Complex words as constructions:
analysability, semantic transparency, and morphological productivity

Dissertation

(kumulativ)

Zur Erlangung des akademischen Grades doctor philosophiae (Dr. phil.)

vorgelegt dem Rat der Philosophischen Fakultät der Friedrich-Schiller-Universität Jena

von M.A. **Sergei Monakhov**

geboren am 05.08.1982 in Primorskij Krai (Russland)

Dissertation, Friedrich-Schiller-Universität Jena, 2024

Gutachter/Gutachterinnen:

1. Prof. Dr. Holger Diessel (Friedrich-Schiller-Universität Jena)

2. Prof. Dr. Volker Gast (Friedrich-Schiller-Universität Jena)

3. Dr. Martin Schäfer (Universität Leipzig)

Tag der Verteidigung: 07.05.2024

Abstract

Multi-morphemic or morphologically complex words are most simply defined as lexical items composed of more than one morpheme. However, this simplicity is deceptive because, in order for this definition to work, one must endorse the notion of morphemes as independent meaningful units. This is problematic in many respects, one being that in terms of both motivation and production, complex words under this account can only be represented in either decomposed or holistic fashions. The construction morphology framework, which treats complex words as constructions on the word level, allows for more fine-grained distinctions between the possible routes of word recognition due to its ability to identify fixed elements and slots (variables) in lexical structure.

Historically, this word-based morphology has been slow to develop, which has had important ramifications in both the empirical and theoretical domains. On the empirical side, the repertoire of languages and expressions that have appeared on the radar of constructional grammatical work remains meagre. From a theoretical point of view, several influential ideas inspired by the dual-route theory and pertaining to the problems of morphological analysability and productivity persist to the present day and, strangely, coexist with modern connectionist approaches.

The above considerations shape the motivation behind this thesis and its main research goals. From an empirical perspective, I aim to broaden the repertoire of languages for which multi-morphemic words have been studied within the framework of construction morphology. To this end, in the first part of the thesis, I provide a thorough analysis of Russian prefixed verbs as instantiations of two different types of prefix-base constructions: one with an open slot for the prefix and a fixed base verb and another with a fixed prefix and an open slot for the base verb.

From a theoretical perspective, I suggest a substantial rethinking of some still-popular approaches to the problems of the parsability and productivity of linguistic expressions. Specifically, I strive to leave behind what has been inherited in these domains from the dual-route theory of word recognition and to find a firm construction-morphology footing for the analysis of these phenomena. This part of the work is substantiated by corpus and experimental data from both Russian and English.

In the second part of the thesis, I address the problem of the morphological productivity of prefixes in Russian and English. I propose a new method of evaluating productivity that, unlike the popular hapax-based measure, is not immediately dependent on token frequency. The probabilistic

estimation of the linguistic productivity reveals that token frequency as such, contrary to common beliefs, cannot be considered a stumbling block for derivational patterns.

The third part of the thesis focuses on the morphological analysability and semantic transparency of multi-morphemic words. I propose distinguishing between two meaning processing models for complex words: one based on the principle of compositionality and another on the principle of parsability. I show that the words characterised by a greater discrepancy between transitional probabilities from affix to base and from base to affix are more likely to be treated as parsable, while those with more comparable (low) transitional probabilities are more likely to to be processed in a compositional manner.

Deutsche Zusammenfassung

Multimorphemische oder morphologisch komplexe Wörter werden am einfachsten als lexikalische Elemente definiert, die aus mehr als einem Morphem bestehen. Diese Einfachheit ist jedoch trügerisch, denn damit diese Definition funktioniert, muss man die Auffassung vertreten, dass Morpheme unabhängige Bedeutungseinheiten sind. Dies ist in vielerlei Hinsicht problematisch, unter anderem, weil komplexe Wörter nach dieser Auffassung sowohl in Bezug auf die Motivation als auch auf die Produktion nur entweder in zerlegter oder in holistischer Form dargestellt werden können. Der Rahmen der Konstruktionsmorphologie, der komplexe Wörter als Konstruktionen auf Wortebene behandelt, ermöglicht eine feinere Unterscheidung zwischen den möglichen Wegen der Worterkennung, da er in der Lage ist, feste Elemente und Slots (Variablen) in der lexikalischen Struktur zu identifizieren.

Historisch gesehen hat sich diese wortbasierte Morphologie nur langsam entwickelt, was sowohl im empirischen als auch im theoretischen Bereich erhebliche Auswirkungen hatte. Auf der empirischen Seite ist das Repertoire an Sprachen und Ausdrücken, die auf dem Radar der konstruktionsgrammatischen Arbeit erschienen sind, nach wie vor dürftig. Aus theoretischer Sicht bestehen einige einflussreiche Ideen, die von der Dual-Route-Theorie inspiriert wurden und sich auf die Probleme der morphologischen Analysierbarkeit und Produktivität beziehen, bis heute fort und koexistieren seltsamerweise mit modernen konnektionistischen Ansätzen.

Die oben genannten Überlegungen bilden die Motivation für diese Arbeit und ihre wichtigsten Forschungsziele. Aus einer empirischen Perspektive möchte ich das Repertoire der Sprachen erweitern, für die multimorphemische Wörter im Rahmen der Konstruktionsmorphologie untersucht wurden. Zu diesem Zweck analysiere ich im ersten Teil der Arbeit eingehend die russischen präfixierten Verben als Instanzen von zwei verschiedenen Arten von Präfix-Basis-Konstruktionen: eine mit einem offenen Slot für das Präfix und einem festen Basisverb und eine andere mit einem festen Präfix und einem offenen Slot für das Basisverb.

Aus theoretischer Sicht schlage ich vor, einige immer noch populäre Ansätze zur Problematik der Parsifizierbarkeit und Produktivität von sprachlichen Ausdrücken grundlegend zu überdenken. Insbesondere bemühe ich mich darum, das, was in diesen Bereichen von der Zwei-Wege-Theorie der Worterkennung übernommen wurde, hinter sich zu lassen und eine solide konstruktionsmorphologische Grundlage für die Analyse dieser Phänomene zu finden. Dieser Teil der Arbeit wird durch Korpus- und experimentelle Daten sowohl aus dem Russischen als auch aus dem Englischen untermauert.

Im zweiten Teil der Arbeit befasse ich mich mit dem Problem der morphologischen Produktivität von Präfixen im Russischen und Englischen. Ich schlage eine neue Methode zur Bewertung der Produktivität vor, die im Gegensatz zu dem beliebten hapax-basierten Maß nicht unmittelbar von der Token-Häufigkeit abhängig ist. Die probabilistische Schätzung der Produktivität zeigt, dass die Token-Häufigkeit als solche, entgegen der landläufigen Meinung, nicht als Stolperstein für Derivationsmuster angesehen werden kann.

Der dritte Teil der Arbeit befasst sich mit der morphologischen Analysierbarkeit und semantischen Transparenz von multimorphemischen Wörtern. Ich schlage vor, zwei Modelle der Bedeutungsverarbeitung komplexer Wörter zu unterscheiden: eines, das auf dem Prinzip der Kompositionalität beruht, und ein anderes, das auf dem Prinzip der Parsabilität basiert. Ich zeige, dass die Wörter, die durch eine größere Diskrepanz zwischen den Übergangswahrscheinlichkeiten von Affix zu Basis und von Basis zu Affix gekennzeichnet sind, eher als parsierbar behandelt werden, während diejenigen mit vergleichbareren (niedrigen) Übergangswahrscheinlichkeiten eher kompositionell verarbeitet werden können.

Contents

List of tables

List of figures

1 Introduction

1.1 Motivation and research goals

It has been noted (Audring, 2022) that while construction grammar arose in the late 1980s (Fillmore, Kay, and O'Connor, 1988), for a considerable time, it was viewed as a theory of syntax — morphology was slow to develop a complementary approach. Booij's seminal monograph on construction morphology only appeared in 2010 (Booij, 2010b), and its 'sibling' theory, relational morphology, was introduced a decade later (Jackendoff and Audring, 2020). Whatever the reasons for this time gap, it had important ramifications, in both empirical and theoretical domains.

On the empirical side, the repertoire of languages and expressions that have appeared on the radar of construction grammatical work remains meagre. Probably the best-studied phenomenon is compounding (Cetnarowska, 2020; Radimský, 2020; Gaeta and Angster, 2019; Arcodia and Basciano, 2018; Bagasheva, 2015, among others). Affixoids have also received some attention; construction-based accounts of these morphological patterns are attested, for example, for German, Dutch (Hartmann, 2019; Hüning and Booij, 2014), Italian (Masini and Micheli, 2020), and Hungarian (Kenesei, 2007). As for multi-morphemic words, the view that complex words instantiate morphological constructions can be found in Croft (2001) and Goldberg (2006). Some examples of the constructional analysis of complex words include those of English *be*-verbs in Petre and Cuyckens (2008) and of the phrasal verbs of Germanic languages in Booij (2010a). Nevertheless, the understanding of the constructional aspects of multi-morphemic word structure is still in its early stages overall.

From a theoretical point of view, this somewhat late development of constructional morphology account resulted in the persistence of several influential ideas from earlier periods and their peculiar coexistence with more modern approaches. The problem is epitomised in how the notions of morphological parsability and productivity of derivational patterns are discussed in the literature. To put this in context, one must consider two types of lexical models that diverge on the question of how to represent morphologically complex words and the units of which they are composed. Specifically, models differ as to whether knowledge about morphemes is explicitly represented as lexical knowledge (Forster and Taft, 1975; Butterworth, 1983; Smolka, Preller, and Eulitz, 2014; Milin, Smolka, and Feldman, 2017).

Lexicon-based (or words-and-rules) models assume the explicit representation of morphemes and of a word's morphological structure. In order to account for the fact that not all complex linguistic forms can be computed by applying combinatoric rules to simpler meaningful elements, a dual-mechanism account has been proposed that allows for the listedness (whole-form storage) of such complex items. The choice of mechanism depends on the degree of compositionality of a given lexeme. If its formation follows a general pattern (*work—work-ed*, *gnome—gnome-s*), then this specific morphological structure need not be stored separately for each lexical entry. If the formation is irregular and obeys no general rule (*break—broke*, *child—children*), then this structure is listed separately (Marslen-Wilson and Tyler, 1998; Pinker, 1999; Pinker and Ullman, 2002, 2003; Ullman, 2004).

Learning-based (or connectionist) models assume no explicit representation of morphemes and instead put forward the idea of non-symbolic patterns of form and meaning processing. These patterns, broadly speaking, may arise in either top-down or bottom-up fashions. The former implies deriving representations based on the connectivity of lexical items (Bybee, 1985, 1995; Harm and Seidenberg, 2004), while the latter implies deriving them based on the connectivity of sublexical units, such as bigrams and trigrams (Baayen et al., 2011; Baayen et al., 2015). Learning-based models are, by definition, single-mechanism models since they assume no divergence in the type of morphological processing for compositional and non-compositional forms. The difference between these forms is believed to be of a quantitative rather than qualitative or categorical nature (Plaut et al., 1996; Patterson et al., 2001; Bird et al., 2003; Bybee and McClelland, 2005; Smolka et al., 2013). For example, irregular forms with significant orthographic and phonological overlap (*draw—drawn*) have shown facilitation comparable to that of regular compositional forms in priming studies (Basnight-Brown et al., 2007).

Now, an interesting point is that construction morphology is closely related to the connectionism with its word-based approach (Audring, 2022). However, in discussing the morphological parsability and semantic transparency of complex words, many researchers draw on the ideas of relative lexical frequency developed by Hay (2001), who worked with a dual-route model of perception. In her 2001 article, 'Lexical frequency in morphology: Is everything relative?', Hay proposed a simple and elegant way of assessing complex words' parsability (decomposability). According to Hay, the degree of parsability of a given item depends on the frequency of the derived word relative to its base. With most complex words, the base is more frequent than the derived form, so this relative frequency is less than one. Such words, Hay argued, are more easily decomposed. In the opposite case, when the derived form is more frequent than the base, a whole-word bias in parsing is expected. This has consequences for semantics (such words become less transparent and more polysemous), affix ordering, phonetics (Hay, 2001, 2002, 2003), and morphological productivity (Hay and Baayen, 2002, 2003).

This approach is intuitively appealing and, up to the present day, has been highly accepted in the field (see, for example, Berg, 2013; Pycha, 2013; Diessel, 2019; Saldana, Oseki, and Culbertson, 2021; Zee et al., 2021). However, many researchers who have examined relative frequency effects have noted that they exhibit inconsistency and may not hold up across contexts or languages. In fact, over the years, contradicting evidence has been accruing in every domain where relative frequency was believed to play a role (see Chapter 7 for discussion). The problem, as I see it, is rooted in the fact that the morpheme-based approach and relative frequency account do not make an allowance for one additional meaning processing mechanism, which construction morphology can identify due to its ability to distinguish between fixed elements and slots (variables) (Culicover and Jackendoff, 2005; Jackendoff, 2008; Booij, 2010; Diessel, 2019).

Simply, for a two-element complex expression — for example, a prefix or particle verb — one can have four possible combinations: (1) both elements are fixed; (2) both elements are variable; (3) the first element is variable, and the second element is fixed; and (4) the first element is fixed, and the second element is variable. Linguistic items of type (1) are non-analysable, non-compositional, and non-productive. They are listed diachronic relics that are not assembled on the fly but are, rather, retrieved from the lexicon. Linguistic items of type (2) are, in contrast, analysable, fully compositional, and productive. For types (1) and (2), there is little divergence between the dual-route model and construction morphology accounts. However, with types (3) and (4), the situation is more complicated.

In a sense, the very design of these constructions predetermines the relative frequency relation between the whole form and the base. Since one fixed element normally appears in many words, combined with different elements that fill the respective construction's empty slot (as in Russian *na-pisatj* 'write on', *v-pisatj* 'write in', *nad-pisatj* 'write above', *pod-pisatj* 'write under'), it is expected that in complex words of type (3), where the base is fixed, the derivation to base frequency ratio will tend to be less than one. In contrast, complex words of type (4), where the base serves as a filler (as in German *auf-klären* 'clear up', *auf-bessern* 'polish up', *auf-schaukeln* 'build up', *auf-modeln* 'spruce up'), will most likely reveal derivation to base frequency ratios greater than one. Therefore, one can expect that for the dual-route model, expressions of type (3) will be indistinguishable from those of type (2), and expressions of type (4) will be conflated with those of type (1).

The problem carries over to the domain of morphological productivity. One influential theory claiming that the relationship between affixes' productivity and analysability is a strong positive correlation was first formulated by Hay and Baayen (2002). They used Baayen's hapax-based measure (Baayen, 1991, 1992, 1994, 2009; Baayen and Lieber, 1991; Baayen and Renouf, 1996; Plag, 2021) to evaluate affixes' productivity and proposed the notion of parsing ratio to evaluate affixes' analysability. For each affix, its parsing ratio gives the probability that a certain word with this affix will be decomposed by a language user during access (Hay and Baayen, 2003). Mathematically, a parsing ratio is defined as the proportion of forms (types or tokens) that fall above a so-called parsing line given by the following equation: log(base frequency) = 3.76 + .76 * log(derivation frequency) (Hay and Baayen, 2002).

It is clear that the notion of parsing ratio builds upon the logic of the relative frequency account of analysability and, thus, reveals the same dual-route mode of thinking. However, the logic behind this approach is not easy to reconcile with the tenets of construction morphology. While Hay and Baayen's way of estimating linguistic productivity seems perfectly justified for words of type (1) (which are non-analysable and thus cannot add anything to the productivity of their affixes) and words of type (2) (which are compositional and hence bear witness to their affixes' wide applicability), the picture is not so clear with expressions of types (3) and (4).

For example, the derivational elements in multi-morphemic words or multi-word expressions of type (4) in German, Russian, and English, being fixed by construction, are often called 'semiproductive' in the literature (Jackendoff, 2002) in the sense that they have input limitations. In other words, they do not accommodate every base that is semantically compatible with the preverb, prefix, or particle (McIntyre, 2001; Blom, 2005). Nevertheless, these constructions often instantiate large groups of words (see the detailed discussions of specific preverb, prefix, and particle uses in German, Russian, and English in Kühnhold and Wellmann, 1973; Stiebels, 1996; Krongauz, 1998; Larsen, 2014), which, notably, are open to new members. Hence, it seems unwise to treat them on a par with listed diachronic relics of type (1) as the relative frequency account would, in most cases, suggest.

The above considerations shaped the motivation for this thesis and its main research goals. First, from an empirical perspective, I aim to broaden the repertoire of languages for which the multi-morphemic words have been studied within the framework of construction morphology. To this end, I provide a thorough analysis of Russian prefixed verbs as instantiations of two different types of prefix-base constructions: one with an open slot for the prefix and a fixed base verb and another with a fixed prefix and an open slot for the base verb. Second, from a theoretical perspective, I suggest a substantial rethinking of some still-popular approaches to the problems of morphological parsability and productivity of linguistic expressions. Specifically, I strive to leave behind what has been inherited in these domains from the dual-route theory of word recognition and instead find a firm construction-morphology footing for the analysis of these phenomena. This part of the work is substantiated by corpus and experimental data from both Russian and English. More specific problem statements, as well as overviews and discussions of previously published studies, are deferred to the relevant chapters.

1.2 Thesis structure

The thesis is structured as follows. The current introduction serves as **Chapter 1**. In **Chapter 2**, I test the morphological gradience theory on Russian prefixed verbs. Using a specially designed experiment in which participants were asked to evaluate the semantic transparency of a prefixed nonce verb given in minimal context and semanticise it by suggesting an existing Russian verb with the same prefix, I offer evidence that these verbs can be analysed as constructional schemas and that the degree of their morphological decomposition depends upon different levels of activation of their sequential and lexical links. I prove that speakers of Russian are highly sensitive to the etymological connections between verb prefixes and the prepositions to which they are related. Thus, prefix-base constructions with prefixes that correspond to prepositions are more likely to be morphologically decomposed, while prefix-base constructions with prefixes that do not relate to prepositions tend to be regarded as single lexical units. Moreover, the general, highly abstract semantics of Russian prefix-base constructions, especially those that retain their prepositional meanings, is accessible to language users. This is confirmed by the fact that the interpretability of these constructions is affected by priming.

Chapter 3 provides yet another way of testing the morphological gradience theory on Russian prefixed verbs. I offer experimental evidence that verbs with prefixes that have prepositional counterparts and verbs with prefixes that only exist as bound morphemes reveal significant differences in terms of their morphological decomposition. In the pronunciation of native speakers, there is a significantly longer pause between prepositional prefixes and bases than between unprepositional prefixes and bases due to the compositional nature of the former and the non-compositional nature of the latter. Drawing on these findings, I contend that Russian prefixed verbs can be analysed as constructional schemas and that the degree of their morphological decomposition depends upon the different levels of activation of their sequential and lexical links. I also show that the production of prefix-base constructions is governed by audience design. Speakers of Russian, when confronted with a construction that has multiple meanings, try to disambiguate those meanings by employing a greater pause between the prefix and base in order to flag construction-specific meanings. As a result, verbs with prefixes unrelated to prepositions may contextually behave like verbs with prepositional prefixes and vice versa, depending on how the speaker interprets the meaning of a particular construction and how explicit he or she wants to make this meaning for the hearer. Overall, my findings strongly support the ideas that morphological structure is gradient and shaped by language use and that morphological decomposition is a matter of degree.

The follow-up study described in **Chapter 4** was designed to verify the results of the experiment reported in Chapter 3. Here, I contend that the observed differences hold under a very different experimental design. Specifically, the findings are the same if one takes absolute rather than relative lengths of the inter-morpheme periods of silence into account, if one controls for all phonetic differences in target verbs and considers only the variability that is left unexplained by these factors, and even if one replaces real bases in target verbs with nonce bases while retaining the prefixes. I conclude that the observed results pertain not to the participants' familiarity with the target verbs (their parsability, language frequency, etc.) but rather to their familiarity with the relevant prefix-base constructional schemas.

In **Chapter 5**, I provide evidence that the inveterate way of assessing linguistic productivity by calculating the ratio of the number of hapax legomena with a given affix to the total number of tokens with that affix may be biased. As such, it renders any conclusions about the correlation between an affix's productivity and parsability dubious. I propose a new method of evaluating morphological productivity that, unlike the hapax-based measure, is not immediately dependent on token frequency. Specifically, I argue that productivity may be viewed as the probability of an affix to combine with a random base, and I suggest an algorithm for estimating this probability.

Chapter 6 challenges the long-established view that since the morphological structure of frequent words is non-transparent, affixes which are encountered in many frequent items become less parsable and, by virtue of that, lose their ability to combine with new bases. The current chapter shows that the real picture may be more complicated. I argue that high-frequency derivations with an affix, once they are accumulated in a certain number of types, do not block the emergence of new low-frequency coinages but rather facilitate it, paving the way for neologisms. What seems to determine the linguistic productivity of a derivational pattern is not the proportion of infrequent or parsable words among all words with a specific affix but rather the proportion of high-frequency items that strongly collocate with their bases.

In **Chapter 7**, I make a case for distinguishing between two meaning-processing models for complex words: one based on the principle of compositionality and another on the principle of parsability. I point out that the difference between these models might be obfuscated if one assesses complex words' degrees of analysability by calculating their derivation to base frequency ratios. I propose replacing this traditional measure with the ratio of two transitional probabilities: P (affix | base) and P (base | affix). When transitional probabilities are comparably low, each of the elements entering into the combination is equally free to vary. The combination itself is judged by language users to be semantically transparent, and its derivational element tends to be more linguistically productive. On the other hand, multi-morphemic words that are characterised by greater discrepancies between transitional probabilities are similar to collocations in the sense that they also consist of a node (conditionally independent element) and a collocate (conditionally dependent element). Such linguistic expressions, though semantically complex, appear less transparent because the collocate's meaning does not coincide with the meaning of the respective free element (even if it exists) and must be parsed out from what is available.

In **Chapter 8**, I draw on the idea of two different meaning-processing models and propose an account of how complex verbs acquire their construction-specific, idiosyncratic meanings. Complex verbs with the same preverb/prefix/particle that is both linguistically productive and analysable can be compositional as well as non-compositional in meaning. For example, English *on* has compositional spatial uses (*put a hat on*) but also a non-spatial 'continuative' use in which its semantic contribution is consistent with multiple verbs (*we played / worked / talked on despite the interruption*). Comparable examples can be given with German preverbs or with Russian prefixes, which are the main data analysed in this chapter. The preverbs/prefixes/particles that encode non-compositional, construction-specific senses have been studied extensively, but it is still far from clear how their semantic idiosyncrasies arise. Even when one can identify the contribution of the base, it is counterintuitive to assign the remaining sememes to the preverb/prefix/particle. Therefore, on the one hand, there seems to be an element without meaning, and on the other, there is a word sense that apparently comes from nowhere. I suggest analysing compositional and non-compositional complex verbs as instantiations of two different types of constructions: one with an open slot for the preverb/prefix/particle and a fixed base verb and another with a fixed preverb/prefix/particle and an open slot for the base verb. Both experimental and corpus evidence supporting this decision is provided for Russian data. I argue that each construction implies its own meaning-processing model and that the actual choice between these constructions can be predicted by taking into account the discrepancy in probabilities of transition from the preverb/prefix/particle to the base and from the base to the preverb/prefix/particle.

