constructionist research (e.g., Petré 2016; Neels, Hartmann and Ungerer 2023; Hartmann and Ungerer 2024). Popularized by Haspelmath (1999), extravagance can be defined as speakers' desire to be noticed and stand out from their peers (see also Ungerer and Hartmann 2020). This also fits with the finding from our corpus analysis that *shm*-reduplication shows a slight tendency to involve lexemes with relatively high emotional valence, thus increasing the chances that the expressions are relevant for hearers and catch their attention. As illustrated by *shm*-reduplication, extravagance and E-creativity are often cyclically related: Speakers' creative breach of linguistic norms constitutes a mechanism (though perhaps not the only one) through which expressions achieve extravagant effects; while the social motivation of attracting others' attention provides an explanation for why speakers choose to be creative in the first place.

8 As a reviewer points out, however, this mainly pertains to standard registers, while reduplication may be more frequent in non-standard registers.

While *shm*-reduplication appears to be E-creative when compared with other, more canonical constructions, the way in which the pattern is extended to new instances also provides evidence of F-creativity. In particular, the high type frequency together with the results of our semantic analysis indicate that *shm*-reduplication constitutes a productive pattern that is applied across a variety of semantic domains. This suggests that speakers have formed a well-entrenched *shm*-schema that can be used to license new instances in flexible but ultimately predictable (and thus F-creative) ways. The relatively homogeneous nature of these instances is highlighted by the fact that most of them adhere to the specific syntactic and pragmatic characteristics of the construction (i.e., occurring in non-integrated positions and expressing dismissive comments). Interestingly, therefore, the narrow syntactic and pragmatic profile that underlies the pattern's E-creativity relative to other constructions (see above) also seems to scaffold its F-creativity by increasing the internal consistency among its instances and thus arguably facilitating the pattern's productive extensions. F-creativity thus illustrates the role that routines – perhaps more appropriately termed *creative routines* in this context – play in *shm*-reduplication, where the repeated use of an originally E-creative pattern leads to the formation of an increasingly well-entrenched schema in speakers' minds. The fact that an increase in the routine-like nature of *shm*-reduplication may, over time, lead to a decrease in its perceived E-creativity is in line with previous observations that language change often involves a cyclical interaction between speakers' desire for extravagance and conformity (Haspelmath 1999; Neels, Hartmann and Ungerer 2023).

Of course, there are still some pockets of E-creativity left *within* the *shm*-reduplication construction: Certain exemplars definitely appear more creative than others. For example, our analysis indicates that *shm-* is sometimes applied to both constituents of compound words (e.g., *impulse buying, schmimpulse schmuying*), which is likely to attract attention. There are also other examples in which *shm*-reduplication is used in particularly playful ways: In (8a), for example, the speaker adds *schmillion* to an enumeration of number words, thus inducing an ad-hoc interpretation of the word as an extremely large numeral while at the same time expressing sarcasm towards the content of the previous sentence. In (8b), the speaker first transforms *dentists* into a made-up form *dontists* before applying *shm*-reduplication, thus increasing the phonological variation and strengthening the dismissive connotation.

(8) a. *. . . that amount of money out of thin air every week this year.* ***Billion, trillion, quadrillion, schmillion****.* (http://www.housepricecrash.co.uk/newsblog/2009/11/blog-money-illusion-26316.php)
b. ***Dentists, dontists, schmontists****, they are all bloody expensive these days* (http://www.sheffieldforum.co.uk/archive/index.php/t-179797.html)

On the other hand, there are also further signs that elements of routine are at work in *shm*-reduplication. First, while the pattern is semantically relatively unconstrained, our vector-space analysis nevertheless points to some domains in which the construction is particularly frequent, for example internet- and media-related concepts. Together with the informal impression we gained during our analysis that many of our examples stem from internet forums, i.e., semi-private subcommunities, this suggests that computer-mediated communication serves not only as a social context in which *shm*-reduplication is used particularly prolifically but also as a topic to which the construction is applied. The creation of such socio-functional "niches" may contribute to the formation of conversational routines that boost the construction's productivity. Second, our analysis provided evidence of two particularly frequent types, *fancy shmancy* and *holy shmoly*. These may arguably have the status of mini-constructions in their own right, especially given that they diverge from some of the construction's prototypical characteristics (with *fancy shmancy* displaying the syntactic behavior of a typical adjective, and *holy shmoly* lacking the dismissive connotation). Nevertheless, they form salient routines that may potentially scaffold the creation of new *shm*-instances via analogical extension (see, e.g., De Smet 2012: 8). For example, *fancy shmancy* may serve as a model for other attributive and predicative uses of *shm*-reduplicated adjectives.