Finally, **Chapter 9** summarises the main findings and results and concludes this work.

1.3 Main research contributions

The main contributions of this thesis can be summarised as follows. First, I conducted a thorough analysis of Russian complex verbs as prefix-base constructions, which has never been done before. The analysis included an investigation of the pronunciation, semantics, acceptability, analysability, and productivity of these word-level constructions. The pieces of evidence from all these domains are consistent with each other and, taken together, indicate that a construction-morphology approach to Russian complex verbs is justified. This is important because it is only with the notion of construction as a 'conventional schema for creating or motivating well-formed expressions in which there is at least one open slot' (Haspelmath, 2023: 1) in mind that one can account for the idiosyncratic behaviour of verbal prefixes in Russian (Monakhov, 2021).

Second, I showed that several token-frequency-based measures bequeathed to present-day researchers by the proponents of the dual-route model are, in fact, in no good agreement with the fundamental tenets of construction morphology. The derivation-to-base frequency ratio used to evaluate complex words' degree of analysability and the ratio of the number of hapax legomena with a given affix to the total number of tokens with that affix used to evaluate the degree of derivational patterns' productivity only work under the assumption that multi-morphemic and multi-word expressions can be either holistic or compositional.

Unfortunately, this account falls short of explaining the fact that in many languages, there are linguistic items which, with regard to their semantics, can be called neither compositional nor non-compositional. These items cannot be called compositional in the traditional Langacker's (1987) sense because their general meaning cannot be inferred from the meaning of their components. However, it feels incorrect to call them non-compositional because, often, their fixed elements make the same semantic contribution in multiple words (Larsen, 2014; McIntyre, 2002). Moreover, it is well-known that German, Russian, and English non-spatial complex verbs with certain preverbs, prefixes, or particles often come in groups of numerous members such that the meanings of derivations are almost identical, and yet the meanings of the bases might have nothing in common (Stiebel, 1996; Zeller, 2001; Monakhov, 2023a).

Taking this into account, I proposed distinguishing between two meaning processing models for complex words: one based on the principle of compositionality and another on the principle of parsability. The distinction between the two models is not a clear-cut categorical one but rather a probabilistic continuum. One can predict which model — compositional or parsable — is more likely to be chosen for each word by taking into account the word's two morphological families: one for the affix and another for the base. The words that are characterised by a greater discrepancy between transitional probabilities from affix to base and from base to affix are more likely to be treated as parsable than those with more comparable (low) transitional probabilities.

Third, I showed that the difference between the two meaning-processing models has predictable implications for affixes' morphological productivity. A high proportion of parsable words among all derivations with a certain prefix might be taken as a sign of the prefix's constrained productivity. If, among multi-morphemic words with a certain prefix, there are many words whose bases are conditionally dependent upon the prefix — that is, there is a strong sequential link between the elements — the prefix's range of applicability is limited, and the constructional meaning is not general enough to accommodate a wide variety of items in its slot.

This behaviour can hardly be captured by the traditional hapax-based measure of linguistic productivity, so I suggested that linguistic productivity should be viewed as the probability of an affix to combine with a random base. The advantages of this approach include the following: (1) token frequency does not dominate the productivity measure but naturally influences the sampling of bases; (2) one does not just count attested word types with an affix but rather simulates the construction of these types and then checks whether they are attested in the corpus; and (3) a corpus-based approach and randomised design assure that true neologisms and words coined long ago have equal chances to be selected (Monakhov, 2023b).

Among other, more technical contributions made by the current thesis, the following should be mentioned. First, I compiled a variety of datasets that can be used for future work on Russian and English complex words. Second, I presented some new algorithms, methods, and experimental setups that are potentially applicable to a variety of other research questions.

1.4 Publications

Parts of the research in this thesis have been published as separate articles. These are given under my name on the reference list.

2 Russian prefixed verbs as constructional schemas

2.1 Introduction

In 1928, a prominent Russian writer and literary critic Kornej Čukovskij argued that there existed no word in the Russian language that a child could not turn into a verb. In his book *Ot dvux do pjati* 'From Two to Five' ([1928]2001), he cited a great number of such coinages that he encountered in the speech of his grandchildren. Among them were, for example:

(1)	*Ot-skorlupa-j*	*mne*	*jajco.*
	from-eggshell-IMP.2SG	for me	egg.
	'Peel me an egg.'		

(2)	*Za-molotoč-j*	*ėtot*	*gvozdik.*
	at-hammer-IMP.2SG	this	nail.
	'Slam this nail.'		

(3)	*Ja*	*na-makaroni-l-sja.*
	I	on-noodle-PST.MASC-REFL
	'I have eaten enough noodles.'	

(Čukovskij, [1928]2001: 31)

The mechanism of this word formation is the same: verbal derivational (prefix) and inflectional (suffix) morphemes are combined with noun stems. Čukovskij contended that children are equally productive in forming both prefixed and non-prefixed verbs; however, the former constitute the majority of his examples. That should come as no surprise to any speaker of Russian, who will agree that the meaning of *na-makaronitsja* is somehow more intuitively clear than the meaning of *makaronitsja*—a verb with the same base but without a prefix.[1]

[1] By being 'more' or 'less intuitively clear', I mean being able or not being able to be interpreted without a context.

These patterns of verb formation are in no way confined to the language of early childhood. Such examples abound throughout Russian history, the most successful of them even became part of the standard language. In fact, the creative potential of these patterns is so great that they allow the incorporation not only of common nouns but also proper nouns, which is a common way of making Russian bon-mots. Such novel prefixed verbs are created to describe an action that is considered characteristic of a certain person. When English journalists were disappointed by the performance of Russian striker Alexander Kerzhakov during the 2012 UEFA European Football Championship, they coined the verb *to kerzhakov* ‘to miss a wide-open goal, especially in a situation in which it is almost impossible to miss’. It was translated into Russian as *s-kerzhakov-itj*, aligning with many other verbs of the same constructional schema: *s-glup-itj* ‘to make a fool of oneself’, *s-ploch-ovatj* ‘to blunder’, etc.

Going all the way down this path brings me to the possibility of coining prefixed verbs with bases that are semantically void and, when considered apart from their prefixes, have no meaning at all. For example: *za-findilitj* ‘to land a blow’ ← **findil*, *u-khajdokatj* ‘to bring to an end of existence’ ← **khajdok*, and so on.

Russian verb prefixes (more precisely, the use of prefixes, since most of them are polysemous) are usually divided into two types: ‘external/superlexical’ and ‘internal/lexical’. They are distinguished by a whole range of semantic and formal properties, specifically, with regard to their linear position. The prefixes of the first type are said to express modes of action, such as the inchoative (*za-igratj* ‘start playing’, *za-prygatj* ‘start jumping’), the delimitative (*po-igratj* ‘play for some time’, *po-begatj* ‘run for some time’), and the distributive (*pere-igratj vo vse igry* ‘play all games’, *pere-lovitj vsech myshej* ‘catch all mice’). They are more compositional in terms of interaction with the semantics of the base than the prefixes of the second type and usually precede them. The prefixes of the second type encode mostly spatial meanings (*v-bezhatj* ‘run into’), they tend to be closer to the base and show a greater degree of semantic cohesion with it. Importantly, one prefix can convey both types of meaning, cf.: *za-igratj* ‘start playing’ and *za-bezhatj za ugol* ‘run around the corner’ (Babko-Malaya, 1999; Ramchand, 2004; Romanova, 2004; Tatevosov, 2008).

Regardless of the prefix category, traditional grammar has always propagated the idea of the ‘semantic double-centeredness’ of Russian prefixed verbs, treating them as syntagmas consisting of two structural components of which the base bears the main burden of lexical meaning, while the prefix shapes and categorises this meaning in terms of some primitive semantic concepts (Miloslavskij, 1980; Aminova, 1988; Volohina and Popova, 1993; Varaksin, 1996; Biskup, 2019). As for the examples like *zafindilitj* or *uchajdokatj* and other similar constructions, they are traditionally dismissed as occasionalisms that are created and understood due to an analogy with fully semanticised prefixed verbs.

There have been attempts to shift the burden of meaning of Russian prefixed verbs from bases to prefixes, most notably that of Krongauz (1998), who clearly understood the vulnerability of the traditional approach but, regrettably, did not come up with an appropriate methodological framework to convey his ideas and discredited them through some overly simplistic argumentation. Thus, he claimed that the meaning of a prefix is the general meaning of a group of all synonymous verbs with this prefix, which, for example, led him to contend that the prefix *s-* has the meaning 'to steal' (because of the verbs with the same meaning: *s-tyritj*, *s-ljamzitj*, *s-peretj*, *s-tjanutj*, etc.) and the prefix *ot-* has the meaning 'to beatj (because of the verbs with the same meaning: *ot-koloshmatitj*, *ot-dubasitj*, *ot-pizditj*, *ot-mudochatj*, etc.). Naturally, explanations of that sort were deemed incongruous and criticised (Beliakov, 1999: 215–216).

I believe that there is a linguistic theory allowing one to analyse Russian prefix-base constructions in a more logical and effective way. That theory is Construction Grammar, the study of symbolic pairings of form and meaning that are characterised by structural or semantic / pragmatic idiosyncrasies and / or a high level of entrenchment in language (Diessel, 2019; Hilpert, 2014; Langacker, 2009; Goldberg, 2006; Croft, 2001). Construction Grammar has recently become one of the most prominent frameworks of linguistic research. Since 1995, when Goldberg's seminal book outlined the theoretical underpinnings of Construction Grammar, significant progress has been made. Linguists proceeded from compiling an inventory of the different possible types of constructions to charting an entire network of constructions that is arguably capable of embracing the whole language domain and explaining every phenomenon within it.

As bilateral linguistic signs, constructions are believed to form a cline stretching from morphological units consisting of at least one bound morpheme and one slot for a free morpheme to syntactic units consisting of two or more slots for free lexemes. In relation to complex words, an important part of the Construction Grammar framework is the idea of morphological gradience. It implies that complex words can be accessed in discourse either via a route of morphological decomposition or via a direct-access, non-decomposed route, depending on their absolute frequency and the relative frequency of their parts (Baayen and Schreuder, 2000; Hay, 2001).

According to this view, the processing of complex words is determined by two types of links. On the one hand, it involves a sequential (syntagmatic, combinatorial) link between a free and a bound morpheme. On the other hand, it activates lexical (paradigmatic, categorising) links to similar words within the network of constructions. It has been demonstrated that the more frequently a word is used, the more automatised and predictable a sequential link between its parts becomes. As a consequence, frequent complex words tend to be structurally and semantically less transparent than infrequent words (Bybee, 1985, 2007; Hay, 2003).

To the best of my knowledge, the theory of morphological gradience has been evaluated predominantly (if not exclusively) against English data. Russian language and specifically Russian prefixed verbs seem to constitute an interesting case in this regard. First, prefixation is very productive in Russian. Some studies demonstrate that up to 90% of all Russian verbs are derived by this means (Tixonov, 1998: 17). Second, it is well known that in Russian and many other Indo-European (especially Slavic) languages, some prefixes are related to prepositions and retain much of their spatial meaning, while others have a different etymological background (Matushansky, 2002; Richardson, 2007; Lehmann, 2009; Markova, 2011; Biskup, 2012; Wiland, 2012).

Provided that a usage-based constructionist approach to Russian prefixed verbs is justified and they can be analysed as prefix-base constructions, one expects to find two things. First, verbs with prefixes which have prepositional counterparts and verbs with prefixes which exist only as bound morphemes should reveal significant difference in terms of their morphological decomposition and degree of semantic transparency. While earlier studies had shown that the total frequency of complex words strengthens their status as lexical units (Bybee, 1985: 117–124), Hay (2003: 88–95) argued that the processing of lexical units is also influenced by the relative frequency of a complex word and its parts.

Second, the meaning of the verbs with prefixes related to prepositions should be accessible to speakers of Russian as a general constructional meaning characterised by a high degree of abstraction, while the meaning of the verbs with prefixes unrelated to prepositions should be contextually inferred as a meaning of a particular lexical item. In other words, prefix-base constructions should constitute cognitive entities in their own right, i. e., their general meanings should be to a certain degree semantically independent from the sum of the meanings of their parts. This notion can be illustrated with a very simple example of coercion. Let me take a famous sentence (4) from Goldberg (1995: 29) showing that constructional schemas can override the argument structures of verbs; see (5) for a translation into Russian:

(4)	*John*	*sneezed*	*the napkin*	*off the table.*
(5)	*John*	*s-chikhnul*	*salfetku*	*so stola.*

The English sentence is traditionally explained as follows: the meaning of the caused motion construction [X cause Y to move Z] interacts here with the semantics of the verb so that the verb contributes the agent role while the construction contributes the theme and the goal. The Russian sentence differs from the English original in that it does not simply contain a word that has never been used in this context (*sneeze*), but actually creates a verb that might not have been used before in any context at all (*s-chikhnutj*). In other words, an English syntactic construction is rendered in Russian as a morphological construction PREFIX–[________]$_{\text{BASE}}$ of the same meaning with the goal encoded by a prefix. This is confirmed by the fact that while the English sentence **John sneezed the napkin* is ungrammatical, the Russian sentence *John s-chikhnul salfetku* is not.

Given all of the above, this study tests the morphological gradience theory on Russian prefixed verbs. With the help of a specially designed experiment, in which participants were asked to evaluate the semantic transparency of a prefixed nonce verb given in a minimal context, as well as to semanticise it by suggesting an existing Russian verb with the same prefix, I offer evidence that these verbs can be analysed as constructional schemas and that the degree of their morphological decomposition depends upon the different levels of activation of their sequential and lexical links.

2.2 Experimental design, data and methods

Prefix-base constructions, like any other constructions, must be stored and processed in a network of associations, and access to them must be determined by the activation level of a construction at a particular moment in time (cf. Diessel, 2019: 24–25; 44). One easy method for activating a construction is through the structural priming of it by means of the same or a similar element preceding it in the discourse (Bock, 1986; Pickering and Ferreira, 2008). With this in mind, I designed and conducted my experiment.

The Russian Grammar of the Russian Academy of Sciences (Shvedova, 1980: § 850) lists 28 verbal prefixes:

- 17 prefixes are not only historically related to prepositions, but also have prepositional counterparts in modern Russian: *v-* (*v* ‘in, atj), *do-* (*do* ‘to, before’), *za-* (*za* ‘for, behind’), *iz-* (*iz* ‘from, out of’), *na-* (*na* ‘on’), *nad-* (*nad* ‘over, above’), *o-* (*o* ‘aboutj), *ob-* (*ob* ‘aboutj), *ot-* (*ot* ‘from’), *po-* (*po* ‘along, by’), *pod-* (*pod* ‘under’), *pred-* (*pered* /

pred ‘before, in front of’), *pri-* (*pri* ‘by, atj), *pro-* (*pro* ‘about, of’), *s-* (*s* ‘with’), *so-* (*so* ‘with’), and *u-* (*u* ‘from, by’);

- 11 prefixes have no prepositional counterparts in modern Russian; this group encompasses morphemic borrowings, prefixes that have unprepositional origin and prefixes derived from prepositions that are no longer part of the Russian language: *de-*, *dis-*, *vz-*, *voz-*, *vy-*, *nedo-*, *niz-*, *pere-*, *pre-*, *raz-*, and *re-*.

Almost all Russian verbal prefixes, both prepositional and unprepositional, are polysemous with the number of meanings ranging from 2 (for example, *v-*) to 10 (for example, *pere-*). For the experiment, all meanings of all prefixes listed by the Russian Grammar were taken into consideration (91 meanings for prepositional prefixes and 34 meanings for unprepositional prefixes, 125 in total). For each meaning, one sentence containing a respective verb was obtained from the Russian National Corpus, all sentences being approximately of the same length. In each of these sentences, the root of the target prefixed verb was substituted with the nonce root *-banksi-*.

Next, two experimental conditions were designed. In the first condition, each of the 125 target sentences was preceded by another sentence obtained from the Russian National Corpus in which the same prefix of the same meaning was used with a different verbal base. In the second condition, the preceding sentences were chosen so that they contained verbs that had different prefixes, or no prefixes at all, but were contextually synonymous to the coded target verb. This procedure is illustrated below with the help of an example. The whole array of target and priming verbs as well as all the meanings of the prefixes can be found in Appendix 1.

One of the meanings of the prefix *pro-* is ‘to perform (bring to fruition) an action identified by the base verb’. As a target sentence, I chose *Ja takim obrazom pro-demonstrirovala, chto legkodostupna!* ‘I have thus demonstrated that I am easily accessible!’ This sentence contains the prefixed verb *prodemonstrirova-l-a* ‘demonstrate-PST-3SG.FEM with the aforementioned general meaning. This verb was coded in the experiment in both conditions as the nonce verb *probanksila.* Two different sentences were chosen as primes in experimental conditions 1 and 2. The former contained the verb *pro-zvuchatj* ‘to sound’ with the same prefix *pro-* and the same constructional meaning, but with a different lexical meaning. The latter contained the verb *po-kazatj* ‘to show’, which is synonymous with *prodemonstrirovatj* ‘to demonstrate’ in its lexical meaning, but includes a different prefix *po-* (Table 1).

In both experimental conditions, in all 125 contexts, priming sentences preceded the target sentences and were separated from them with a <…> sign.

Table 1. Design of experiment

<table>
<tr><th>condition</th><th>priming sentence</th><th></th><th>target sentence</th></tr>
<tr><td>1</td><td>Golos Lidii Timofeevny pro-zvuchal otkuda-to iz-za ugla <…> [‘Lidia Timofeyevna’s voice sounded from somewhere round the corner <…>’]</td><td rowspan="2">+</td><td rowspan="2">Ja takim obrazom pro-banksila, chto legkodostupna! [‘I have thus [demonstrated] that I am easily accessible!’]</td></tr>
<tr><td>2</td><td>Sadisj, ja tebe po-kazhu. Ja sela rjadom, chuvstvuja sebja po-glupomu <…> [‘Sit down, I’ll show you. I sat down next to him, feeling stupid <…>’]</td></tr>
</table>

The instructions for the participants of the experiment were written so as not to reveal the true purpose of study. See the full text translated from Russian:

> Hello! Thank you for agreeing to participate in my experiment. The experiment does not require any special knowledge; the only requirement is to be a native Russian speaker. The purpose of the experiment is to investigate the conditions of semanticization (inference of meaning) of Russian verbs through their immediate context. The experimental material includes short excerpts from works of different genres and different time epochs extracted from the Russian National Corpus. In each excerpt, several parts of the original text were deleted. The places of deletion are marked with this sign: <…>. The bases of the target verbs were consistently replaced with the same nonce base *-banksi-*.
>
> You are asked to do the following:
>
> Part A. Rate on a scale of 1 to 4, how intuitively well you understand the meaning of the nonce word (it is CAPITALISED):
>
> 1—the meaning is absolutely incomprehensible,
>
> 2—the meaning is rather more opaque than clear,
>
> 3—the meaning is rather more clear than opaque,
>
> 4—the meaning is absolutely comprehensible.
>
> Part B. Substitute the nonce word, as you understand it, with any existing Russian verb, replacing the nonce base *-banksi-* and preserving all other elements (beginning and end) of

the verb. For example: *protivobanksitj—protivodejstvovatj* 'counteract' OR *protivostoyatj* 'resist' etc.

In most cases, it is possible to opt for several different words at once. Please choose the one that, in your opinion, is most appropriate in this context. Please note that your answer must contain the same prefix as the nonce word! If the meaning of the nonce word is absolutely unclear to you and you choose '1' in Part A, please still suggest the first verb that comes to your mind with the corresponding prefix in Part B.

Please rely only on your language competence when performing the task; do not use any information sources (corpuses, dictionaries, etc.). It is advisable to carry out the tasks quickly, without thinking about your answers for long periods—your natural reaction to the proposed stimulus is important.

Once you have started the task, please complete it. You should not skip sentences. Please bear in mind that tasks that have been completed in a shorter time than it takes to read them will not be accepted. Answers with non-existing verbs or verbs with prefixes different from those of the nonce word will also be considered inappropriate.

The contexts were randomly shuffled, so that different meanings of the same prefix did not follow each other. Given the abundance of sentences, I decided not to add any filler contexts.

To conduct the experiment, I used Yandex Toloka, a Russian crowdsourcing service analogous to Amazon Mechanical Turk that allows the analysis of large volumes of data in a short time. For example, one can ask users to categorise the wide variety of items in an online store into groups, find or verify specific information, translate texts, and so on. First, I created a special template so that each task included one of the 125 pairs of sentences as input data, as well as two fields for output data: 1) an integer varying from 1 to 4 to rate the 'clearness' (comprehensibility) of a nonce verb and 2) a string field to substitute the nonce word with an existing Russian verb with the same prefix. For each task, a time limit of 10 minutes was imposed.

Second, I assembled four pools of users who met the following criteria: 1) they were native speakers of Russian; and 2) they belonged to the top 10% of all rated active users. Participants of each pool were assigned to one of the four groups of tasks: experimental condition 1 (code 1_1) and 2 (code 1_2) for verbs with prepositional prefixes and experimental condition 1 (code 2_1) and 2 (code 2_2) for verbs with unprepositional prefixes. Each task had to be performed by 33 different users, and no user could see any tasks other than those assigned to their pool.

The null hypothesis H_0 of the experiment was that there would be no significant difference between the two experimental conditions both in terms of comprehensibility and interpretability of the coded verbs. The alternative hypothesis H_1 was that verbs with prefixes related to prepositions would reveal significantly higher scores than verbs with unprepositional prefixes.

2.3 Results and discussion

2.3.1 Clearness scores

The total number of submissions was 8,250 (125 meanings x 2 experimental conditions x 33 participants); on average, each participant performed nine tasks. Out of those submissions, 1,856 were erroneous due to one of the following reasons: either no substitute verb at all was provided (818) or the provided verb had a prefix which did not match that of the nonce verb (1,038). I found a significant association between the experimental condition and the number of right, wrong, and no answers: $\chi^2(6) = 371.99$, $p < 0.001$, Cramer's $V = 0.15$ (R Core Team, 2022).

On average, tasks with unprepositional prefixes (2_1 and 2_2) produced a significantly greater numbers of blanks and wrong submissions than prepositional ones. Conversely, the odds of obtaining a correct answer from the test participants were 2.42 times greater if the task included a prepositional verb (especially in experimental condition 1_2) than those if it did not.

First, I analysed the distribution of clearness scores provided by the participants in two experimental conditions, having preliminarily excluded the ratings that were given alone, without a substitute verb, as this type of submission is indicative of answering the question without proper consideration. The scores were given on a scale of 1 to 4, for which 1 indicated absolute incomprehensibility and 4 perfect comprehensibility of the nonce verb in a given context. My alternative hypotheses were that 1) the median clearness score for prepositional prefix-base constructions would be significantly greater than the median clearness score for unprepositional constructions regardless of experimental condition; and 2) the median clearness score in experimental condition 1 with its structural priming of prefix-base constructions would be significantly greater than the median clearness score in experimental condition 2 with its 'lexical boost' (Pickering and Branigan, 1998) regardless of the type of construction.

For each context, I calculated the sum of all participants' clearness scores, which gave me four numeric vectors, two of 91 numbers for prepositional prefixes in experimental conditions 1_1 and 1_2 and two of 34 numbers for unprepositional prefixes in experimental conditions 2_1 and 2_2, each number ranging between 33 (a hypothetical situation in which each participant submitted a score of 1) and 132 (a hypothetical situation in which each participant submitted a score of 4). The overall distribution of median values can be seen in Figure 1: $M_{1\_1} = 84$, $M_{1\_2} = 87$, $M_{2\_1} = 74$, $M_{2\_2} = 72$.

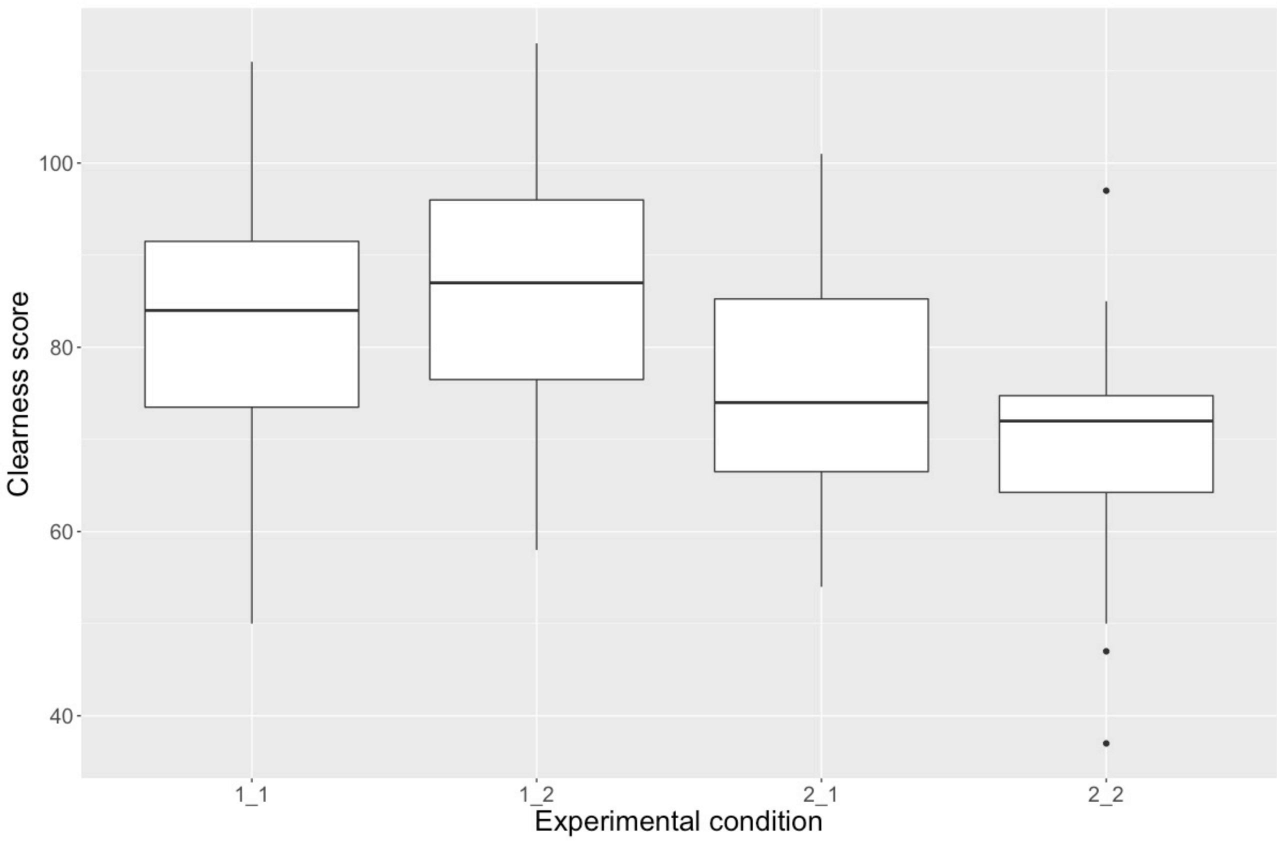

Figure 1. Boxplot of clearness scores

Since my data fall into four groups and are ordinal-scaled, I used a non-parametric ANOVA based on permutation to compare the medians of all groups (Sheskin, 2011: 1002). This was achieved with the help of the *oneway_test()* function from the package *coin* for RStudio and verified by the Kruskal-Wallis one-way ANOVA by ranks implemented in the *stats* package. Both tests showed that the null hypothesis of the true differences of relative effects being equal to 0 can be safely rejected (Approximative K-Sample Fisher-Pitman Permutation Test: $\chi^2(3) = 46.86$, $p < 0.001$; Kruskal-Wallis rank sum test: $\chi^2(3) = 44.42$, $p < 0.001$).

To find out which groups differ significantly, I performed a post hoc non-parametric test of relative contrast effects implemented in the *nparcomp* package for RStudio.

The results provided in Table 2 should be interpreted as follows. Each estimator represents the probability that a randomly chosen subject in treatment group 1 reveals a smaller response value X than a randomly chosen subject from treatment group 2 with response value Y. If this probability is less than 0.5, then the values in group 1 tend to be larger than those in group 2. If the probability equals 0.5, none of the observations tend to be smaller or larger (Konietschke et al., 2015: 2).

Table 2. Non-parametric relative contrast effects (clearness scores)

contrast	difference in medians	statistic	95 % CI		
			lower	estimator	upper
1_1—1_2	84-87	2.00	0.47	0.58	0.68
1_1—2_1	84-74	-2.44	0.22	0.35	0.50
1_1—2_2	84-72	-5.09***	0.12	0.21	0.34
1_2—2_1	87-74	-3.68***	0.16	0.28	0.43
1_2—2_2	87-72	-6.08***	0.08	0.15	0.27
2_1—2_2	74-72	-1.89	0.21	0.36	0.54

Significance codes: ***—$p < 0.001$,*—$p < 0.05$.

It can be observed that of my two initial hypotheses, only the first one has been confirmed. The median clearness score for prepositional prefix-base constructions is indeed significantly greater than the median clearness score for unprepositional constructions in both experimental conditions (see contrasts 1_1—2_2, 1_2—2_1, and 1_2—2_2 in Table 2). However, the difference between experimental conditions 1 and 2 for both prepositional and unprepositional prefixes is negligible (as confirmed both by the *p*-values above the threshold of statistical significance and the fact that the confidence intervals for respective estimators cross 0.5), and, in fact, seems to be quite the opposite of what I had expected. Prepositional prefixes are characterised by higher clearness scores in experimental condition 2, in which a priming sentence contains a verb with a different prefix or no prefix at all, albeit one which is synonymous in meaning to the target verb. In contrast, unprepositional prefixes slightly favour experimental conditional 1, in which a priming sentence contains a verb with the same constructional but different lexical meaning.

One could hypothesise that seeing a verb with the same unprepositional prefix in a priming sentence helped participants of the experiment to 'constructionalise' the respective nonce verb, that is, detach the prefix from the base and thus make the word more semantically transparent. On the other hand, with regard to prepositional prefixes, such a prop turned out to be superfluous or even deluding since the respective prefix-base constructions are easily decomposable as such and have a variety of possible constructional meanings; however, seeing a verb with a different prefix but with a similar lexical meaning helped participants to to arrive at an interpretation.

Both the nuisance of structural priming for prepositional constructions and its importance for unprepositional ones result in the fact that difference between clearness scores in experimental conditions 1_1 and 2_1 is not significant at the conventional 0.05 level ($p = 0.06$).

There is some anecdotal evidence in my data supporting this claim. One of the target sentences for the prefix *vy-* was: *mestnye rybaki vy-banksili v more ne myshonka ne ljagushku a nevedomu zverushku* 'local fishermen vy-banksili into the sea not a mouse not a frog but an unknown animal'. The coded verb was *vy-lovitj* 'catch, fish out'. However, due to the homonymy of locative and accusative case forms of the Russian noun *more* 'sea', the construction could be analysed as meaning both 'to get, obtain, find something by means of an action identified by the base verb' (*vy-lovitj v more* 'catch in the sea', locative form) and 'to move away, to stand out from something, to direct out by means of an action identified by the base verb' (*vy-brositj v more* 'throw into the sea', accusative form). There were two priming sentences, one of them containing the verb *vy-stradatj* 'achieve through suffering', another one the verb *po-jmatj* 'catch'. The results obtained in the two experimental conditions were illuminatingly different (Table 3):

Table 3. Distribution of answers in different experimental conditions (prefix *vy-*)

meanings of substitutes provided for the nonce verb	priming verb	
	vy-stradatj	*po-jmatj*
'to get, obtain, find something'	13	16
'to move away, to direct out'	15	6
no answer	5	11

Note: $\chi^2(2) = 6.41$, $p < 0.05$.

Table 3 shows that though the context suggests the default meaning 'to get, obtain, find something' (after all, fishermen are more likely to be occupied catching something in the sea rather than disposing of something brought there), the *vy*-construction by itself, when activated in the discourse, is primarily connected to the opposite meaning 'to move away, to direct out'.

2.3.2 Correctness scores

It is evident that psychological scaling may be problematic. I cannot be sure that participants treat the distances between the points at the ends of the scale in the same way as the distances between the points in the middle of the scale. Hence, I need a quantitative measure of how well participants actually interpreted the prefix-base constructions. This goal was achieved by manually coding the data and calculating what can be called a 'correctness score'.

The correctness score was designed so that it most closely matched the scale of the clearness score. Each submission was ranked on a scale from 1 to 4 according to the schema provided in Table 4. Cases of no answer were assigned a 1.

Table 4. Correctness scoring schema

score	same prefix	same general meaning	same verb
1	—	—	—
2	+	—	—
3	+	+	—
4	+	+	+

Again, for each context, I calculated the sum of all participants' correctness scores, which provided four numeric vectors, two of 91 numbers for prepositional prefixes in experimental conditions 1_1 and 1_2 and two of 34 numbers for unprepositional prefixes in experimental conditions 2_1 and 2_2, each number ranging from 33 (a hypothetical situation in which each participant provided no answer) to 132 (a hypothetical situation in which each participant provided the exact word from the original context). The overall distribution of values can be seen in Figure 2: $M_{1\_1} = 79$, $M_{1\_2} = 85$, $M_{2\_1} = 75$, $M_{2\_2} = 67$.

The same tests as with the clearness scores were performed, and both of them showed that the null hypothesis of the true differences of relative effects being equal to 0 can be safely rejected (Approximative K-Sample Fisher-Pitman Permutation Test: $\chi^2(3) = 32.32$, $p < 0.001$; Kruskal-Wallis rank sum test: $\chi^2(3) = 30.49$, $p < 0.001$). The results of a post-hoc non-parametric test of relative contrast effects are given in Table 5. The numbers here should be interpreted as those in Table 2. One can see that the distribution of correctness scores is very similar to that of clearness scores, which means that psychological scaling fairly closely mirrored the actual complexity of the picture. The only difference is that the contrast between prepositional and unprepositional prefixes is now restricted to the conditions 1_1—2_2 and 1_2—2_2. As for the pair 1_2—2_1, though participants marked lexically boosted prepositional constructions as more semantically transparent than structurally primed unprepositional constructions, the difference between the numbers of correct substitutions was found insignificant at the conventional 0.05 level ($p = 0.07$).

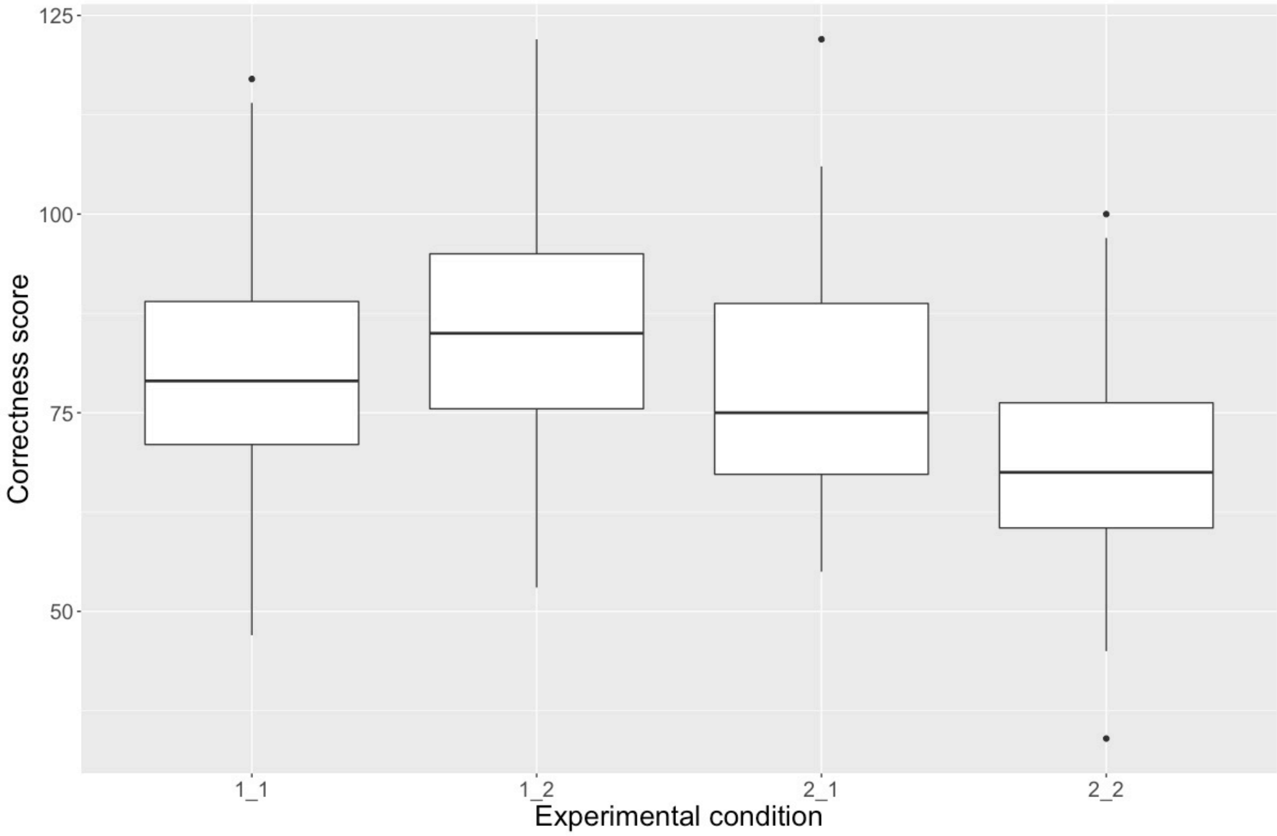

Figure 2. Boxplot of correctness scores

Table 5. Non-parametric relative contrast effects (correctness scores)

contrast	difference in medians	statistic	95 % CI		
			lower	estimator	upper
1_1—1_2	79-85	2.32	0.49	0.60	0.70
1_1—2_1	79-75	-0.85	0.30	0.44	0.60
1_1—2_2	79-67	-3.83***	0.16	0.27	0.41
1_2—2_1	85-75	-2.41	0.22	0.35	0.50
1_2—2_2	85-67	-4.94***	0.10	0.19	0.33
2_1—2_2	75-67	-2.39	0.18	0.32	0.51

Significance codes: ***—p < 0.001.

Table 6. Summary of significant contrasts for clearness and correctness scores

	1_1	1_2	2_1	2_2
1_1		—	—	Clearness + Correctness
1_2	—		Clearness	Clearness + Correctness
2_1	—	Clearness		—
2_2	Clearness + Correctness	Clearness + Correctness	—	

All observed significant contrasts between clearness and correctness scores are summarised in Table 6. I can deduce several conclusions from the above observations. Regardless of experimental condition, prepositional prefixes are distinguished from unprepositional ones. They were rated by the participants as significantly more semantically transparent than their counterparts ($M_{1\_1+1\_2} = 84 > M_{2\_1+2\_2} = 72$). They also produced a greater number of the correct substitutions of coded words ($M_{1\_1+1\_2} = 82 > M_{2\_1+2\_2} = 71$). However, the priming mechanism works very differently with these two types of constructions. The interpretation of the nonce verbs with prepositional prefixes is significantly facilitated by lexical boost (in pairs like *do-bavitj → v-banksitj*), while the interpretation of the nonce verbs with unprepositional prefixes is mostly affected by structural priming (in pairs like *pere-kroitj → pere-banksitj*).

The latter finding is contrary to what I expected and reveals a less straightforward dependence between types of Russian verbal prefixes and complex words' routes of accessibility. A seemingly reasonable explanation of this dependence was provided above, let me reiterate it in more details. When I did my little surgery on prefixed verbs, removing their actual bases and implanting the same nonce base into them, I effectively blocked for these words the direct-access, non-decomposed route.

In agreement with my hypothesis, this operation had more dire consequences for verbs with unprepositional prefixes because it turned them into charades that had to be guessed from the context. It is, then, of little surprise that lexical boost in this situation could not provide the participants of the experiment with sufficient information: they must have experienced troubles even with matching priming verb to the target verb. On the other hand, structural priming of the verbs with unprepositional prefixes helped to constructionalise them, opening the route of morphological decomposition and providing participants with a hint at an interpretation.

Conversely, the verbs with prepositional prefixes did not really require any structural prop because their prefixes, which coincide in form with very frequent prepositions, are easily detachable from the bases on their own. Lexical boost, on the other hand, helped the participants to strengthen the link between general constructional and specific lexical meaning of respective verbs, thus limiting the space of possible interpretations. All of the above can be visualised with the help of the following scheme (Figure 3):

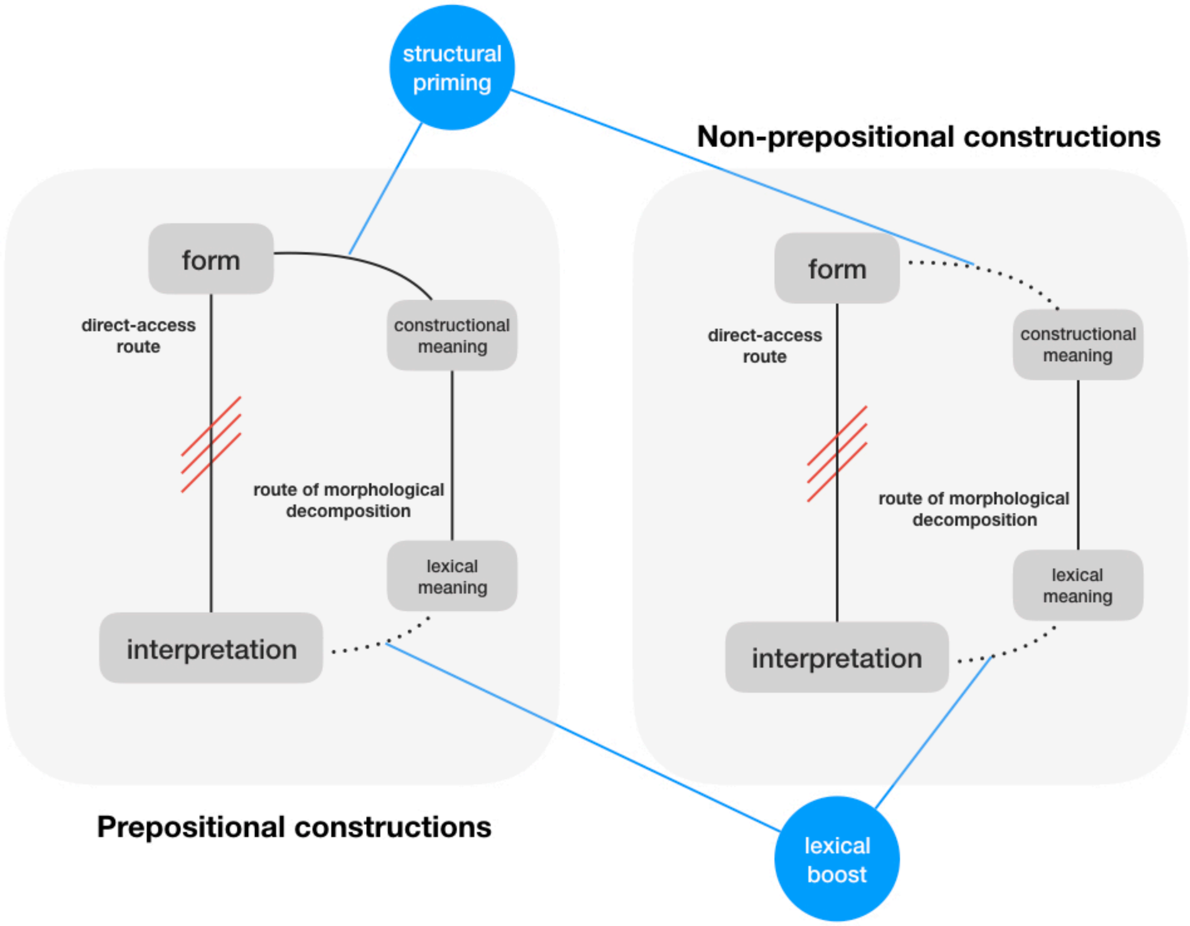

Figure 3. Scheme of Russian prefix-base constructions' routes of accessibility under two types of priming

Notes: 1) three long red dashes symbolise the blocking of a direct-access route during the experiment; 2) black dotted lines represent a weak accessibility along respective lines under experimental conditions; 3) black solid lines represent a strong accessibility along respective lines; 4) blue solid lines point at the links that different types of priming reinforce.

This information seems to provide reliable evidence that priming affects the interpretability of Russian nonce verbs with prepositional and unprepositional prefixes in different ways. This confirms my hypothesis that the former should be considered constructional schemas with a fixed element and a slot that can be filled with certain other elements, that is, prefix-base constructions that are stored and processed in a network of associations, while the latter should be analysed as one-chunk lexical units whose constructional nature is opaque to language users if not activated in the context.

2.3.3 Idiosyncratic behaviour of prefixes

An important question to answer is whether an interaction exists between a prefix and the type of priming, in other words, whether any prefixes reveal idiosyncratic behaviour under different experimental conditions. The interaction plots for clearness and correctness scores are presented in Figure 4 for prepositional prefixes and in Figure 5 for unprepositional prefixes.