5 Conclusion

In this paper, we have presented a corpus-based analysis of *shm*-reduplication based on a comparatively large sample drawn from the ENCOW corpus. *shm*-reduplication can be seen as a paradigm example of an "extravagant" pattern at the interface of word-formation and syntax. We have focused on the morphosyntactic usage of the construction as well as on its semantic characteristics. Regarding the former, we have shown that *shm*-reduplication is usually syntactically non-integrated, most prototypically occurring in separate sentences. The most frequent instance *fancy shmancy* makes up the bulk of attributive and predicate uses, which suggests that it can be considered a construction in its own right (similarly to the second most frequent instance, *holy shmoly*). Regarding the semantics of *shm*-reduplication, we have shown that the construction is not constrained to any particular semantic domain, even though it combines particularly frequently with terms from the socio-political realm, science, and technology. In terms of its semantic prosody, the pattern combines with lexemes across the emotional valence scale, even though it shows a slight preference for words with positive emotional valence.

Based on these characteristics, the pattern behaves quite similar to other extravagant constructions such as "snowclones" like *X is the new Y* or *the mother of all X* (Hartmann and Ungerer 2024), which also display preferences for some semantic domains but are still open for a semantically virtually unconstrained inventory of slot fillers. In addition, the productivity pattern of *shm*-reduplication shows some similarities to these snowclones as well as to extravagant word-formation patterns like pseudo-participles (e.g., *bebrillt* 'be-glass-ed', see Kempf and Hartmann 2022): Apart from the outliers *holy shmoly* and especially *fancy shmancy* with many attestations, most instantiations of the pattern are attested only a few times, or even just once. This indicates that the instances of the construction are usually ad-hoc coinages that are strongly tied to the specific contexts in which they are used. The latter usage profile may be a common characteristic of linguistic patterns that can be characterized as extravagant, i.e., that are somehow unusual and salient and thus, at least partly, serve the purpose of gaining the hearers' attention.

Whether or not this is the case would be an interesting question for a potential follow-up study that systematically compares a set of extravagant constructions across different domains (e.g., morphology, syntax, phraseology) and ideally across different languages. A contrastive perspective could be particularly fruitful for *shm*-reduplication, as the pattern is attested in other languages (e.g., German) as well. A comparison with the above-mentioned *m*-reduplication could also prove insightful, especially given that the semantic domains of both constructions overlap but are not fully congruent. A further potentially promising way of extending the analysis of *shm*-reduplication could be to add a multimodal perspective: Given its dismissive semantics, we could expect that in spoken language, it is often combined with dismissive gestures such as members of the family of AWAY gestures (Bressem and Müller 2014).[9] In addition, given our assumption that *shm*-reduplication is a conceptually more oral phenomenon, taking multimodal data into account can potentially provide more insights into the use of *shm*-reduplication in communicative interaction.

These examples show that there is still a lot of uncharted territory to explore when it comes to the interplay of creativity and routine in the use of *shm*-reduplication, and of expressive patterns in word formation and syntax in general. With this paper, we hope to have shed some new light on the use of *shm*-reduplication in authentic (written) data, thus laying the groundwork for follow-up studies of this pattern and related phenomena.

9 A pilot study based on data from the TV News Archive (https://archive.org/details/tv, accessed 27 February 2024) shows that this is the case in 13 out of 23 attestations.

6 Data Availability

The data, R scripts, and additional plots referred to in the text are available via the Open Science Framework (OSF): https://osf.io/f2cu3/

References

Bergen, Benjamin K. 2004. The psychological reality of phonaesthemes. *Language* 80(2). 290–311. https://doi.org/10.1353/lan.2004.0056.

Bergs, Alexander. 2018. Learn the rules like a pro, so you can break them like an artist (Picasso): Linguistic aberrancy from a constructional perspective. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 277–293. https://doi.org/10.1515/zaa-2018-0025.

Bergs, Alexander. 2019. What, if anything, is linguistic creativity? *Gestalt Theory* 41(2). 173–183. https://doi.org/10.2478/gth-2019-0017.

Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6(1). 213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303.

Booij, Geert E. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Bressem, Jana & Cornelia Müller. 2014. The family of Away gestures: Negation, refusal, and negative assessment. In Cornelia Müller, Alan J. Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Sedinha Tessendorf (eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, 1592–1604. Berlin & Boston: De Gruyter.

Bybee, Joan L., Revere Perkins & William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Cienki, Alan. 2017. Utterance Construction Grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard* 3(s1). https://doi.org/10.1515/lingvan-2016-0048.

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

De Smet, Hendrik. 2012. *Spreading Patterns: Diffusional Change in the English System of Complementation*. Oxford: Oxford University Press.

Dekalo, Volodymyr & Beate Hampe. 2017. Networks of meanings: Complementing collostructional analysis by cluster and network analyses. *Yearbook of the German Cognitive Linguistics Association* 5(1). 143–176. https://doi.org/10.1515/gcla-2017-0011.

Donabedian, Anaïd & Ioanna Sitaridou. 2021. Anatolia. In Evangelia Adamou & Yaron Matras (eds.), *The Routledge Handbook of Language Contact*, 404–433. London & New York: Routledge.

Downing, Laura J. & Sharon Inkelas. 2015. What is reduplication? Typology and analysis part 1/2: The typology of reduplication. *Language and Linguistics Compass* 9(12). 516–528. https://doi.org/10.1111/lnc3.12152.

Finkbeiner, Rita. 2012. Naja, normal und normal. Zur Syntax, Semantik und Pragmatik der x-und-x-Konstruktion im Deutschen. *Zeitschrift für Sprachwissenschaft* 31(1). 1–42. https://doi.org/10.1515/zfs-2012-0001.

Finkbeiner, Rita. 2017. “Argumente hin, Argumente her.” Regularity and Idiomaticity in German N Hin, N Her. *Journal of Germanic Linguistics* 29(3), 205–258. https://doi.org/10.1017/S1470542716000234.

Finkbeiner, Rita, Jörg Meibauer & Heike Wiese. 2016. What is pejoration, and how can it be expressed in language? In Rita Finkbeiner, Jörg Meibauer & Heike Wiese (eds.), *Pejoration*, 1–18. Amsterdam & Philadelphia: John Benjamins.

Firth, John. 1930. *Speech*. Oxford: Oxford University Press.

Fischer, Olga. 2011. Cognitive iconic grounding of reduplication in language. In Pascal Michelucci, Olga Fischer & Christina Ljungberg (eds.), *Semblance and Signification*, 55–81. Amsterdam/ Philadelphia: John Benjamins.

Frankowsky, Maximilian. 2022. Extravagant expressions denoting quite normal entities: Identical constituent compounds in German. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant Morphology. Studies in Rule-Bending, Pattern-Extending and Theory-Challenging Morphology*, 156–179. Amsterdam & Philadelphia: John Benjamins.

Grandi, Nicola. 2017. Evaluatives in morphology. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.250 (last modified 29 March 2017).

Grandi, Nicola & Livia Körtvélyessy. 2015. Introduction: Why evaluative morphology? In Nicola Grandi & Livia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: Edinburgh University Press.

Gries, Stefan Th. 2023. Overhauling collostructional analysis: Towards more descriptive simplicity and more explanatory adequacy. *Cognitive Semantics* 9(3). 351–386. https://doi.org/10.1163/ 23526416-bja10056.

Grohmann, Kleanthes K. & Andrew Ira Nevins. 2004. Echo reduplication: When too-local movement requires PF-distinctness. *University of Maryland Working Papers in Linguistics* 13. 84–108.

Hartmann, Stefan. 2024. Open Corpus Linguistics – or How to overcome common problems in dealing with corpus data by adopting open research practices. In Mark Kaunisto & Marco Schilk (eds.), *Challenges in Corpus Linguistics: Rethinking Corpus Compilation and Analysis*, 89–105. Amsterdam & Philadelphia: John Benjamins.

Hartmann, Stefan & Tobias Ungerer. 2024. Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns. *Journal of Linguistics* 60(3). 599–634. https://doi.org/10.1017/ S0022226723000117.

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

Hoffmann, Thomas. 2018. Creativity and Construction Grammar: Cognitive and psychological issues. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 259–276. https://doi.org/10.1515/zaa-2018-0024.

Hoffmann, Thomas. 2020. Construction grammar and creativity: Evolution, psychology, and cognitive science. *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2018.

Hohenhaus, Peter. 2004. Identical constituent compounding – a corpus-based study. *Folia Linguistica* 38(3–4). https://doi.org/10.1515/flin.2004.38.3-4.297.