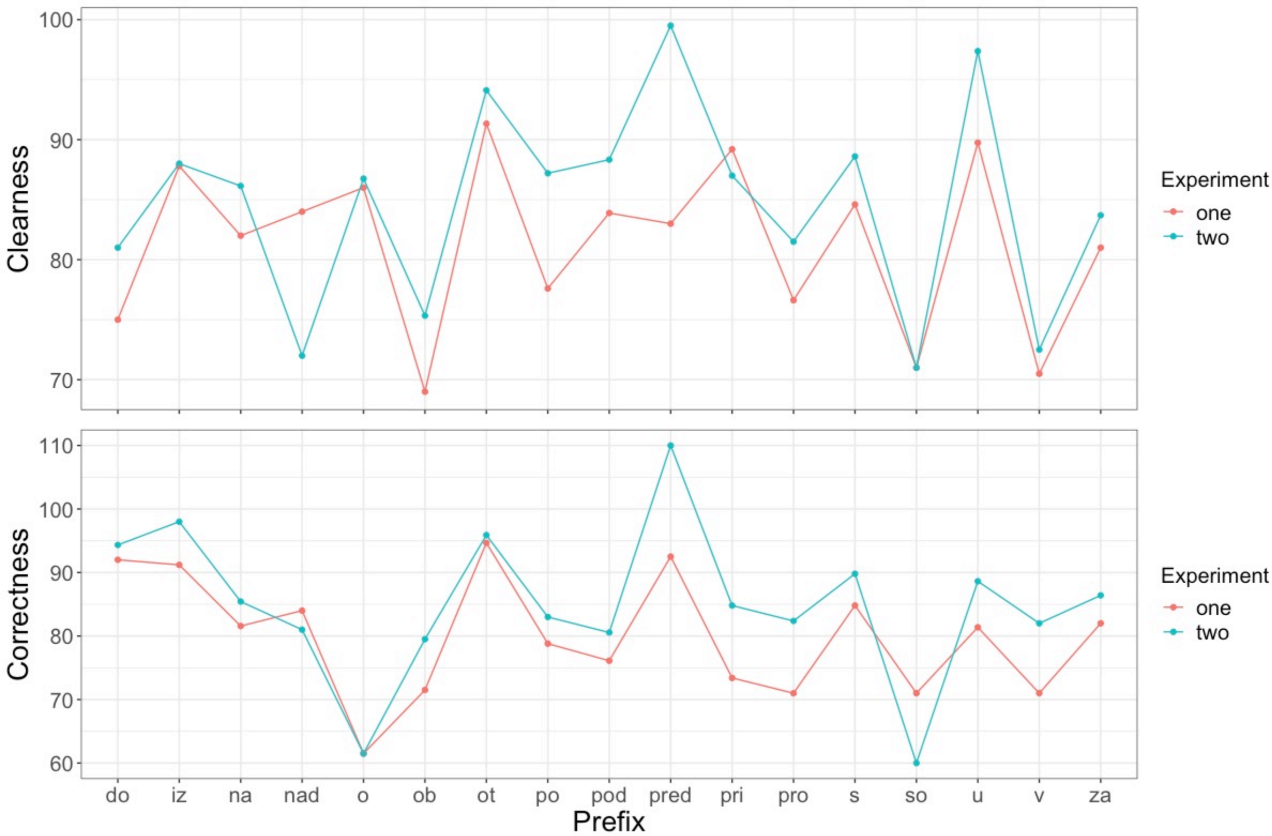

Figure 4. Interaction plot of clearness and correctness scores in experimental conditions 1 and 2 (prepositional prefixes)

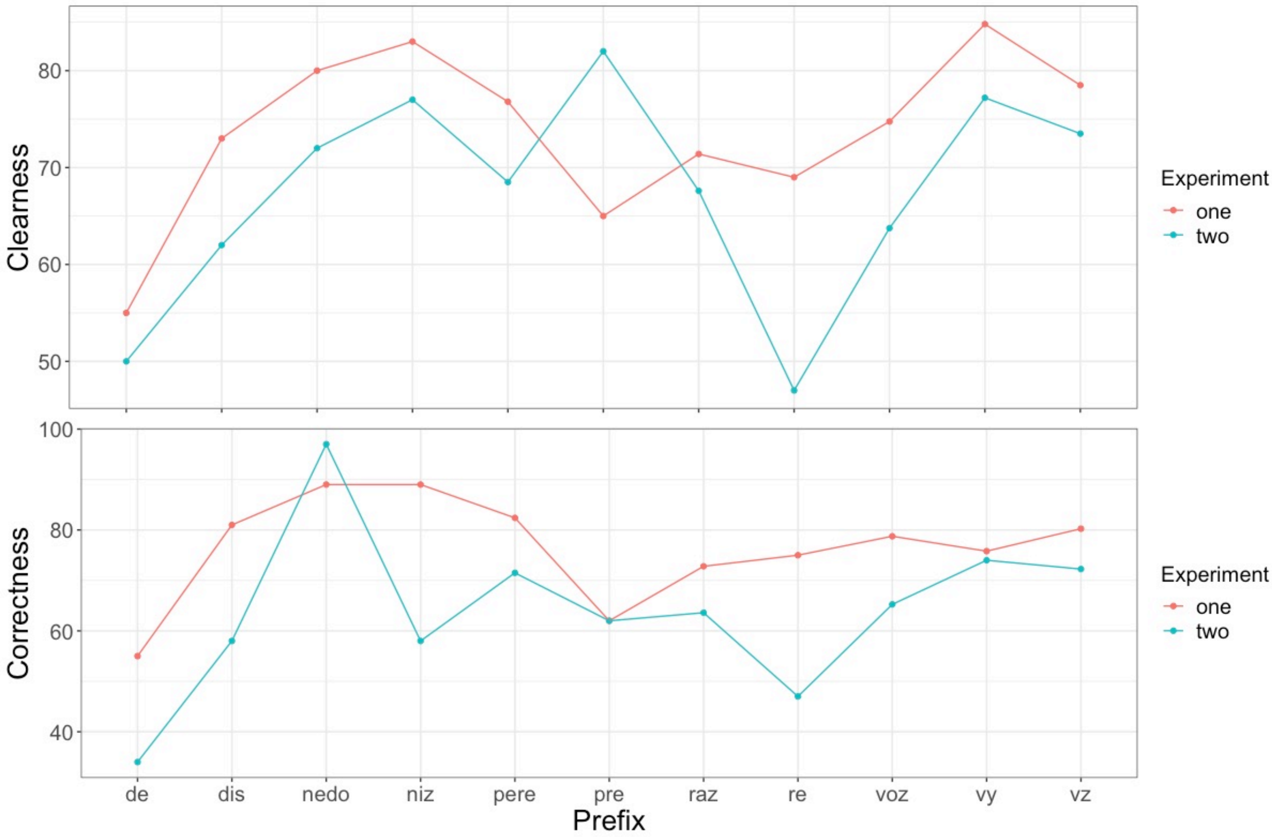

Figure 5. Interaction plot of clearness and correctness scores in experimental conditions 1 and 2 (unprepositional prefixes)

A characteristic criss-cross pattern suggests that there is an interaction. As I have observed earlier, prepositional prefixes provide, on average, higher clearness and correctness scores under the experimental condition 2, in which a priming sentence contains a verb with a different prefix, or no prefix at all, but which is synonymous in meaning to the target verb, while unprepositional prefixes, in contrast, favour experimental condition 1, in which a priming sentence contains a verb with the same constructional but different lexical meaning.

However, it is now clear that this trend does not hold for all prefixes. The difference in both clearness and correctness scores, for some of them, almost disappears (prepositional *o-* and *ot-*; unprepositional *pre-* and *vy-*) and for some, is reversed (prepositional *nad-*, *pri-*, and *so-*; unprepositional *nedo-* and *pre-*). In order to establish whether the differences between the prefixes and experimental conditions, as well as their interaction, are statistically significant, I employed a two-way (factorial) ANOVA. Although my data are quasi-interval and thus violate the interval data assumption, many studies have shown that the F-test is entirely robust to these violations and can be used to perform a statistical analysis of data collected using a Likert-type response format with no resulting bias (Glass, Peckham, and Sanders, 1972; Carifio and Perla, 2007; Boone and Boone, 2012).

The assumption of homogeneity of variance for both prepositional and unprepositional prefixes was met, as confirmed by the Levene test performed with the help of the *leveneTest()* function in the package *car* for RStudio (F = 0.63, p = 0.93 for clearness scores of prepositional prefixes; F = 0.97, p = 0.51 for clearness scores of unprepositional prefixes; F = 0.74, p = 0.51 for correctness scores of prepositional prefixes; F = 0.61, p = 0.83 for correctness scores of unprepositional prefixes). The orthogonal Helmert contrasts for the prefixes and experimental conditions variables were calculated with the help of the *contr.helmert()* function in RStudio. The results of a factorial ANOVA for the clearness and correctness scores of prepositional and unprepositional prefixes are provided in Tables 7–8.

Table 7. Results of factorial ANOVA for clearness scores

	Prepositional prefixes			Unprepositional prefixes		
	SS	df	F	SS	df	F
Intercept	835833	1	6296.75***	172006	1	1160.22***
Prefix	6978	16	3.28***	2405	10	1.62
Experiment	350	1	2.63**	357	1	2.4
Prefix:Experiment	749	16	0.35	486	10	0.32
Residuals	19646	148		6820	46	

Significance codes: ***—p < 0.001, **—p < 0.01.

Table 8. Results of factorial ANOVA for correctness scores

	Prepositional prefixes			Unprepositional prefixes		
	SS	df	F	SS	df	F
Intercept	823610	1	4643.8***	170193	1	764.03***
Prefix	12674	16	4.46***	3651	10	1.63
Experiment	754	1	4.25*	1368	1	6.14*
Prefix:Experiment	826	16	0.29	945	10	0.42
Residuals	26249	148		10247	46	

Significance codes: ***—p < 0.001, *—p < 0.05.

Since some researchers will argue that the data collected using a Likert-type response format cannot be analysed by means of a factorial ANOVA, I double-checked my findings by performing Kruskal-Wallis one-way ANOVAs by ranks for each variable and each type of scores. The results, presented in Table 9, are absolutely compatible with those obtained from two-way ANOVAs.

Table 9. Results of Kruskal-Wallis one-way ANOVAs by ranks for prepositional and unprepositional prefixes

	Prepositional prefixes		Unprepositional prefixes	
	Clearness scores	Correctness scores	Clearness scores	Correctness scores
Prefix	$\chi^2(16) = 44.66$***	$\chi^2(16) = 51.49$***	$\chi^2(10) = 16.97$	$\chi^2(10) = 13.76$
Experiment	$\chi^2(1) = 4.05$*	$\chi^2(1) = 5.47$*	$\chi^2(1) = 3.71$	$\chi^2(1) = 5.9$*

Significance codes: ***—$p < 0.001$, *—$p < 0.05$.

One can see that prepositional and unprepositional prefixes, when viewed as separate groups, display very different properties. Thus, with prepositional prefixes, both independent variables (prefix and experimental condition) have highly significant effects on both clearness and correctness scores. However, with unprepositional prefixes, neither of the independent variables affects the clearness scores and, for the correctness scores, only the change in experimental conditions evokes significant differences, while the change in prefix does not.

An important deduction which can be made from the aforementioned results is that unprepositional prefixes were considered by the participants of the experiment to be homogeneous, while at least some pairs of prepositional prefixes revealed statistically significant idiosyncratic behaviour under both experimental conditions.

2.3.4 Meaning of prefixes

To determine which prefixes differ significantly in terms of both clearness and correctness scores, I performed a post hoc Tukey Honest Significant Difference test. The results for those pairs of prefixes for which the test provided significant adjusted *p*-values and 95% confidence intervals not crossing zero are given in Table 10.

Table 10. Tukey multiple comparisons of clearness and correctness scores' means for different prepositional prefixes

Clearness scores				
Pair of prefixes	Estimate	95 % CI lower	95 % CI upper	Adjusted *p*-value
ot- — ob-	20.6	5.44	35.7	< 0.001
u- — ob-	21.4	5.91	36.9	< 0.001
u- — pro-	14.5	0.16	28.8	< 0.05
Correctness scores				
Pair of prefixes	Estimate	95 % CI lower	95 % CI upper	Adjusted *p*-value
o- — do-	-31.7	-57	-6.35	< 0.01
o- — iz-	-33.1	-55.3	-10.9	< 0.001
o- — na-	-22	-42.8	-1.22	< 0.05
ot- — o-	33.8	13.9	53.7	< 0.001
pred- — o-	39.8	11	68.5	< 0.001
s- — o-	25.8	3.56	48	< 0.01
u- — o-	23.5	3.20	43.8	< 0.01
za- — o-	22.7	3.09	42.3	< 0.01
ot- — ob-	19.8	2.31	37.2	< 0.05
pod- — ot-	-16.9	-32.6	-1.32	< 0.05
pro- — ot-	-18.6	-34.7	-2.48	< 0.01

One can observe a striking contrast between the interpretations of some prefixes. The reason for this can be uncovered through examining the arrays of meanings of the prefixes *ot-* and *ob-* since this pair is marked off by both clearness and correctness scores. Having labelled each meaning as 'prepositional' or 'non-prepositional', based on whether the corresponding preposition can or cannot be used in a paraphrase of the target verb in accordance with the procedure proposed by Bergsma et al. (2010) and modified and extended by Biskup (2015),[2] I obtained the results provided in Table 11.

[2] My 'prepositional' label corresponds to Biskup's classes 1 and 2, while 'non-prepositional' label corresponds to Biskup's classes 3 and 4.

Table 11. Prepositional and non-prepositional meanings of prefixes *ob-* and *ot-*

Prefix	Prefix meaning	Target verb	Paraphrase	Type of meaning
ob-	‘to surpass another performer of an action identified by the base verb’	*ob-igratj* ‘outplay’	—	non-prepositional
	‘to extend an action identified by the base verb to many objects (or to many places within a single space)’	*ob-ezditj* ‘go everywhere’	—	non-prepositional
	‘to direct an action identified by the base verb around an object in the path of movementj	*ob-exatj* ‘drive around something’	—	non-prepositional
	‘to perform (bring to fruition) an action identified by the base verb’	*ob-venčatj* ‘wed’	—	non-prepositional
	‘to direct an action identified by the base verb around something or towards all sides of something’	*ob-žaritj* ‘fry’	—	non-prepositional
	‘to harm someone (sometimes, cheat someone) through an action identified by the base verb’	*ob-vorovatj* ‘rob of’	—	non-prepositional
ot-	‘to perform an action identified by the base verb intensively, completely, and finally’	*ot-repetirovatj* ‘rehearse’	—	non-prepositional
	‘to separate something that was previously attached as a result of an action identified by the base verb; to annul of the result of such action’	*ot-lepitj* ‘detach’	*ot-delitj ot* ‘separate from’	prepositional
	‘to head somewhere by means of an action identified by the base verb’’	*ot-vezti* ‘drive to'	*ot-dalitj [ot etogo mesta]* ‘move to [from the deictic center]’	prepositional
	‘to perform an action identified by the base verb in response to another action’	*ot-blagodaritj* ‘give creditj	*ot-platitj [ot polučatelja]* ‘pay back [from the deictic center]’	prepositional

'to refuse or to force the refusal of something by performing an action identified by the base verb'	*ot-govoritj* 'talk outj	*ot-sovetovatj ot* 'advise againstj	prepositional
'to bring to an undesirable state (of damage, fatigue) as a result of an action identified by the base verb'	*ot-davitj* 'tread on one's footj	—	non-prepositional
'to perform (bring to fruition) an action identified by the base verb'	*ot-iskatj* 'find after some searching'	—	non-prepositional
'to remove, to separate from something by means of an action identified by the base verb'	*ot-brositj* 'throw away'	*ot-švyrnutj ot* 'hurl away from'	prepositional
'to end an action identified by the base verb that has lasted for a certain period of time'	*ot-gremetj* 'stop rumbling'	—	non-prepositional

The difference is clear. The meaning of the prefix *ob-* has undergone a long development moving away from the meaning of its corresponding preposition (six out of six meanings are non-prepositional), while the prefix *ot-* has remained fairly close to its preposition (five out of nine meanings are prepositional). Even with verbs like *ot-vezti* and *ot-blagodaritj*, for which my paraphrases may seem artificial, the directedness of action away from the deictic center is evident.

To assess whether this is truly a factor in the distribution of clearness and correctness scores, I coded all 91 meanings of prepositional prefixes as prepositional (30 meanings) or non-prepositional (61 meanings). The interaction plots for clearness and correctness scores can be found in Figure 6. Since my data were found to violate the assumption of homogeneous variance as confirmed by the Levene test, and the sample sizes are not equal, I resorted to a non-parametric ANOVA based on permutation. The summary is provided in Table 12.

Interestingly, the change in type of meaning from non-prepositional to prepositional most significantly boosted the number of correct interpretations of the nonce verbs with prepositional prefixes under experimental condition 1 with its structural priming. In other words, when a construction is activated in discourse (experimental condition 1), the difference in the helpfulness of prepositional versus non-prepositional clues is much greater than when a construction is not activated (experimental condition 2). In summary, I can confirm a significant difference in the accessibility of Russian prefix-base constructions that have prepositional and non-prepositional meanings. This is what one would intuitively expect because the former type of construction is more naturally morphologically decomposed than the latter.

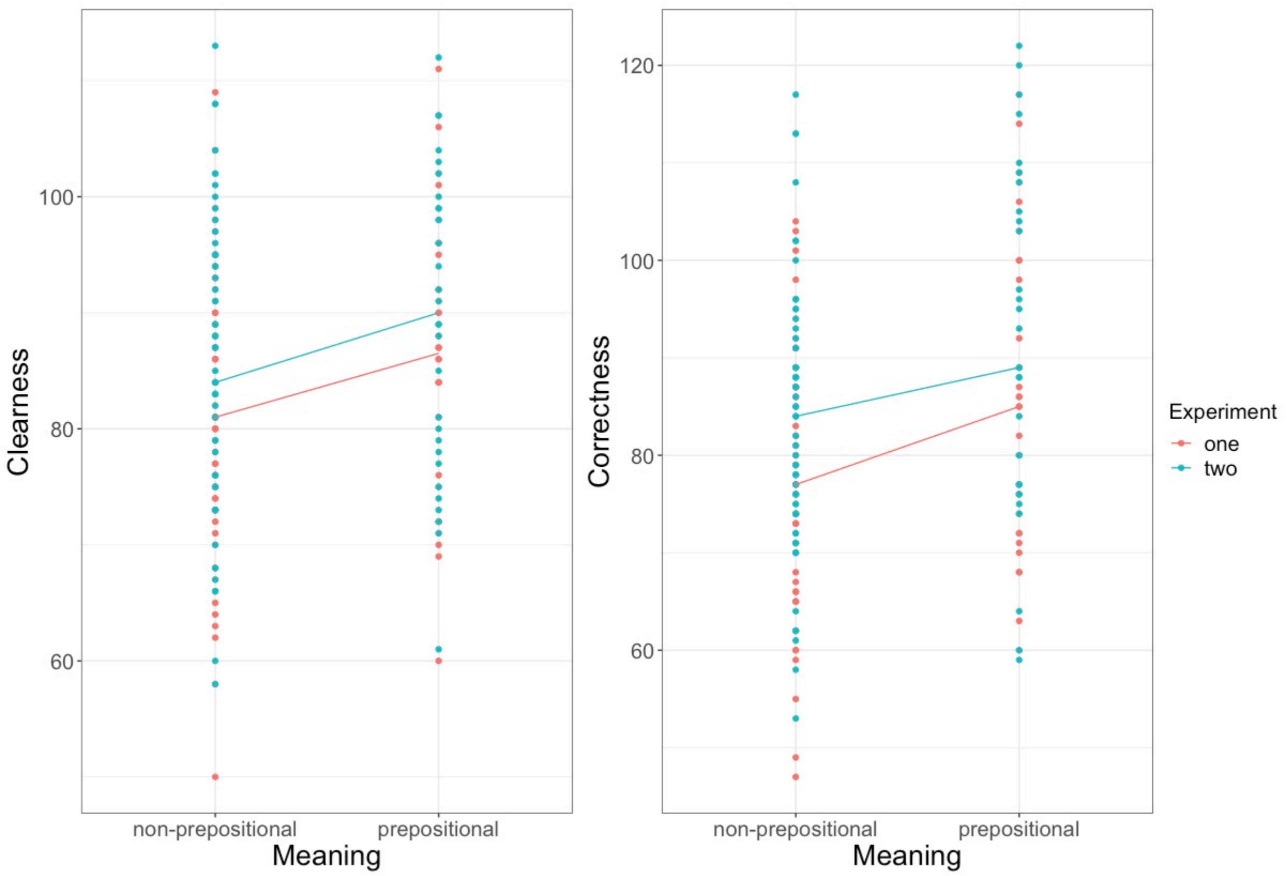

Figure 6. Interaction plot of the clearness and correctness scores in experimental conditions 1 and 2 (meanings of prepositional prefixes)

Note. Dots represent prefix-specific scores, lines connect the median scores for two types of meaning.

Table 12. Results of non-parametric ANOVA based on permutation for prepositional and non-prepositional meanings of prepositional prefixes

	Clearness scores	Correctness scores
Meaning (unprep., prep.)	Z = -2.12*	Z = -3.37***
Experiment (one, two)	Z = -2.09*	Z = -2.47*

Significance codes: ***—p < 0.001, *—p < 0.05.

2.4 Conclusion

This study has yielded a number of important results that can be summarised as follows. Speakers of Russian are very sensitive to the etymological connection between verb prefixes and the prepositions they are related to. Thus, prefix-base constructions with prefixes that correspond to prepositions are more likely to be morphologically decomposed, while the prefix-base constructions with prefixes that do not relate to prepositions tend to be regarded as a single lexical unit. Moreover, the general, highly abstract semantics of Russian prefix-base constructions, especially of those that retain their prepositional meaning, is undoubtedly accessible to language users, which is confirmed by the fact that the interpretability of these constructions is affected by priming.

All of this can be presented in the form of a hierarchy: borrowed prefixes and native prefixes unrelated to prepositions → native prefixes related to prepositions but with non-prepositional meaning → native prefixes related to prepositions and with prepositional meaning. The closer a prefix is to the left extremity of the scale, the higher the chances that the respective prefix-base construction is accessed via lexical link, that is, directly as one chunk. Conversely, the closer a prefix is to the right extremity of the scale, the higher the chances that the respective prefix-base construction is accessed via a sequential link between its morphological parts. Thus, my findings speak strongly in favour of the idea that morphological structure is gradient and shaped by language use and that morphological decomposition is a matter of degree.

3 How morphological decomposition manifests itself in the duration of the inter-morpheme period of silence in Russian prefixed verbs

3.1 Introduction

In the previous chapter, I reported the results of the experiment in which participants were asked to evaluate the semantic transparency of a prefixed nonce verb given in minimal context as well as to semanticise it by suggesting an existing Russian verb with the same prefix. These results suggest that speakers of Russian are very sensitive to the etymological connection between verb prefixes and the prepositions to which they are related (Monakhov, 2021). Thus, prefix-base constructions with prefixes that correspond to prepositions are more likely to be morphologically decomposed, while prefix-base constructions with prefixes that do not relate to prepositions tend to be regarded as single lexical units. Moreover, the general, highly abstract semantics of Russian prefix-base constructions, especially of those that retain their prepositional meanings, is undoubtedly accessible to language users, which is confirmed by the fact that the interpretability of these constructions is affected by priming.

The current chapter offers yet another way of testing the morphological gradience theory on Russian prefixed verbs. In the literature, there is abounding evidence that if linguistic expressions can be easily predicted from the context, speakers tend to reduce them, but if predictions cannot be made so easily, the tendency to produce linguistic expressions with less articulatory effort, which is inherent in automatisation, may be overridden in order to avoid misunderstandings (Lorenz and Tizón-Couto, 2017; Kuperman and Bresnan, 2012; Bybee, 2010; Bybee, 2001; Haspelmath, 2008; Pluymaekers et al., 2005; Aylett and Turk, 2004; Bell et al., 2003; Jurafsky et al., 2001).

Phonetic reduction as the phenomenon in which linguistic units are realised with relatively less acoustic-phonetic substance (Clopper and Turnbull, 2018) naturally implies reducing pausing. It is well known that frequent regular multi-word strings such as *I think that* and *I don't know* are more likely to be repeated as holistically processed chunks and that pauses frequently occur at their boundaries but not within those boundaries (Tremblay and Baayen, 2010). Given the idea of the lexicon-syntax continuum in Construction Grammar (Hoffmann and Trousdale, 2013), the same should apply to word-internal structure. For example, in a recent study, Bundgaard-Nielsen and Baker (2020) examined the effect of intra-word pausing on word acceptability in the polysynthetic language Wubuy. Wubuy listeners were presented with pairings of Wubuy words which were either unmodified or into which 500 ms of pause had been inserted at morphologically transparent legal boundaries, morphologically opaque illegal boundaries, or morpheme-internally, also illegal.

What Bundgaard-Nielsen and Baker found was that while the participants dispreferred words into which pauses have been inserted at illegal positions, they did not show a preference for unmodified words over words with legal pauses. The reported preference results were consistent with the locations of pauses in the speech sample. Here, the speaker produced much longer silent periods at morphologically transparent junctures, than she did at morphologically opaque junctures.

Bundgaard-Nielsen and Baker attributed this behaviour exclusively to the nature of wordhood in polysynthetic languages. They speculated that such a behaviour would not be attested in analytic languages, such as English, where morphologically complex '*mis*-[silence]-*place*' is likely to be dispreferred over '*misplace'*, just as morphologically simple '*ti*-[silence]-*ger*' is likely to be dispreferred over '*tiger'*, even when speakers are aware that '*mis*-' is a productive bound morpheme in English. This assumption strikes me as somewhat dubious. In any case, a more interesting comparison to make would be not that between *misplace* and *tiger* but, for example, between *misplace* and *outplace*, so that the degree of these complex words' morphological decomposition, manifested by the duration of the inter-morpheme period of silence, could be correlated with the relative frequency of their component parts.

Without any relevant English data at hand, I cannot pursue this line of argumentation any further. I contend, however, that Russian, though not a polysynthetic language, exhibits with regard to the intra-word pausing much Wubuy-like behaviour. In what follows I present evidence that in Russian pronunciation, there tends to be a longer silent period between prepositional prefixes and bases (morphologically transparent junctures) than between unprepositional prefixes and bases (morphologically opaque junctures). The reason for that is shown to be grounded in the compositional nature of the former and non-compositional nature of the latter.

3.2 Experimental design, data, and methods

Russian Grammar, created by the Russian Academy of Sciences in 1980, lists 28 verbal prefixes (Shvedova, 1980). These are comprised of the following:

- 17 prefixes are not only historically related to prepositions, but also have prepositional counterparts in modern Russian: *v-* (*v* ‘in, atj), *do-* (*do* ‘to, before’), *za-* (*za* ‘for, behind’), *iz-* (*iz* ‘from, out of’), *na-* (*na* ‘on’), *nad-* (*nad* ‘over, above’), *o-* (*o* ‘aboutj), *ob-* (*ob* ‘aboutj), *ot-* (*ot* ‘from’), *po-* (*po* ‘along, by’), *pod-* (*pod* ‘under’), *pred-* (*pered* / *pred* ‘before, in front of’), *pri-* (*pri* ‘by, atj), *pro-* (*pro* ‘about, of’), *s-* (*s* ‘with’), *so-* (*so* ‘with’), and *u-* (*u* ‘from, by’);
- 11 prefixes have no prepositional counterparts in modern Russian; this group encompasses morphemic borrowings, prefixes that have unprepositional origin and prefixes derived from prepositions that are no longer part of the Russian language: *de-*, *dis-*, *vz-*, *voz-*, *vy-*, *nedo-*, *niz-*, *pere-*, *pre-*, *raz-*, and *re-*.

Almost all Russian verbal prefixes, both prepositional and unprepositional, are polysemous with the number of meanings ranging from 2 (for example, *v-*) to 10 (for example, *pere-*). For the experiment, all the meanings of all the prefix-base constructions discussed by Russian Grammar were taken into consideration (91 meanings for prepositional prefixes and 34 meanings for unprepositional prefixes; 125 in total). For each meaning, one verb was randomly chosen from the list of paradigm examples with which this meaning is illustrated in Russian Grammar and one sentence with this verb used in this meaning was randomly obtained from the Russian National Corpus. I did not want to construct any sentential templates to make sure that my subjects remain unaware of the true purpose of the study and do not try to pronounce the relevant verbs in a somewhat affected manner. However, all the sentences were constrained to be of approximately the same length.

To conduct the experiment, I used Yandex Toloka, a Russian crowdsourcing service analogous to Amazon Mechanical Turk. First, I created a special template so that each task included one of the 125 sentences as input data. As output data, an audio file containing the record of the specific sentence being pronounced by a participant was requested. For each task, a time limit of 10 minutes was imposed. After completing the tasks, all participants were rewarded in the amount of $0.04 USD for each accepted submission.

Second, I assembled two pools of users who met the following criteria: 1) being a native speaker of Russian and 2) being in the top-rated 10% of all active users on the platform. Participants of each pool were assigned to one of two groups of tasks: 1) pronouncing sentences that included verbs with prepositional prefixes or 2) pronouncing sentences that included verbs with unprepositional prefixes. Each sentence had to be recorded by 30 different users and no user could see any tasks other than those assigned to their pool. The number of sentences one user could pronounce was limited: those having recorded as many as five sentences were automatically discarded from the project.

The instructions for the participants of the experiment were written so as not to reveal my research hypothesis. The full text translated from Russian is provided below.

> Hello! Thank you for agreeing to participate in my experiment. The experiment does not require any special knowledge; the only requirement is to be a native Russian speaker. I ask you to read short Russian sentences aloud at your usual pace and record your reading.
> Please bear in mind:
> 1 Audio recordings that do not contain the specified sentence, contain only pieces of it, or contain any extraneous words not present in it will not be accepted.
> 2. Start talking after you have pressed the button on the recorder so that the beginning of the sentence is not lost. Press the end button after you have finished speaking so that the end of the sentence is not lost.
> 3. It is not necessary to record in complete silence, but the level of external noise must be kept to a minimum.

The contexts were randomly shuffled so that different meanings of the same prefix did not follow each other. Given the abundance of sentences, I decided not to add any filler contexts.

The null hypothesis H_0 of the experiment was that there would be no significant difference in the duration of the periods of silence between verbal prefixes and bases across the two groups of tasks. The alternative hypothesis H_1 was that there would be a longer period of silence between prepositional prefixes and bases than between unprepositional prefixes and bases.

Overall, 3,696 recorded sentences were obtained: 125 unique prefix-base construction each recorded by 30 native speakers of Russian, minus 54 submissions that were excluded due to errors (empty files, high levels of noise, omission of target verbs). In total, 883 people took part in the experiment with, on average, four unique sentences being recorded by each speaker.

The acoustic waveforms of the target verbs were hand-segmented in Praat (Boersma & Weenink, 2020). Visually identifiable periods of silence at the boundaries between verbal prefixes and bases were manually coded (see Figure 7 for illustration) by two annotators. For each prefixed verb, two values were extracted: A) the total duration of the pronunciation of the given verb in milliseconds and B) the duration of the silent period between the verbal prefix and the base in milliseconds. As a measure of interest, the simple ratio of B to A was calculated in order to control for varying speech rates (cf. Matzinger, Ritt, and Fitch, 2020).

The intraclass correlation coefficient was computed to assess the agreement between two annotators in measuring the duration of the silent periods. The ICC for ‘single fixed raters’ was found to be equal 0.80 ($p < 0.0001$, CI 95% 0.80—0.82), indicating a good agreement (Koo, Terry, and Mae Li, 2016). From the Bland and Altman plot below (Figure 8), it can be seen that the mean difference between two methods of annotating is equal to -1.60 ms (CI 95% -29.32—26.13)), and the differences are normally distributed with no discernible pattern in the data. Since only two annotators were employed, I decided not to average the results but rather use one set of measurements, obtained from the first annotator.

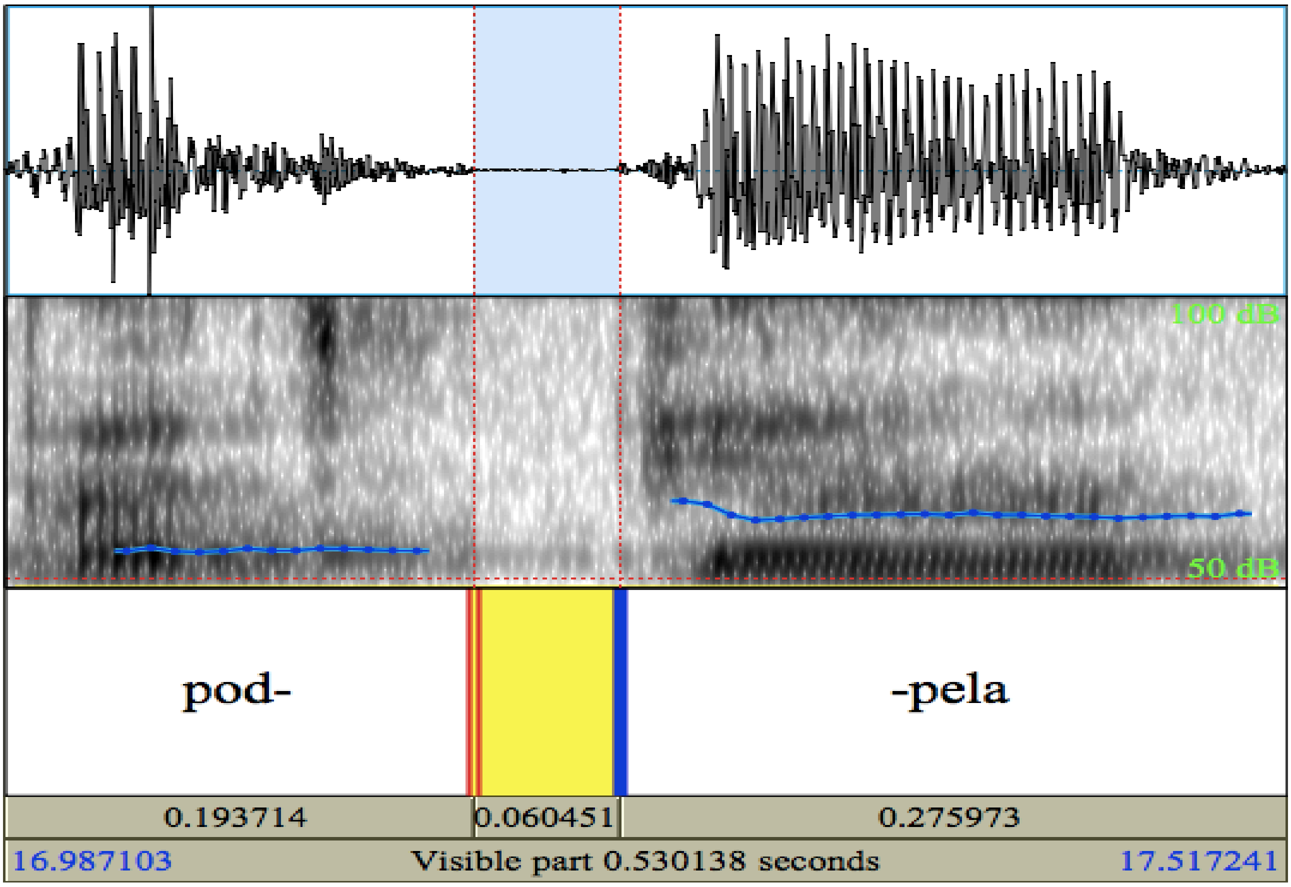

Figure 7. Waveform of the Russian prefixed verb *pod-pela* (‘<she> sang along’) containing a period of silence of 60.4 ms between the prefix and the base

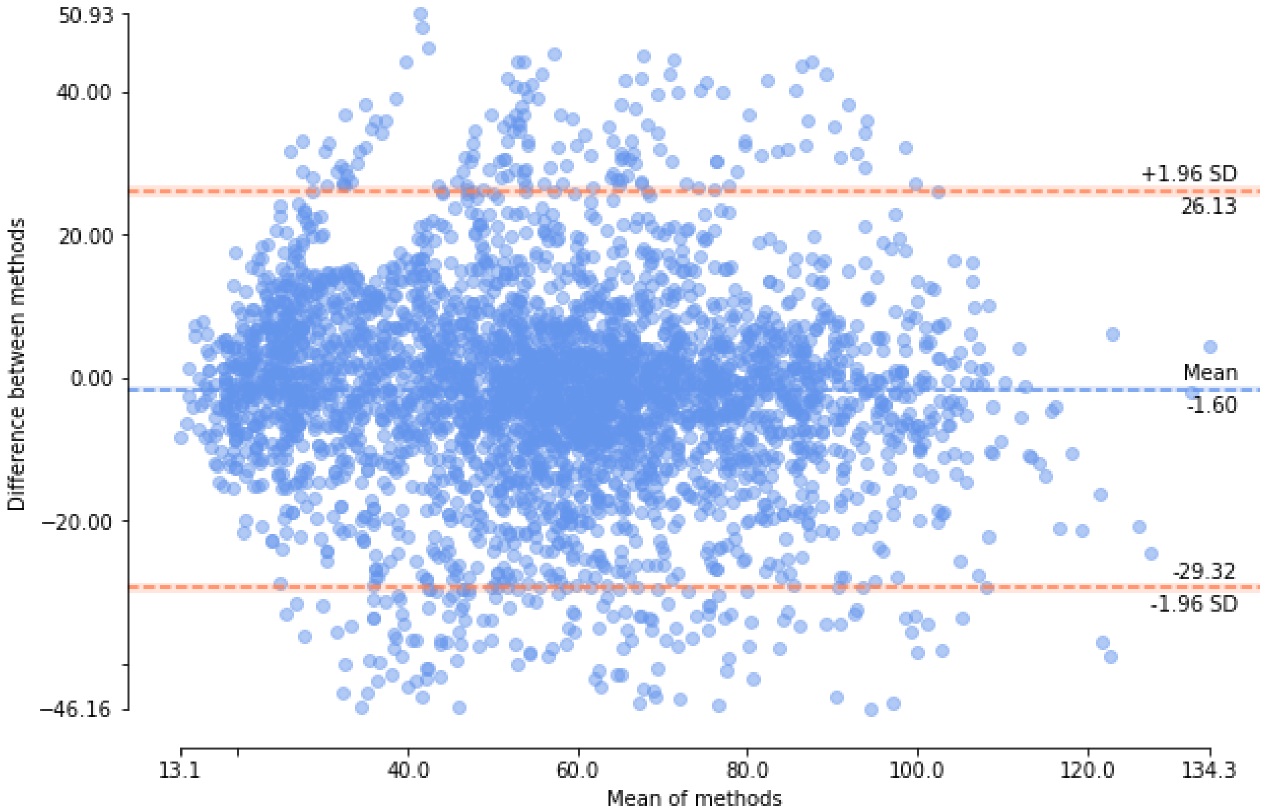

Figure 8. Bland and Altman plot of differences between two annotating methods

3.3 Results and discussion

The resulting distributions of the pause ratios for the two types of prefixes are plotted in Figure 9. As it is evident from this figure, my null hypothesis H_0 that there would be no significant difference in the duration of silent periods between verbal prefixes and bases across two types of prefixes can be safely rejected. The prefixes related to prepositions produced, on average, significantly greater ratios (M = 0.11, SE = 0.0007) than the prefixes unrelated to prepositions (M = 0.07, SE = 0.001), t = 25.5, $p < 0.001$. The effect size was large: $r = 0.52$ (Field, 2012: 58).

One possible contradiction about the observed difference being influenced by the fact that verbs with unprepositional prefixes tend to be longer than their counterparts (mostly due to the borrowed lexemes with the prefixes of Latin origin *re-*, *dis-*, and *de-*) is resolved by directly comparing the durations of pauses. The prefixes related to prepositions produced, on average, significantly greater silent periods (M = 58.54, SE = 0.40) than the prefixes unrelated to prepositions (M = 48.08, SE = 0.71), t = 12.7, p < 0.001. In addition, there is no significant correlation between durations of pauses and total durations in my data ($r = 0.02$, $p = 0.12$), so I will stick with relative rather than absolute values for the reason mentioned above.

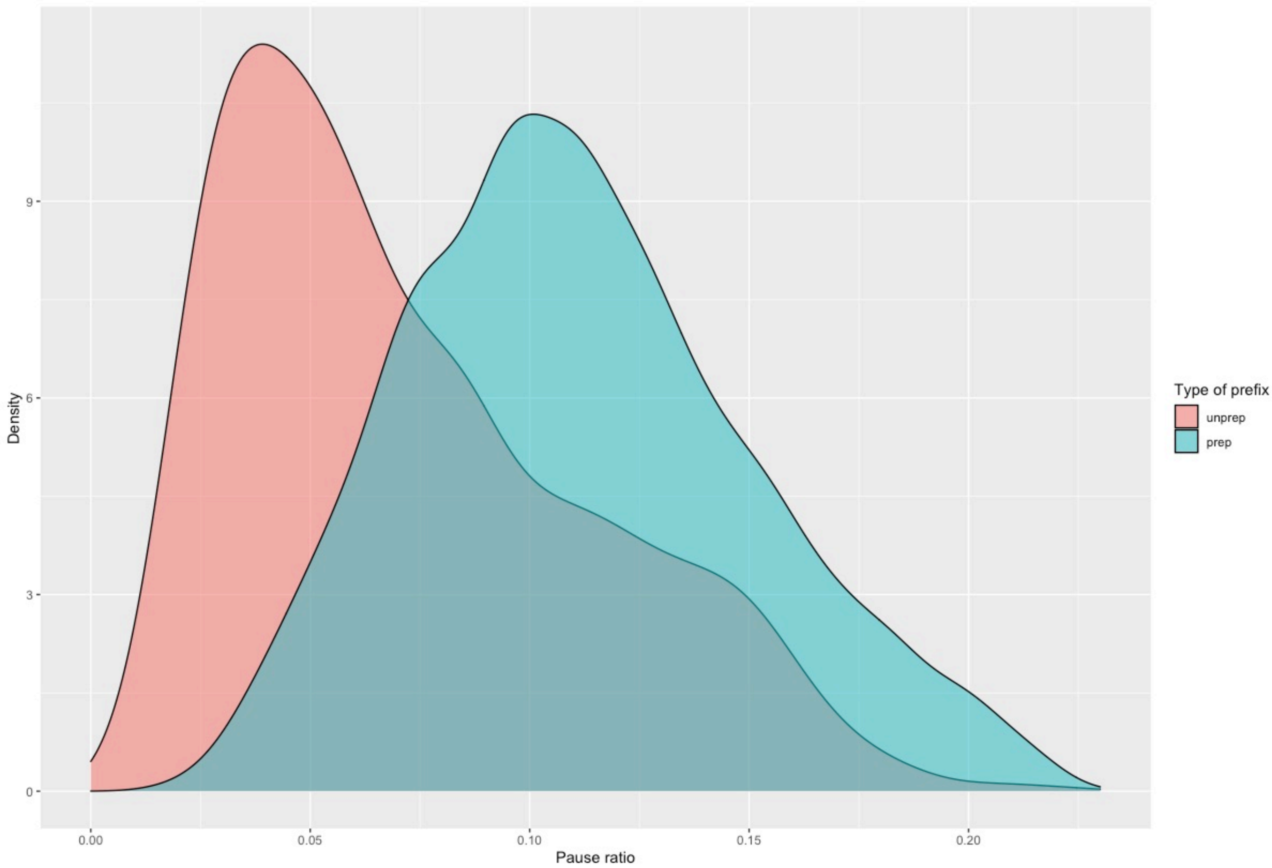

Figure 9. Distributions of pause ratios for verbs with unprepositional and prepositional prefixes

3.3.1 The mystery of the third population

If one looks at the distributions of pause ratios for different prefixes without grouping them into any categories, it becomes evident that the real picture is much more complicated than the initially proposed dichotomy. There is a lot of variation, and while two clusters of unprepositional and prepositional prefixes are clearly detectable, with the centers of distributions located around the two mean pause ratios that I identified, there seems to be some indication of a hidden third population underlying the data (Figure 10). Most notably, a group of prefixes, namely *pere*-, *raz*-, and *vy*- (unrelated to prepositions) and *na*-, *ob*-, and *pred*- (related to prepositions), appear to form a camp of their own lying in the area between the unprepositional and prepositional strongholds.