Inkelas, Sharon. 2014. Non-concatenative derivation: Reduplication. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 169–189. Oxford: Oxford University Press.

Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication: Doubling in Morphology*. Cambridge: Cambridge University Press.

Jurafsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72(3). 533–578. https://doi.org/10.2307/416278.

Kempf, Luise & Stefan Hartmann. 2022. What's extravagant about *be-sandal-ed feet*? Morphology, semantics and pragmatics of German pseudo-participles. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant Morphology: Studies in Rule-Bending, Pattern-Extending and Theory-Challenging Morphology*, 19–50. Amsterdam & Philadelphia: John Benjamins.

Kentner, Gerrit. 2017. On the emergence of reduplication in German morphophonology. *Zeitschrift für Sprachwissenschaft* 36(2). 233–277. https://doi.org/10.1515/zfs-2017-0010.

Kentner, Gerrit. 2022. do not repeat: Repetition and reduplication in German revisited. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant Morphology: Studies in Rule-Bending, Pattern-Extending and Theory-Challenging Morphology*, 182–205. Amsterdam: John Benjamins.

Kentner, Gerrit. 2023. Reduplication as expressive morphology in German. In J. P. Williams (ed.), *Expressivity in European Languages*, 1st edn., 103–120. Cambridge: Cambridge University Press.

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36. 15–43. https://doi.org/10.1515/9783110244922.15.

Kołłątaj, Andrzej. 2016. Reduplication in English – typology, correlation with slang and metaphorisation. *Philolog. Studia Neofilologiczne* 6. 237–248. https://doi.org/10.34858/polilog.6.2016.021.

Kuperman, Victor, Zachary Estes, Marc Brysbaert & Amy Beth Warriner. 2014. Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General* 143(3). 1065–1081. https://doi.org/10.1037/a0035669.

Lensch, Anke. 2018. Fixer-uppers. Reduplication in the derivation of phrasal verbs. In Rita Finkbeiner & Ulrike Freywald (eds.), *Exact Repetition in Grammar and Discourse*, 158–181. Berlin & Boston: De Gruyter.

Levenshtein, Vladmir I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10(8). 707–710.

Levshina, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam/Philadelphia: John Benjamins.

Maaten, Laurens van der & Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9(86). 2579–2605.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert & Kurt Hornik. 2023. cluster: Cluster analysis basics and extensions. Manual. https://CRAN.R-project.org/package=cluster.

Mattiello, Elisa. 2013. *Extra-Grammatical Morphology in English: Abbreviations, Blends, Reduplicatives, and Related Phenomena*. Berlin & Boston: De Gruyter.

McCarthy, John & Alan Prince. 1986. Prosodic morphology. https://hdl.handle.net/20.500.14394/32398.

Meibauer, Jörg. 2013. Expressive compounds in German. *Word Structure* 6(1). 21–42. https://doi.org/10.3366/word.2013.0034.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado & Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In Christopher J. Burges, Leon Bottou, Max Welling, Zoubin Ghahramani & Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 3111–3119. Red Hook, NY: Curran Associates.

Minkova, Donka. 2002. Ablaut reduplication in English: The criss-crossing of prosody and verbal art. *English Language & Linguistics* 6(1). 133–169. https://doi.org/10.1017/S1360674302001077.

Nagaya, Naonori. 2020. Reduplication and repetition from a constructionist perspective. *Belgian Journal of Linguistics* 34. 259–272. https://doi.org/10.1075/bjl.00051.nag.

Neels, Jakob, Stefan Hartmann & Tobias Ungerer. 2023. A quantum of salience: Reconsidering the role of extravagance in grammaticalization. In Hendrik De Smet, Peter Petré & Benedikt Szmrecsanyi (eds.), *Context, Intent and Variation in Grammaticalization*, 47–78. Berlin & Boston: De Gruyter.

Nevins, Andrew & Bert Vaux. 2003. Metalinguistic, shmetalinguistic: The phonology of shm-reduplication. *Chicago Linguistic Society* 39(1). 702–721.

Norde, Muriel & Sarah Sippach. 2019. *Nerdalicious scientainment*: A network analysis of English libfixes. *Word Structure* 12(3). 353–384. https://doi.org/10.3366/word.2019.0153.

OED. 2023. *Oxford English Dictionary*. https://oed.com

Paley, Andrew. 2019. "Seeing the world as it is: The importance of honest sight." In *Texas Jewish Post*. https://tjpnews.com/seeing-the-world-as-it-is-the-importance-of-honest-sight (last modified 22 July 2019).