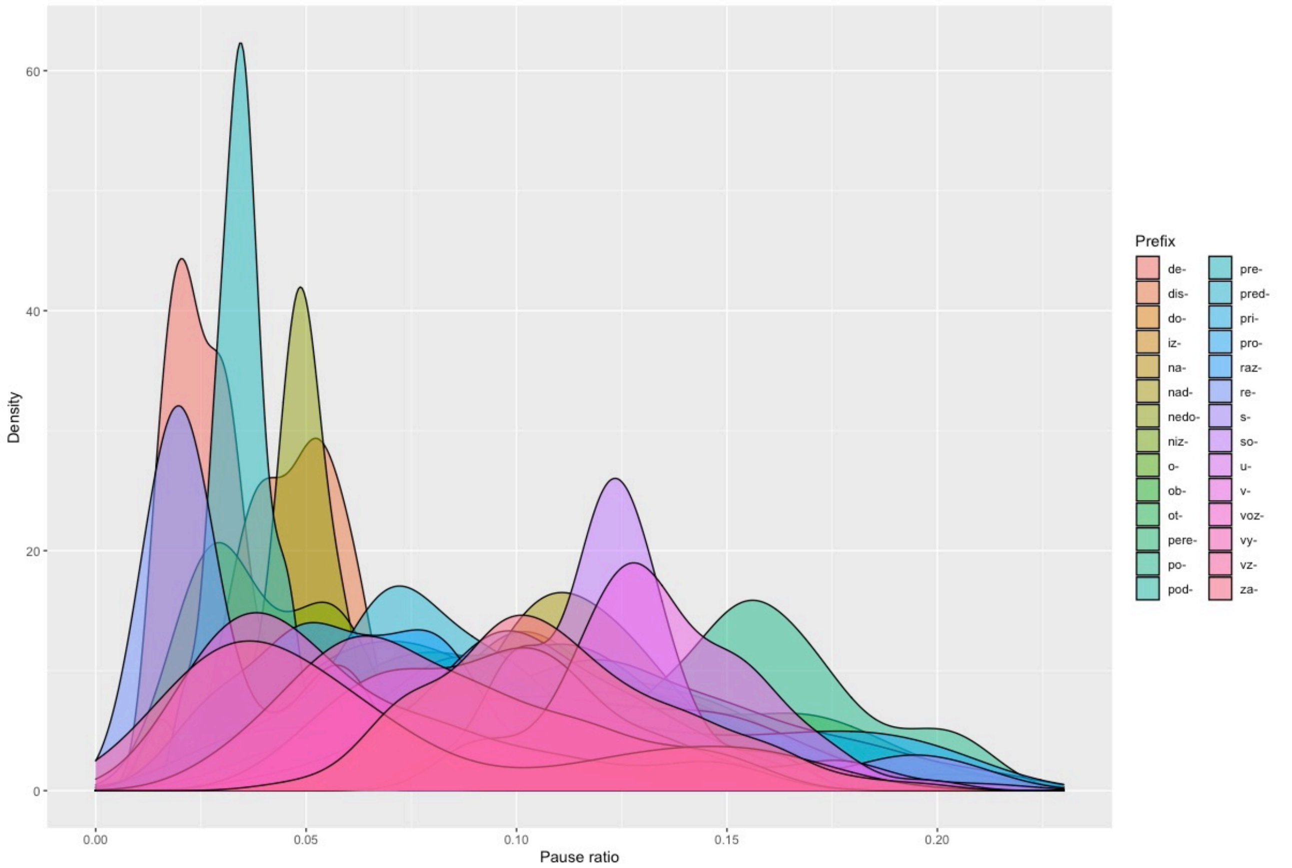

Figure 10. Distributions of pause ratios for constructions with different prefixes

It might be the case that there are three populations of pause ratios instead of two, but I do not know exactly which population each observation belongs to. These latent variables can be treated as parameters in a hierarchical model and identified with the help of Bayesian inference and the Markov chain Monte Carlo algorithm. To obtain posterior distributions for p (population membership of the observations) and θ (population-specific parameter, i.e., its mean), I used a mixture of three normal distributions, each with $\mu = 0$ and $\sigma = 15$. Since my prior information about the locations of the populations' means and their standard deviations was very limited, it was reasonable to associate with parameter θ a weakly informative prior allowing for a great degree of variability, so that the real data would dominate the posterior distribution and overwhelm the prior. The means were ordered so that, in posterior distribution, the first mean was less than the second and the second was less than the third. The model was instructed to use for p a Dirichlet prior parameterised by a vector $\alpha = [1, 1, 1]$ so that all memberships were initially considered equally likely. Two chains of 3,000 candidate samples each were drawn after the initial 5,000 draws were discarded as a burn-in period that the model needed to reach the stationary distribution.

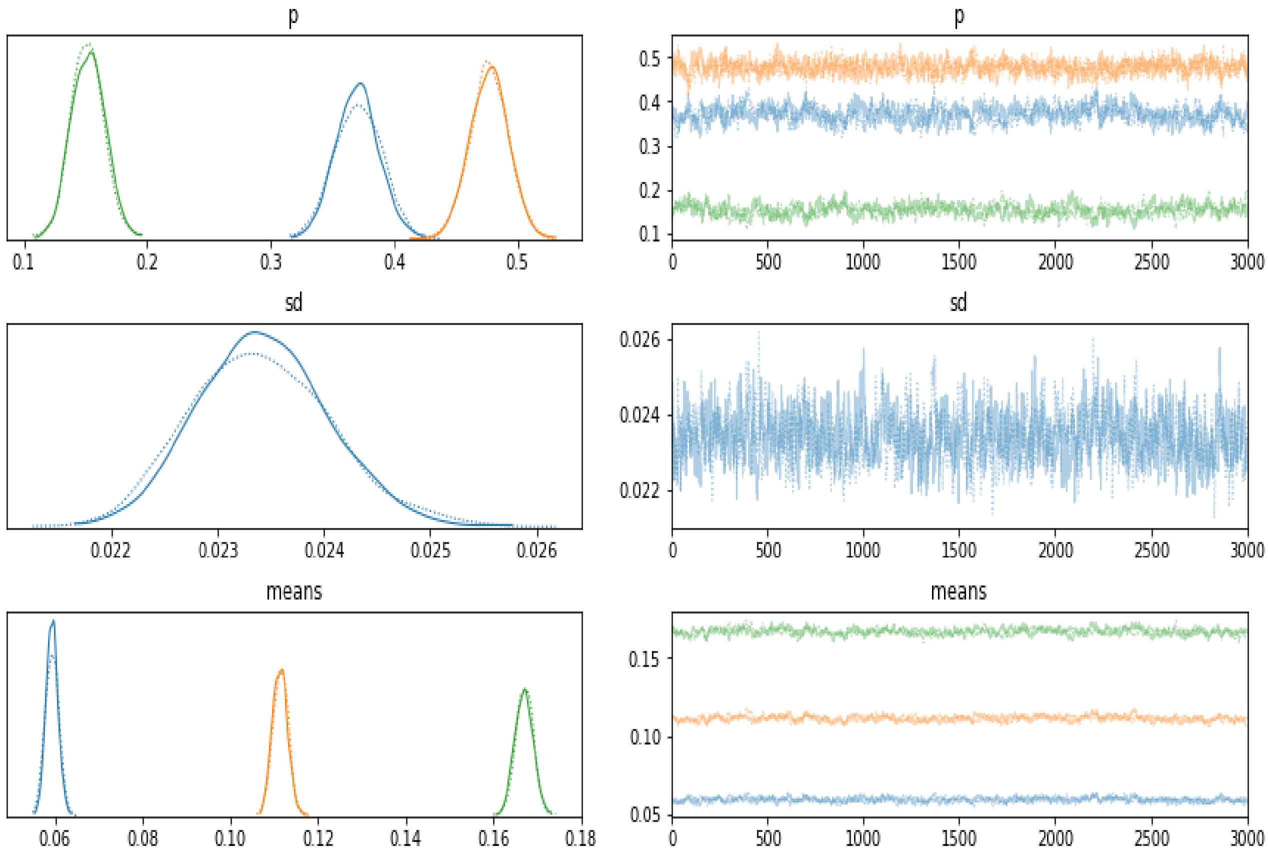

Figure 11. Gaussian mixture model of the pause ratios: posterior distributions and trace plots

As can be read from the trace plots (Figure 11), the algorithm converged successfully, having effectively identified three underlying populations with the actual observations distributed between them in the following proportion: 37% of the data came from the first population, with a mean equal to 0.06; 47% of the data came from the second population, with a mean equal to 0.11; and 16% of the data came from the third population with a mean equal to 0.17. Surprisingly, the parameter θ of the third population turned out to be very different from what I had expected: instead of lying between the centers of the distributions of unprepositional and prepositional prefixes, it occupied the right-most position on the x-axis. This means that some relatively small group of prefixed verbs in my sample is characterized by a very large pause ratio.

Bayesian statistical modelling allows me to directly calculate the probabilities of each specific observation coming from the particular population. The probabilities for four observations, chosen for the sake of example, are provided in Figure 12. Using these probabilities, I can estimate the overall, as well as prefix-wise, accuracy of the model in attributing the observations to one of the three populations.

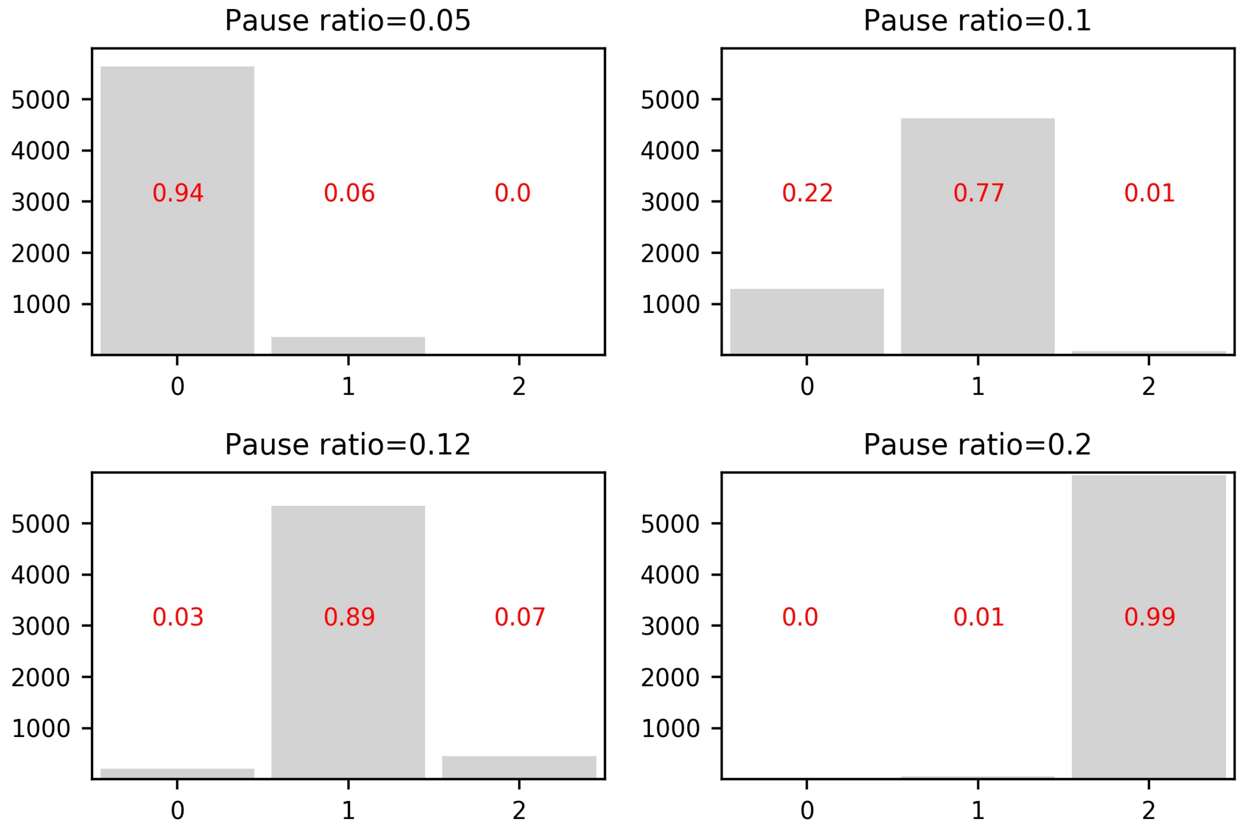

Figure 12. Probabilities of specific observations coming from three different populations of pause ratios

My best guess, given all the information available at the moment, will be that the first population (the one with the mean of 0.06) will comprise constructions with unprepositional prefixes and the second population (the one with the mean of 0.11) will comprise constructions with prepositional prefixes to the exclusion of prefix *po-*, which for now I will consider the only representative of the third population (the one with the mean of 0.17) due to its larger than prepositional average pause ratio (see Figure 10).

The overall accuracy of the model calculated as the simple percentage of correct predictions among all predictions is 61%. It is noteworthy that the prefixes show substantial variability with regard to this score, ranging from 100% (*dis-*, *de-*, *pre-*) to 35% (*pere-*, *pred-*) (Figure 13). Not surprisingly at all, among the prefixes occupying the lowest positions in this accuracy hierarchy are the same prefixes I have earlier identified as problematic due to their intermediate location between the unprepositional and prepositional camps, namely *pere-*, *pred-*, *vy-*, *na-*, and *ob-*.

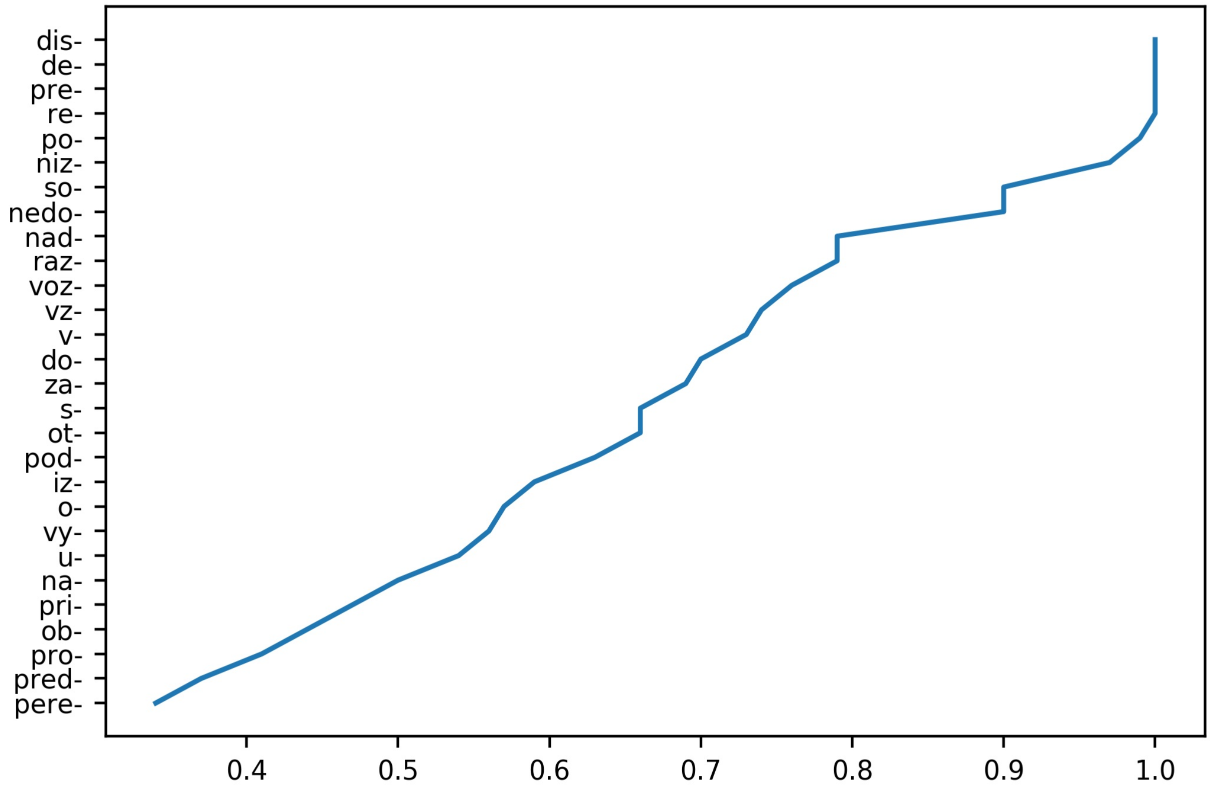

Figure 13. Accuracy of predicting the population of a pause ratio averaged across constructions with different prefixes

One main conclusion that can be drawn so far is that the dividing line between the three identified populations of pause ratios does not clearly separate types of prefixes or single prefixes but rather cuts across them, demarcating several groupings of prefix-base constructions that share some common features in their functioning and meanings.

3.3.2 Prefix-base constructions' meanings

A simple dichotomous division of the Russian verbal prefixes that I have adopted initially does not take into account that the 'prepositonality' of a prefix-base construction is a matter of degree. In fact, the rich variety of these constructions as form-meaning pairings can be subdivided two-dimensionally.

The first dimension deals with the level of compositionality/idiomaticity. It has been observed that different combinations of verbal prefixes and bases attested in Russian form a cline stretching between the poles of (1) total compositionality, where the meaning of the whole unit is composed of the meaning of the prefix plus that of the base, and (2) total idiomaticity, where the meaning of the construction is completely unpredictable from the meanings of its parts.

(1) *za-brositj*
behind-throw
'to throw something behind something, far away'

(2) *ot-rjaditj*
from-rule
'to send someone somewhere to fulfil an assignment'

The usual way to assess whether a prefixed verb has a compositional or non-compositional meaning is by checking whether the prefix-related preposition can or cannot be used in a paraphrase alongside a verb with the same base (Biskup, 2015; Bergsma et al., 2010). For example, it is possible with (1): *Kuda on **za**-brosil mjach?* 'Where did he throw the ball?' → *On brosil mjach **za** dom* 'He threw the ball behind the house', but not with (2). Of course, this test cannot tell us anything about prefixes that are unrelated to prepositions since there is simply no corresponding preposition with which to form a paraphrase.

Thus, three groups of prefix-base constructions emerge: prepositional (constructions with compositional meaning), non-prepositional (constructions with non-compositional meaning), and unprepositional (constructions with prefixes unrelated to prepositions). In other words, of the two types initially proposed, constructions with prefixes that do not have analogous free lexemes in modern Russian still form one group, but constructions with prefixes that do have such counterparts are further subdivided into two types—prepositional and non-prepositional—based on the results of a Bergsma-Biskup test.

In the second, semantic dimension, the continuum of the prefix-base constructions can be divided into eight groups, seven of which are analogous to the Aristotelian ontological categories (Jansen, 2007). The eighth is an additional category of *limit* that I employed to designate the cases where a prefix does not bear any particular meaning but simply turns the grammatical aspect of the verb from imperfective into perfective (Arkadiev and Shluinsky, 2015). The overview of the categories is provided in Table 13.

Table 13. Ontological categories of prefix-base constructions

group of prefix-base constructions	Aristotelian category	Russian example	construction meaning	English translation
Direction	the where	*iz-gnatj*	‘to remove something from somewhere’	to exile from
Place	the where	*u-mestitj*	‘to make something fit in somewhere’	to place in
Time	the when	*po-kuritj*	‘to perform an action within a certain period of time’	to smoke for a while
Relation	that which is related to something	*ot-blagodaritj*	‘to perform an action in response to another action’	to pay back
State/Posture	the positioning	*u-kachatj*	‘to bring someone or something to an undesirable state (extreme fatigue, powerlessness, exhaustion)’	to make seasick
Quality/Manner	the how constituted	*nedo-otsenitj*	‘to perform an action incompletely, fail to achieve the necessary standard’	to underestimate
Quantity	the how much	*o-prositj*	‘to extend an action to many objects (or to many places within a single object)’	to poll
Limit	—	*na-smeshitj*	‘to perform an action, bring it to fruition’	to make someone laugh

It is clear that this variability can be reduced even further by merging categories that encode spatial, topographic relations (*direction*, *place*), on the one hand, and categories that can be derived from the former via metonymical, metaphorical, or other semantical links (*relation*, *state/posture*, *quality/manner*, *quantity*; see, for example, Table 14), on the other hand (Šarić, 2003; Janda, 1986).

Table 14. Hypothetical routes of the development of meanings encoded by prefix *na-*

literal meaning	extension by metonymy	extension by metaphor	extension by contrast
'to direct an action to a surface of something; place something on the surface, bump into something' (*na-kleitj* 'to glue on')	'to accumulate in a certain amount by means of a surface-oriented action' (*na-soritj* 'to litter on')		
	'to accumulate in a certain amount' (*na-lovitj* 'to catch a lot')	'to perform an action intensively' (*na-bezobraznichatj* 'to mess up')	'to perform an action in a gentle, unobtrusive manner' (*na-igratj* 'to play music gently')
		'to teach someone something' (*na-mushtrovatj* 'to train, prime')	

The category of *time* most naturally belongs to the first group, since time is known to be conceptualised across cultures in terms of space (Núñez and Cooperrider, 2013). Of course, this conceptualisation also presupposes metaphorical extension, but what sets the category of time aside from the others is the fact that this extension is (almost) universal: all Russian prepositional prefix-base constructions with temporal meanings relate to prepositions that encode the same meanings, while the meanings within my second group of categories are predominantly construction-specific. The only exception to this rule is constituted by the prefix *pro-*, whose constructional meaning 'to perform an action for some time (often for a long time)' is evidently derived from another meaning of the same construction: 'to direct an action through something inward'.[3]

[3] Given the potential controversiality of this decision, I ran a series of tests parallel to the ones described below but with the category of *time* being included in the second group of categories. It did not affect the nature or the significance of the results, though it made the observed distinctions less extreme. This is why I will not report these parallel results here.

By applying the aforementioned transformations, I get three groups of categories that will further on be referred to as literal (*direction*, *place*, *time*), metaphorical (*relation*, *state*/*posture*, *quality*/*manner*, *quantity*), and conventional (*limit*). I use the word 'literal' instead of 'prepositional' for two different reasons. First, I do this to avoid overlap with another dimension of categorising that was discussed above. Second, it would be somewhat awkward to talk about the 'prepositional' meaning of prefix-base constructions where prefixes do not have their prepositional counterparts. Of course, it may be argued that in such cases, spatial and temporal categories do not necessarily lie in the center of the constructional network of meanings. Discussing this in detail is really beyond the scope of the present study, but it can be shown that for most native unprepositional prefixes, the coding of topographic relations is both the most prototypical and the earliest from a historical point of view.

This tripartite semantic division is well-established in the literature with regard to the verb-particle constructions of several Indo-European languages (Dehé et al., 2002; Iacobini and Masini, 2007). It usually takes the following form: (1) locative meanings, (2) idiomatic meanings, and (3) aspectual meanings. As can be seen, the only change I propose is that of including the category of *time* into group (1) for the reasons described above.

The resulting matrix filled with some examples is provided in Table 15. Each of the 125 prefix-base constructions in my data was assigned to the semantic category whose inherent idea it conveys and to the compositional type that it represents.

Table 15. Matrix of cross-referenced semantic categories and compositional types of prefix-base constructions

category / type	prepositional	non-prepositional	unprepositional
literal	*pod-plytj* 'to swim under'	*o-bezhatj* 'to run around'	*niz-vergnutj* 'to bring down, overthrow'
metaphorical	*ot-govoritj* 'to talk out of'	*pro-dumatj* 'to think carefully'	*raz-morozitj* 'to defrost'
conventional	*do-chitatj* 'to finish reading'	*ob-venchatj* 'to pronounce married'	*voz-muzhatj* 'to mature'

3.3.3 Factoring in categories and types

The locations of pause ratios for different categories and types of prefix-base constructions are presented in Figure 14. Two linear hierarchies can be distinguished: one for types (unprepositional [M = 0.06] > non-prepositional [M = 0.10] > prepositional [M = 0.11]) and one for categories (conventional [M = 0.08] > metaphorical [M = 0.09] > literal [M = 0.10]). All these differences in locations are significant, as confirmed by the approximative k-sample Fisher-Pitman permutation tests ($\chi^2(2) = 581.71$, $p < 0.0001$ for types and $\chi^2(2) = 42.778$, $p < 0.0001$ for categories). However, most interesting are interactions between categories and types.

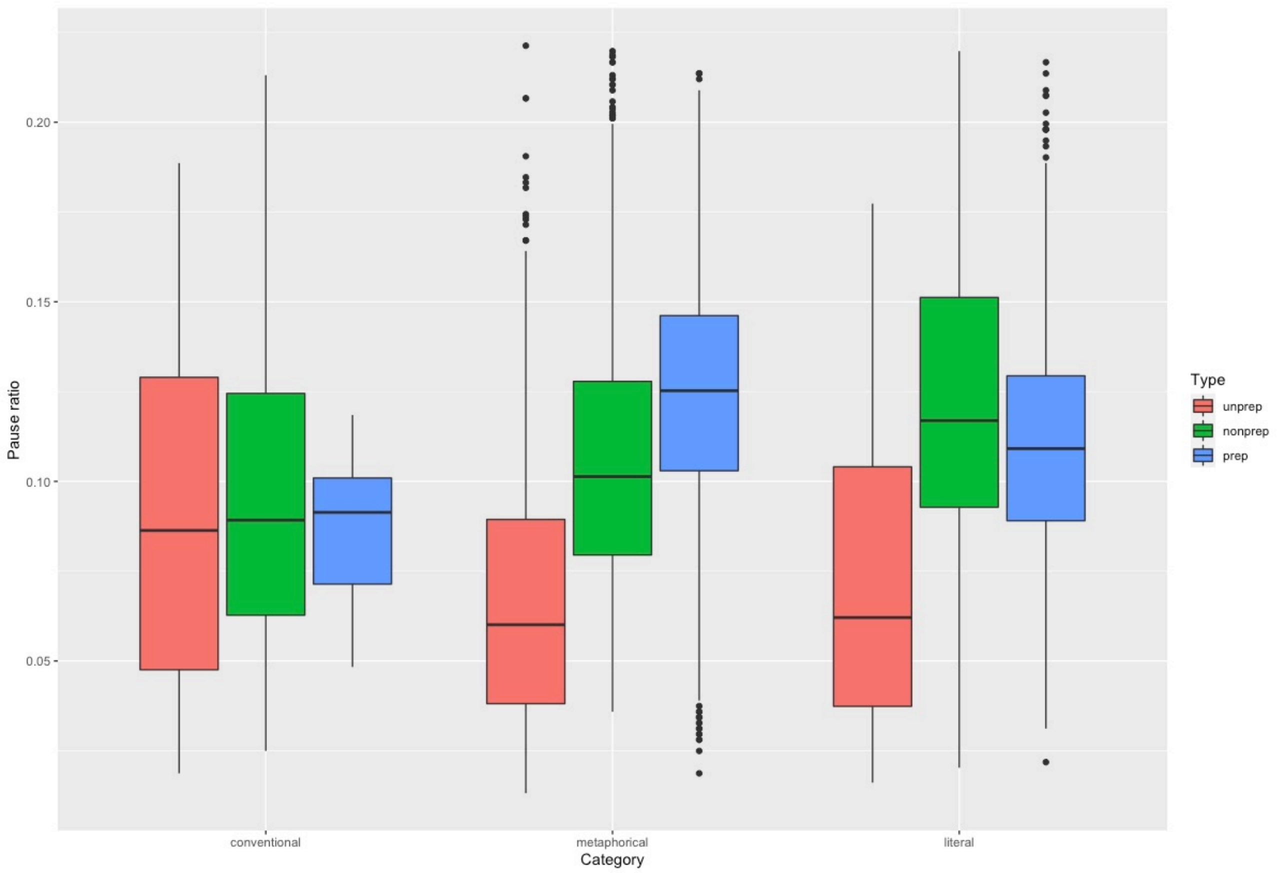

Figure 14. Locations of pause ratios for different categories and types of prefix-base constructions

To analyse these interactions, I fitted a linear regression model (Model 1) to the data with pause ratio as the response and semantic category and compositional type as independent factors. Since my data violate the assumption of homoscedasticity, I used bootstrapping that involved 5,000 resamplings to obtain confidence intervals for the constant, main effects, and interaction terms. The original statistics lie within the confidence intervals based on the bootstrapping; therefore, the results of the model can be considered reliable. A summary is provided in Table 16.

Table 16. Coefficients of Model 1

	B (SE)	95 % CI (based on 5,000 resamplings)	
		Lower	Upper
Constant	0.08*** (0.003)	0.08	0.09
category: metaphorical	-0.01*** (0.003)	-0.02	-0.01
category: literal	-0.01*** (0.004)	-0.02	-0.06
type: non-prepositional	0.006 (0.004)	-0.002	0.01
type: prepositional	-0.001 (0.008)	-0.01	0.009
metaphorical*non-prepositional	0.03*** (0.004)	0.02	0.04
literal*prepositional	0.03*** (0.008)	0.03	0.05
literal*non-prepositional	0.04*** (0.004)	0.04	0.06
metaphorical*prepositional	0.05*** (0.008)	0.02	0.05

Significance codes: *** $p < 0.001$. Notes: Adjusted $R^2 = 0.19$, $p < 0.0001$.

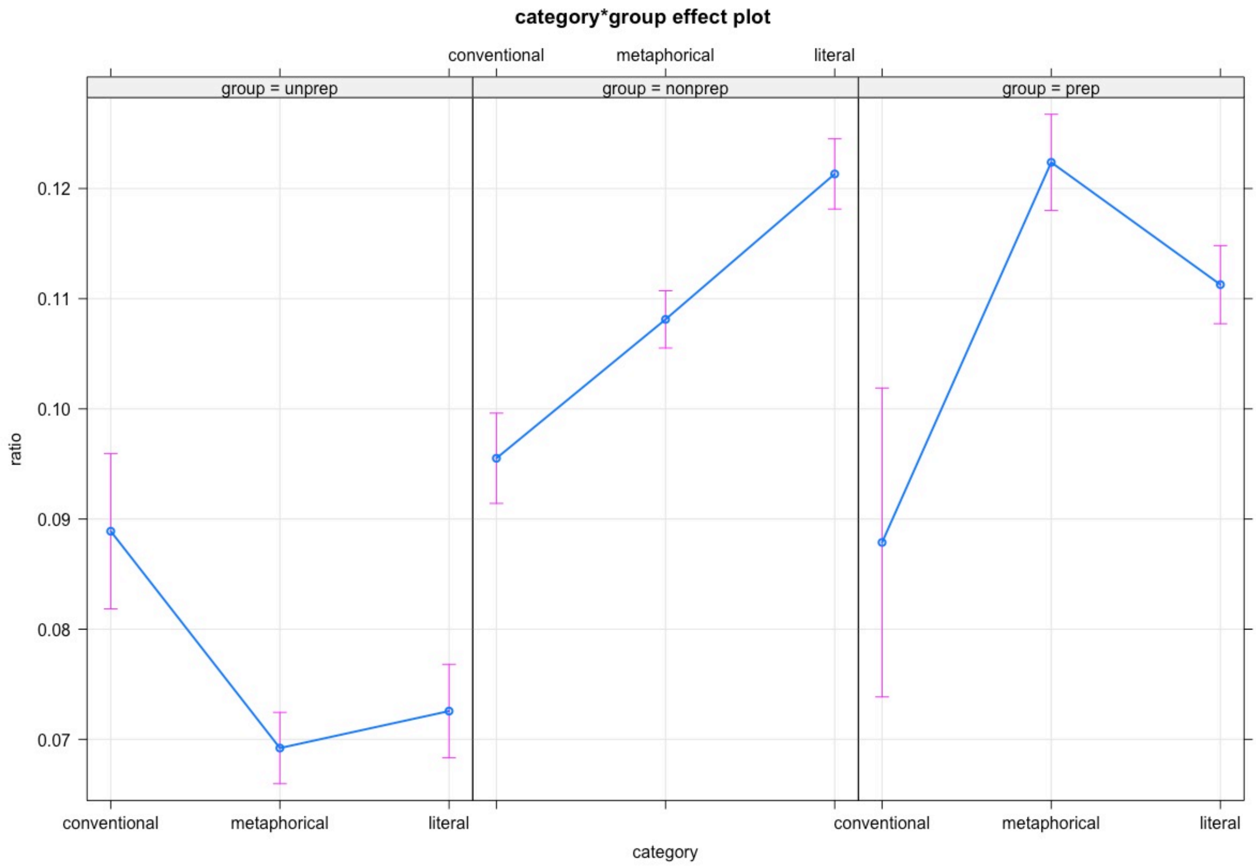

Figure 15. Main effects plot of Model 1

To make sense of the coefficients, it is better to look at the main effects plot of Model 1 (Figure 15). An interesting pattern can be observed. With different compositional types of prefix-base constructions, different semantic categories are characterized by the largest pause ratio: unprepositional constructions give preference to conventional meanings, non-prepositional constructions give preference to literal meanings, and prepositional constructions give preference to metaphorical meanings. Moreover, the locations of the pause ratios seem to align across different categories and types in accordance with the following pattern: metaphorical and literal meanings for unprepositional constructions > conventional meanings for all types of constructions > metaphorical meanings for non-prepositional types and literal meanings for prepositional types > literal meanings for non-prepositional types and metaphorical meanings for prepositional types.

While fitting Model 1 to the data, I did not take into consideration any phonetic factors. Meanwhile, it is highly probable that some amount of the pause ratios' total variance can be explained by certain segmental and suprasegmental features of particular prefixed verbs. To control for these factors, I fitted to the data another model, Model 2, where alongside the semantic category and compositional type, the following predictor variables were used: 1) the presence of vowel in the prefix (yes or no), 2) the type of the prefix's final phoneme (fricative, stop, vowel), 3) the type of the base's initial phoneme (affricate, approximant, fricative, nasal, stop, trill, vowel), and 4) the index of the stressed syllable. A summary of Model 2 is provided in Table 17. The main effects plot can be found in Figure 16.

Table 17. Coefficients of Model 2

	B (SE)	95 % CI (based on 5,000 resamplings)	
		Lower	Upper
Constant	0.08*** (0.005)	0.07	0.09
category: metaphorical	-0.01*** (0.004)	-0.02	-0.008
category: literal	-0.006 (0.004)	-0.01	0.003
type: non-prepositional	0.009* (0.004)	0.0005	0.01
type: prepositional	0.01 (0.008)	-0.002	0.02
metaphorical*non-prepositional	0.02*** (0.004)	0.01	0.03
literal*prepositional	0.005 (0.008)	-0.007	0.01
literal*non-prepositional	0.01*** (0.004)	0.007	0.02

metaphorical*prepositional	0.02*** (0.008)	0.01	0.04
vowel in prefix	0.002 (0.002)	-0.002	0.007
prefix ends: stop	0.01*** (0.002)	0.005	0.01
prefix ends: vowel	0.01*** (0.002)	0.008	0.01
base begins: approximant	0.003 (0.005)	-0.007	0.01
base begins: fricative	-0.002 (0.005)	-0.01	0.006
base begins: nasal	0.004 (0.005)	-0.005	0.01
base begins: stop	0.02*** (0.005)	0.02	0.03
base begins: trill	0.01 (0.006)	-0.0008	0.02
base begins: vowel	-0.006 (0.006)	-0.01	0.002
syllable stressed	-0.007*** (0.0007)	-0.008	-0.005

Significance codes: *** $p < 0.001$, * $p < 0.05$. Notes: Adjusted $R^2 = 0.32$, $p < 0.0001$.

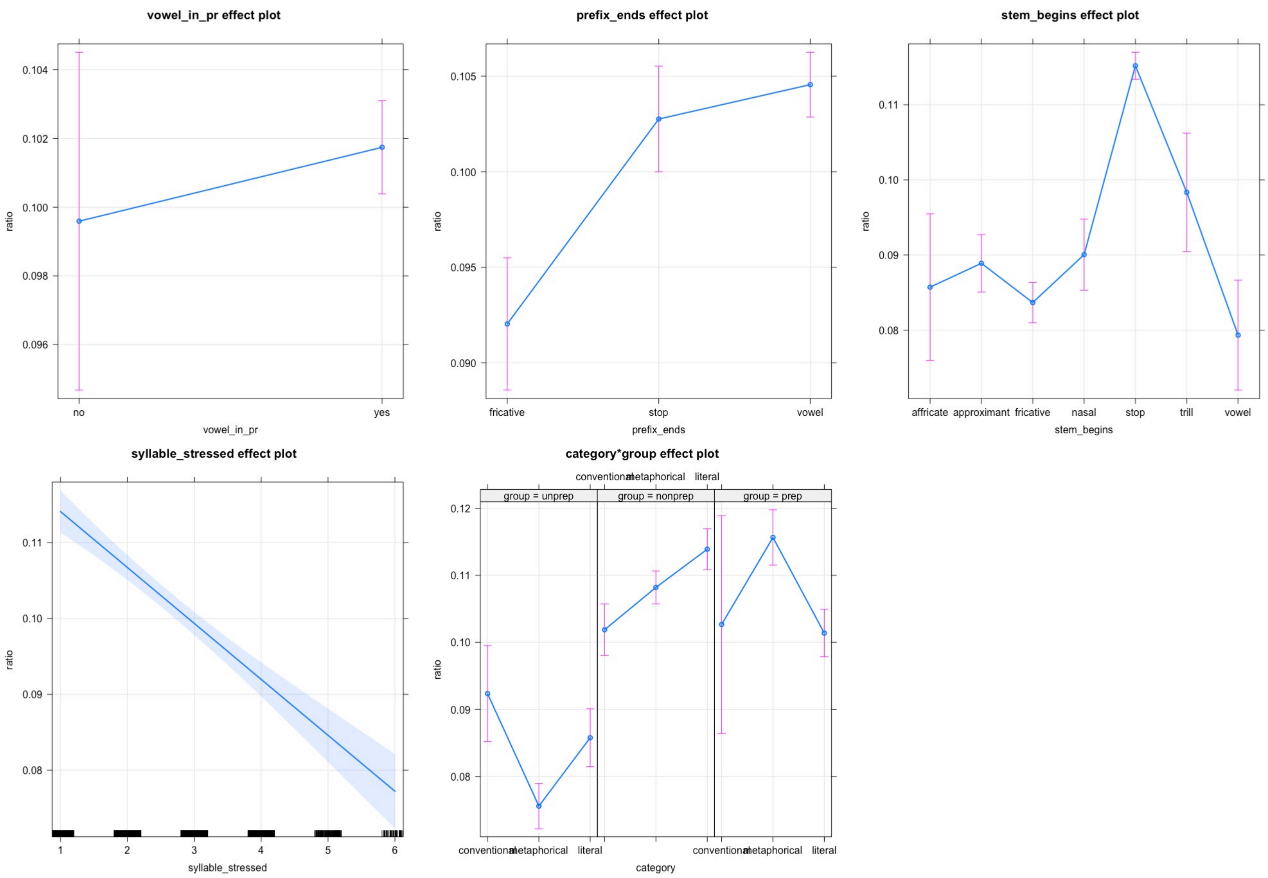

Figure 16. Main effects plot of Model 2

In line with my expectations, Model 2 was able to account for a greater amount of variance than Model 1. Several phonetic features, such as prefix-final and base-initial stops, were found to boost the pause ratio. There was also a predictable negative effect of the index of the stressed syllable: the further to the right the stress is from the prefix, the smaller the pause ratio tends to be. However, the most important thing to note is that, having controlled for all phonetic variation, I still found essentially the same pattern of pause ratio alignment across different semantic categories and compositional types. Some levels collapsed together, and in a sense, the new picture became more symmetrical, which can be represented with the help of Table 18.

Table 18. Groupings of semantic categories and compositional types based on the coefficients of Model 2

M[U]		C[U]	L[U]		
		C[N]	M[N]		L[N]
		C[P]	L[P]		M[P]

Note. [U] — unprepositional type, [N] — non-prepositional type, [P] — prepositional type, C — conventional meaning, M — metaphorical meaning, L — literal meaning.

This picture seems to make intuitive sense. Non-prepositional constructions do not have corresponding prepositions to encode the most basic, literal meanings, and that is why the L[N] group is characterized by the greatest pause ratio among all [N]: the prefix tends to be further detached from the base in pronunciation to accentuate this prefix's preposition-like behaviour. In contrast, prepositional constructions [P] do have corresponding prepositions for their literal meanings and so tend to employ a greater pause ratio for flagging construction-specific, metaphorical meanings M[P]. Given that unprepositional constructions do not relate to any prepositions at all, it should not surprise me that they are the mirror image of prepositional constructions.

Thus, instead of my previous two, one linear hierarchy of the locations of pause ratios for interacting categories and types of the prefix-base constructions can be proposed (see Figure 17):

M[U] (‘first’; M = 0.06) > C[U,N,P] / L[U] / M[N] / L[P] (‘second’; M = 0.09) > L[N] / M[P] (‘third’; M = 0.12). All these differences in locations are significant, as confirmed by the approximative k-sample Fisher-Pitman permutation test ($\chi^2(2) = 500.17$, $p < 0.0001$).

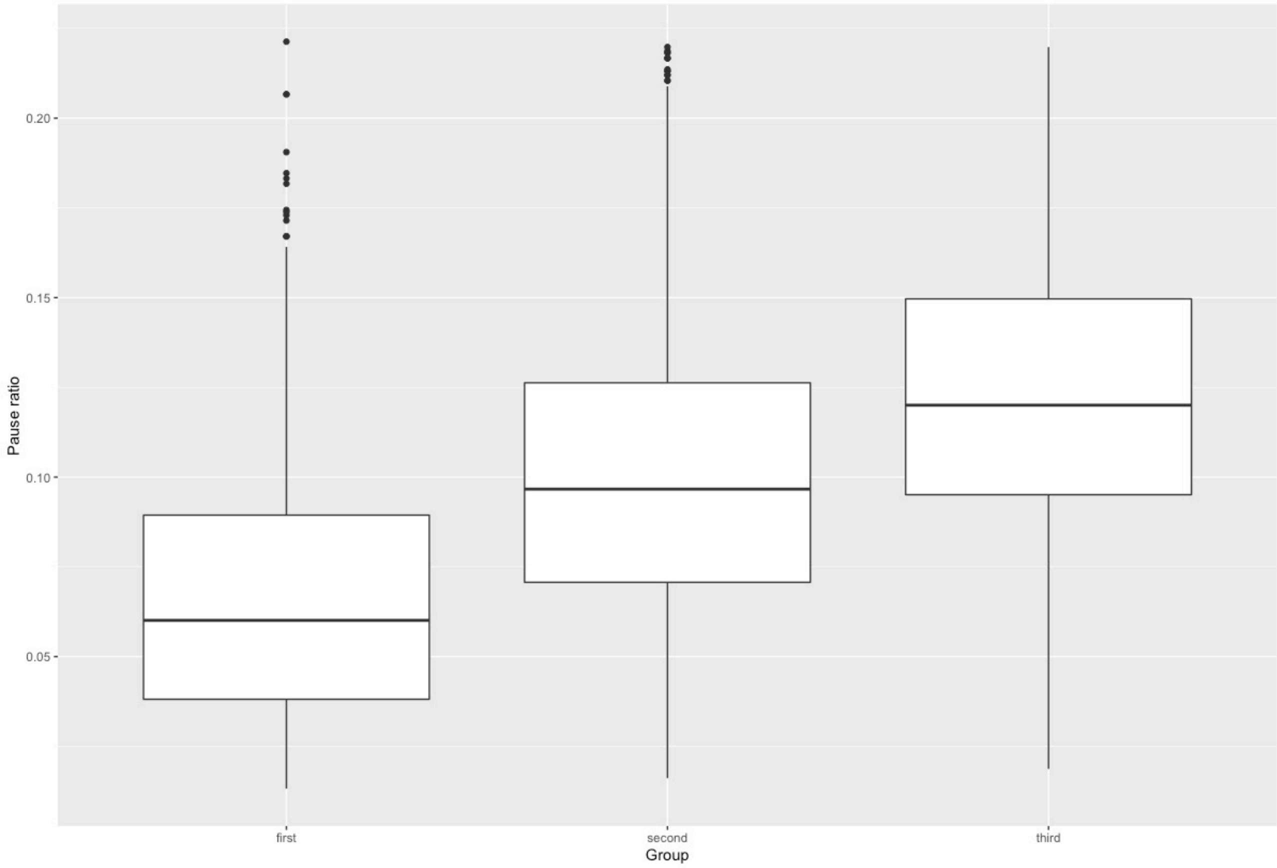

Figure 17. Locations of pause ratios for interacting categories and types of the prefix-base constructions

However, trying to connect these groups with the three populations discovered by the Markov chain Monte Carlo algorithm and use these groups for predicting the population membership of the observations in the same way as I did before (see Figure 12) leads to an unsatisfactory result: the accuracy of the new classification is worse than that of the previous one. What may be the reason for this? Presumably, one should search for the variance-inducing factors even deeper, on the level of specific semantic categories and compositional types’ pairings that accommodate multiple meanings.

3.3.4 The mystery of the third population revisited

Many Russian prefix-base constructions have more than one meaning per semantic category / compositional type unit. For example, two meanings of the prefix *na-* ‘to perform an action intensively’ and ‘to perform an action in a gentle, unobtrusive manner’ are both of the metaphorical category (*quality*/*manner*) and non-prepositional type. Similarly, two meanings of the prefix *pere-* ‘to extend an action to a specific, usually necessary or predetermined period of time’ and ‘to cease an action, usually after a long or intensive performing’ belong together to the literal category (*time*) and unprepositional type.

It is illuminating to see how the prediction errors of my Markov chain Monte Carlo algorithm are distributed across the constructions that have and do not have multiple meanings within the same semantic category / compositional type slot. The matrix of prediction errors is given in Table 19.

Table 19. Matrix of prediction errors across semantic categories and compositional types of different prefix-base constructions

prefix/ category	place	direction	time	relation	quality	quantity	state	limit
de-				1				
dis-				1				
do-				1 (II→III) P			1	1 (I←II) P
iz-		1			1 (I←II) N	2		1 (II→III) N
na-	1 (I←II) P	1			2 (II→III) N	1	1 (I←II) N	1 (I←II) N
nad-	2							
nedo-					1			
niz-		1						
o-		2 (II→III) N				1		1
ob-		2 (I←II) N				1	2 (I←II) N	1 (I←II) N
ot-		2 (I←II) P	1	2 (I←II) N	1		2	1 (I←II) N

pere-	1	1 (I→II) U	2 (I→II) U	1 (I→II) U	3 (I→III) U	1 (I→II) U	1 (I→II) U	
po-			2 (II→III) N		1 (II→III) P	1 (II→III) P		1 (II→III) N
pod-		3 (II→III) P		2 (II→III) N	2	1		1
pre-					1			
pred-	1 (I←II) P		1 (I←II) P					
pri-		1 (II→III) N		2 (I←II) P	1			1 (I←II) N
pro-		3 (II→III) N	1		1 (II→III) P	1 (II→III) P	1 (I←II) N	1 (I←II) N
raz-				1	1	1	1 (I→II) U	1
re-				1				
s-	1		1 (I←II) N				2 (II→III) P	1
so-				1				
u-	2 (I←II) N	1		1		1 (I←II) N	2 (I←II) N	1 (I←II) N
v-		2 (II→III) P						
voz-		1 (I→II) U	1	1				1
vy-		1	1		1 (I→II) U		1	1 (I→II) U
vz-		1	1		1			1 (I→III) U
za-	2 (II→III) N	1	2	1	1		2 (I←II) P	1

Note. U — unprepositional type, N — non-prepositional type, P — prepositional type; I — first population (smallest mean pause ratio), II — second population (medium mean pause ratio), III — third population (largest mean pause ratio); {1, 2, 3} — number of meanings encoded by the construction within a particular category/type slot.