Pedersen, Thomas Lin. 2024. ggforce: Accelerating "ggplot2". Manual. https://CRAN.R-project.org/package=ggforce.

Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1). 149–188. https://doi.org/10.1515/ling-2015-0043.

Petré, Peter. 2016. Unidirectionality as a cycle of convention and innovation: Micro-changes in the grammaticalization of [be *going to* INF]. *Belgian Journal of Linguistics* 30. 115–146. https://doi.org/10.1075/bjl.30.06pet

R Core Team. 2023. R: A language and environment for statistical computing. Manual. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

Regier, Terry. 1994. A preliminary study of the semantics of reduplication. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=183743bf7da4da277d9ddf2f4aa1d34a58e110f8 (accessed 27 February 2024).

Rubino, Carl. 2013. Reduplication. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/27.

Saba Kirchner, Jesse. 2010. *Minimal reduplication*. Santa Cruz, CA: University of California, Santa Cruz dissertation. https://doi.org/doi:10.7282/T3VQ31K8.

Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, Experiment and Argument in Linguistics and Philosophy of Language*, 15–26. Bern: Peter Lang.

Schäfer, Roland. 2015. Processing and querying large corpora with the COW14 architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lüngen & Andreas Witt (eds.), *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3)*, 28–34.

Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Nicoletta Calzolari, Khalid Choukri, Terry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 486–493. Istanbul: European Language Resources Association.

Schlechtweg, Marcel. 2018. *Memorization and the Compound-Phrase Distinction: An Investigation of Complex Constructions in German, French and English*. Berlin & Boston: De Gruyter.

Schneider, Ulrike. 2020. ΔP as a measure of collocation strength: Considerations based on analyses of hesitation placement in spontaneous speech. *Corpus Linguistics and Linguistic Theory* 26(2). 249–274. https://doi.org/10.1515/cllt-2017-0036.

Schmidt, Ben & Jian Li. 2022. wordVectors: Tools for creating and analyzing vector-space models of texts. Manual. http://github.com/bmschmidt/wordVectors.

Schwaiger, Thomas. 2015. Reduplication. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word Formation: An International Handbook of the Languages of Europe*, vol. 1, 467–484. Berlin & New York: De Gruyter.

Schwaiger, Thomas. 2018. The derivational nature of reduplication and its relation to boundary phenomena. In Rita Finkbeiner & Ulrike Freywald (eds.), *Exact Repetition in Grammar and Discourse*, 67–88. Berlin & Boston: De Gruyter.

Southern, Mark R. V. 2005. *Contagious Couplings: Transmission of Expressives in Yiddish Echo Phrases*. Westport: Praeger.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243. https://doi.org/10.1075/ijcl.8.2.03ste.

Stroebel, Liane. 2017. Can Macromania be explained linguistically? Beneath the morphological boundary: A sketch of subconscious manipulation strategies in Emmanuel Macron's political discourses. *Yearbook of the German Cognitive Linguistics Association* 5(1). 57–76. https://doi.org/10.1515/gcla-2017-0005.

Trousdale, Graeme. 2020. Creativity, reuse, and regularity in music and language. *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2021.

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

Ungerer, Tobias & Stefan Hartmann. 2020. Delineating extravagance: Assessing speakers' perceptions of imaginative constructional patterns. *Belgian Journal of Linguistics* 34. 345–356. https://doi.org/10.1075/bjl.00058.ung.

Ungerer, Tobias & Stefan Hartmann. 2023. *Constructionist Approaches: Past, Present, Future*. Cambridge: Cambridge University Press.

Warriner, Amy Beth, Victor Kuperman & Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45(4). 1191–1207. https://doi.org/10.3758/s13428-012-0314-x.

Weinreich, Max. 1980. *History of the Yiddish Language*. Chicago: Chicago University Press.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*, 2nd edn. Berlin: Springer.

Wiese, Heike & NilginTanış Polat. 2016. Pejoration in contact: m-reduplication and other examples from urban German. In Rita Finkbeiner, Jörg Meibauer & Heike Wiese (eds.), *Pejoration*, 243–267. Amsterdam & Philadelphia: John Benjamins.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1987. Plain morphology and expressive morphology. *Annual Meeting of the Berkeley Linguistics Society* 13. 330–340. https://doi.org/10.3765/bls.v13i0.1817.