Colour coding: yellow — the population of the construction's pause ratio is predicted correctly, red — the population of the construction's pause ratio is characterized by a smaller mean than predicted, green — the population of the construction's pause ratio is characterized by a larger mean than predicted, black bold frame — the construction's pause ratios are split between two different populations.

The coding scheme is complicated, so some comments are needed. For each prediction, I returned to my earlier and more accurate guess, but this time took even more agnostic approach. I hypothesised that the first population of pause ratios (the one with the mean of 0.06) comprises constructions with unprepositional prefixes, while the second population (the one with the mean of 0.11) comprises constructions with prepositional prefixes including prefix *po-*. As for the third population (the one with the mean of 0.17), I preferred to make no assumptions about it in the absence of better evidence and assigned no constructions to it.

According to these simple criteria, each observation in my data was labelled as coming from the first or second population of pause ratios. Then, it was checked to which population the observation actually belonged. To make sure I did not take into account any random fluctuations, only those observations of each prefix-base construction were retained that belonged to the same population of pause ratios as at least nine other observations, thus constituting together no less than 1/3 of the whole sample of 30 observations.

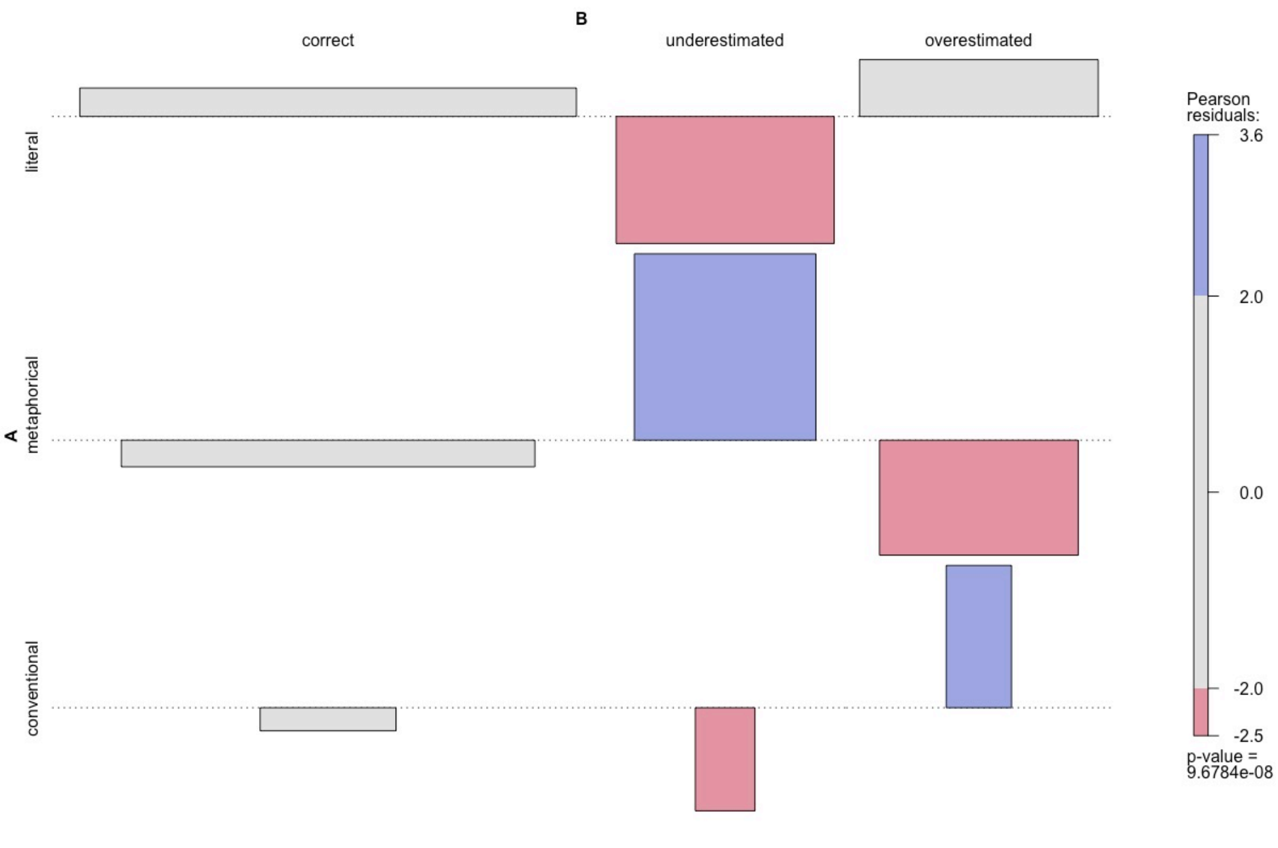

Figure 18. Prediction results for different semantic categories of prepositional constructions

After summarising the results, four different scenarios were possible for any construction: 1) it was labelled as correctly predicted if at least 20 of its observations came from the expected population, 2) it was labelled as underestimated if at least 20 of its observations came from the population with the mean pause ratio that was larger than expected, 3) it was labelled as overestimated if at least 20 of its observations came from the population with the mean pause ratio that was smaller than expected, or 4) it was labelled as split if at least 10 of its observations came from a different population than the rest.

The results are conspicuous. First of all, if one compares the association plots in Figures 18–20, one can once again see how the same semantic categories shift along the scale of three pause ratios' populations depending on the compositional type of the construction by which they are encoded. For the prepositional type, it is prefix-base constructions with metaphorical meanings that, according to my simplistic classification results, tend to be underestimated — that is, are characterized by many observations coming from the third population with a mean pause ratio of 0.17. At the same time, prepositional constructions with conventional meaning are significantly overestimated — that is, many of their observations come from the first population with a mean pause ratio of 0.06 (Figure 18).

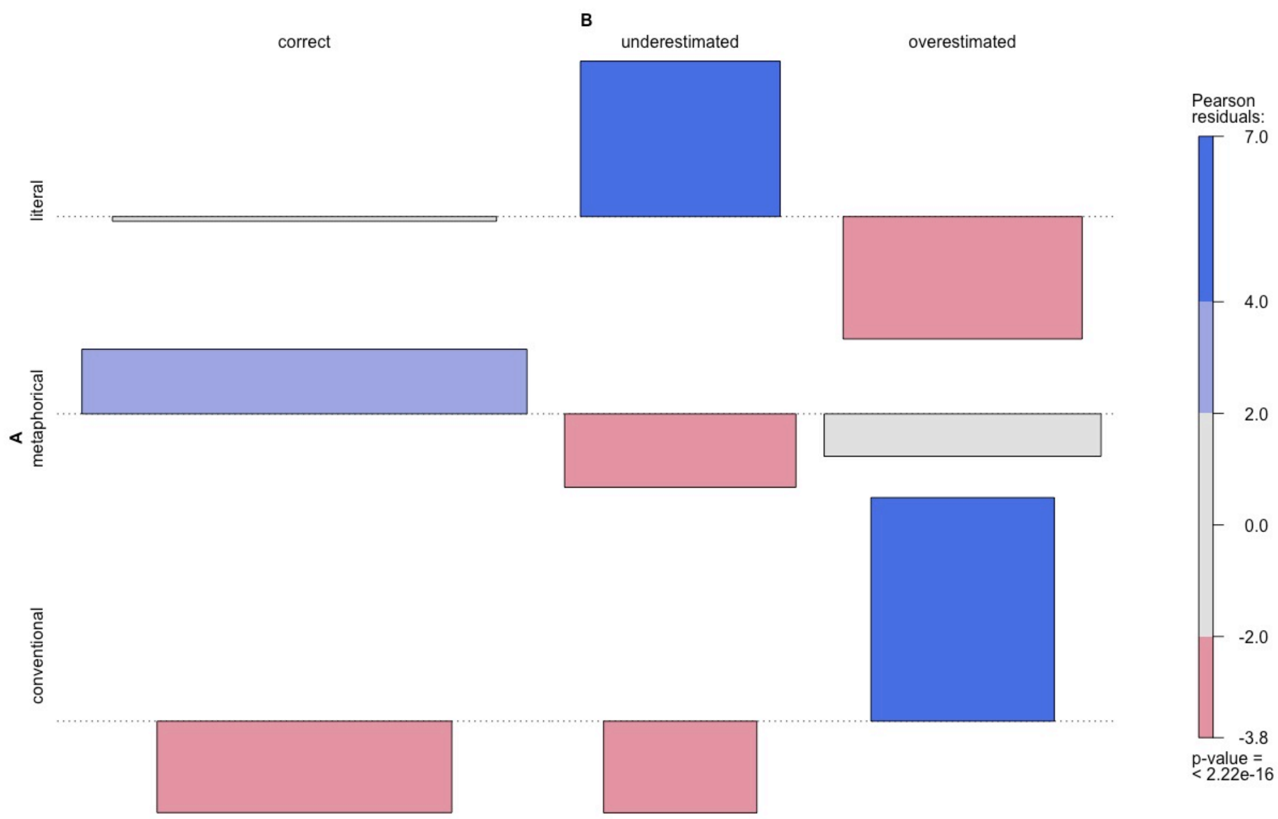

Figure 19. Prediction results for different semantic categories of non-prepositional constructions

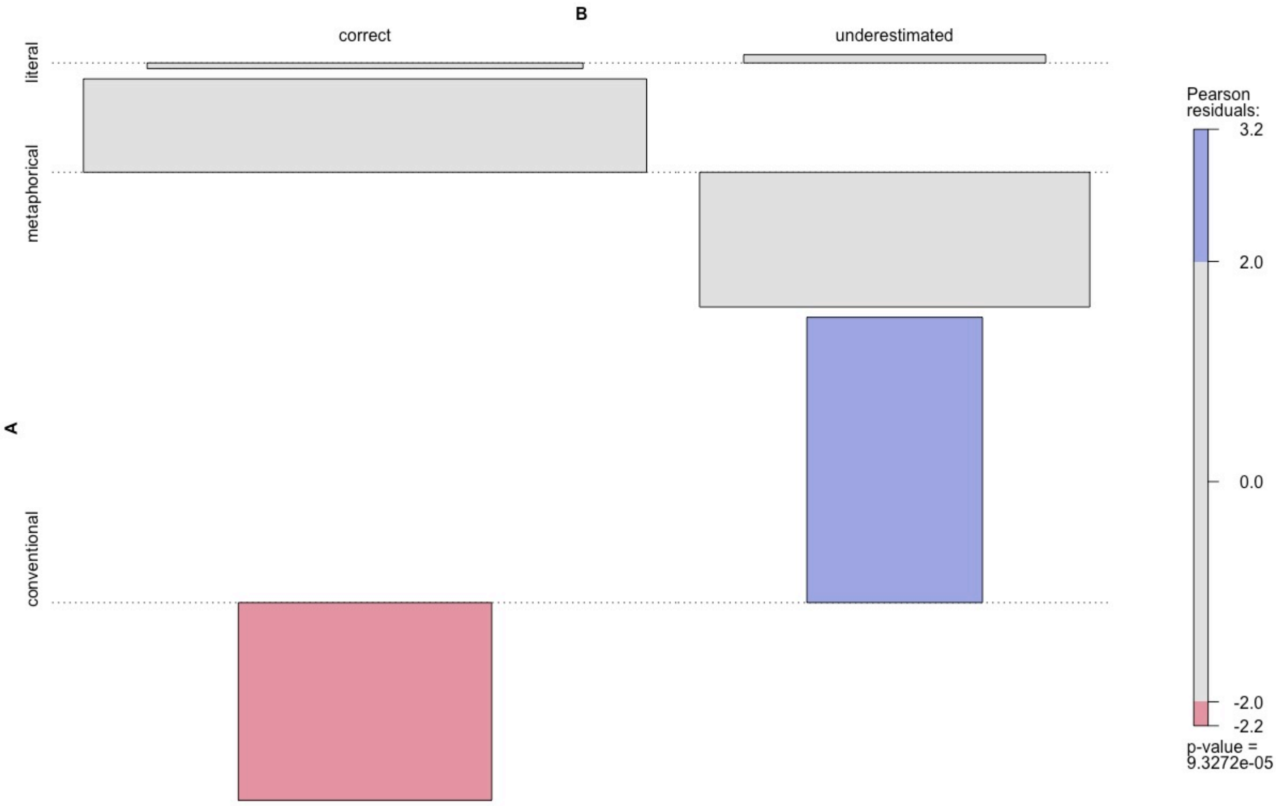

Figure 20. Prediction results for different semantic categories of unprepositional constructions

With the non-prepositional type, the situation is very different. While constructions with conventional meanings are also predominantly first population dwellers, here it is literal rather than metaphorical observations that tend to be underestimated by my classification procedure — that is, have pause ratios greater than the mean of the second population (Figure 19).

Finally, the observations of the unprepositional type cannot be overestimated, since the first population was chosen as the default one. Here, I can see that constructions with conventional meaning tend to be underestimated — that is, align with non-prepositional metaphoric and prepositional literal counterparts as representatives of the second population of pause ratios (Figure 20).

All of this, however, only confirms what I already know. It doesn't explain why the overall results of the classification that takes all these factors into account are so dissatisfying. To understand it, one should look at how dissimilar the results of classification are for the constructions with one and more than one meaning per category/type slot. First, as confirmed by Pearson's chi-squared test, the odds of a construction being classified correctly are 4.65 times higher if the construction has one meaning per category/type slot than if it has multiple meanings ($\chi^2(2) = 12.03$,

$p = 0.002$). In other words, all unprepositional constructions with one meaning tend to come from the first population with a mean pause ratio equal to 0.06, while all monosemous (within category/type slot) constructions with prefixes related to prepositions tend to come from the second population with a mean pause ratio equal to 0.11. Second, in the constructions that have more than one meaning per category/type slot, split distributions are significantly overrepresented compared to homogeneous ones (Figure 21).

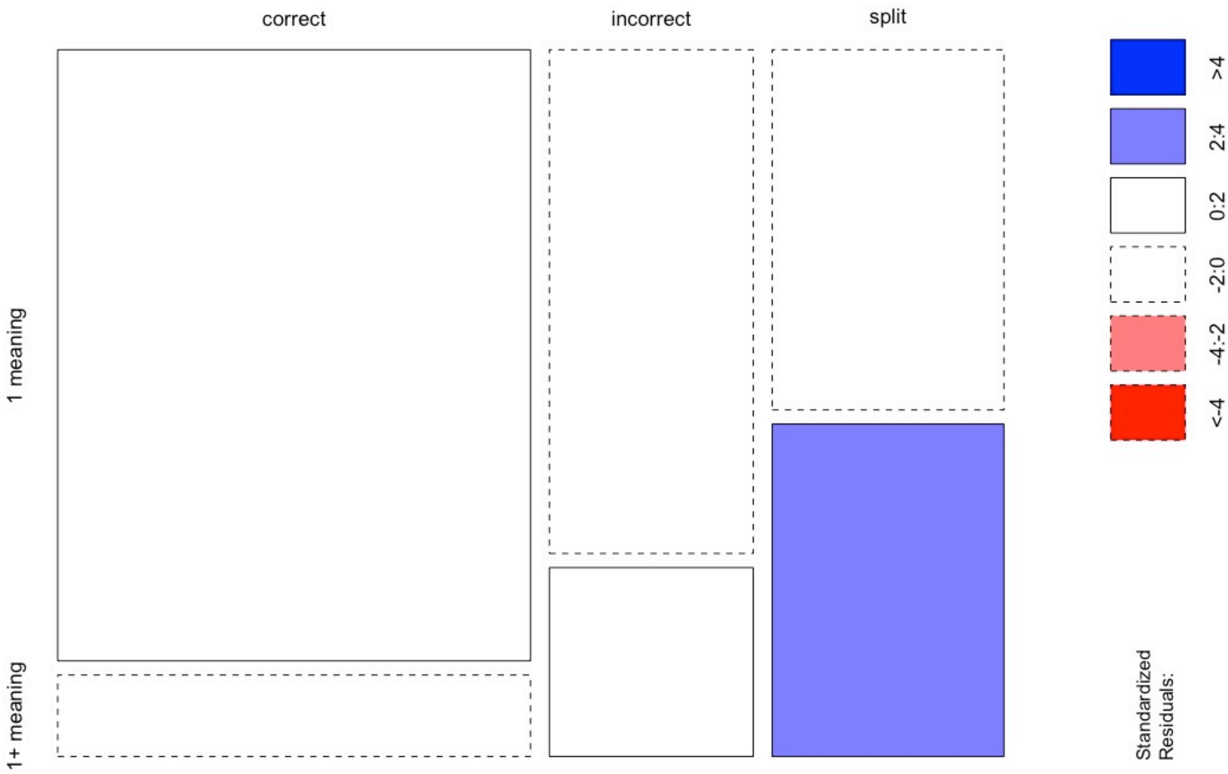

Figure 21. Prediction results in relation to the number of constructional meanings per category/type slot

These results suggest an insight on why my classification attempts failed. Speakers of Russian, when confronted with a prefix-base construction that has multiple meanings for a single category/type slot, try to disambiguate those meanings by keeping one pause ratio as a reference level and shifting another to a lower (if possible) or higher population, depending on the preferences of a particular category/type slot.

This phenomenon can be illustrated with the help of the following example. In Figure 22, the densities of the pause ratios of four different prefix-base constructions with prefix *za-* are plotted. All of them represent the non-prepositional compositional type. As for the semantic categories, two constructions encode metaphorical meanings (*state*, label M[N] in Figure 22's legend): (1) 'to bring someone to an undesirable state' (*za-draznitj* 'to humiliate by taunting') and (2) 'to get, earn, grab something' (*za-voevatj* 'to conquer'). Two other constructions encode literal meanings (*place*, label L[N] in Figure 22's legend): (3) 'to apply an action to a part of the object' (*za-tesatj* 'to make thinner by cutting') and (4) 'to cover up, close with something' (*za-pudritj* 'to powder').

Three vertical dotted lines labeled I, II, and III mark the means of the three populations of pause ratios respectively. Given my initial guess, I would expect to see that the means of all four constructions would be located near the second population's mean. However, this is true only for constructions (2) and (3). Distributions (1) and (4) are, in fact, split: some of their observations obviously belong to the second population, but most are shifted either to the left, towards the first population's mean, which is the case with (1), or to the right, towards the third population's mean, which is the case with (4).

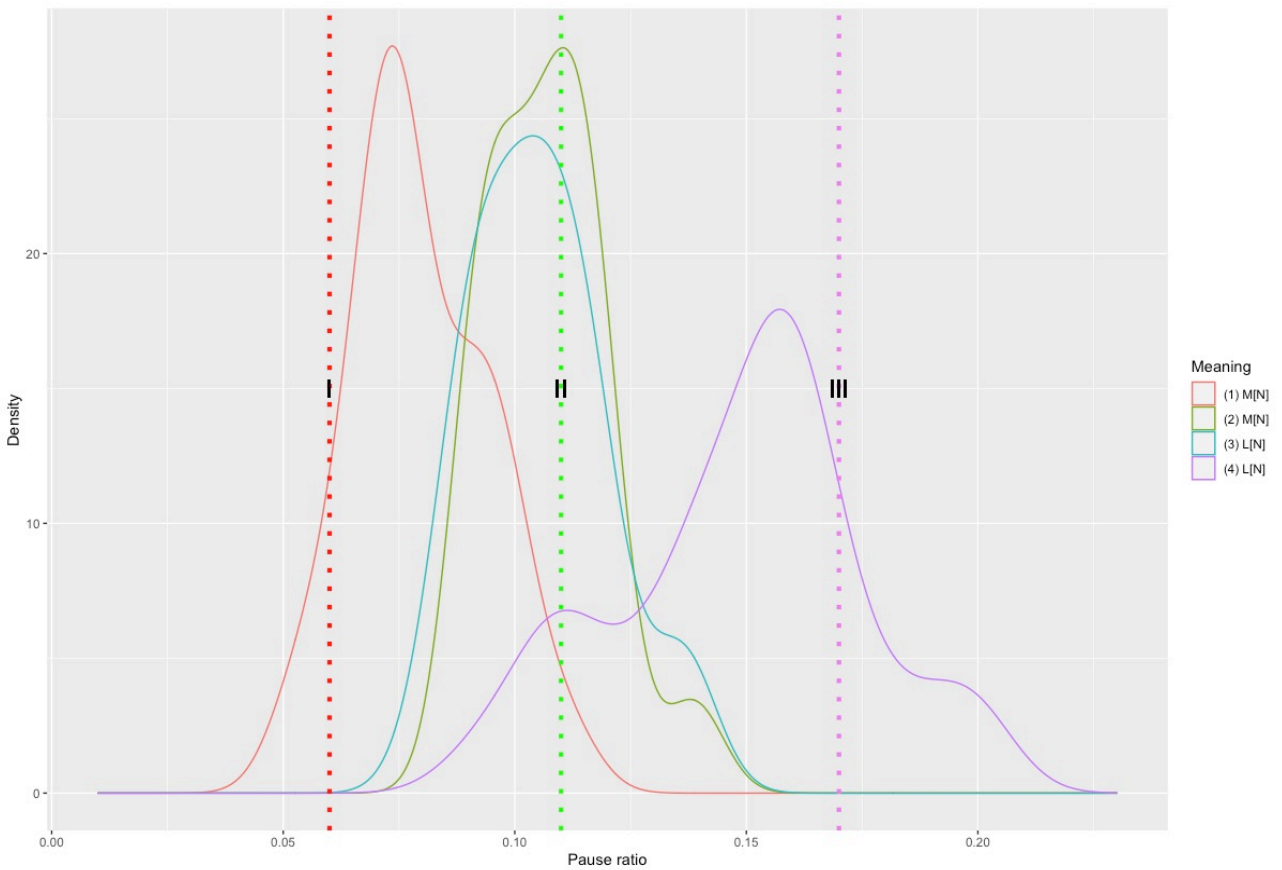

Figure 22. Pause ratio distributions of four constructions with prefix *za-* sharing two semantic categories

Taking into account that (1) and (2), on the one hand, and (3) and (4), on the other hand, share the same semantic categories while being of the same compositional type, I can see some evidence of audience design in these prefix-base constructions. As I have noted before, in general, constructions of the non-prepositional type tend to employ a greater pause ratio for flagging literal meaning, presumably to accentuate the prefix's preposition-like behaviour. However, when a non-prepositional construction can convey several literal meanings of the same category, only one of its realisations gets this promotion, and the other is treated as the default. Conversely, when a non-prepositional construction encodes two different metaphorical meanings, only one of its realisations is left as the default, while the other one is downgraded to the lowest level of prefix-base pronunciational detachment, on which this construction actually becomes actually desemanticised and its prefix starts being treated as just a marker of perfectivity.

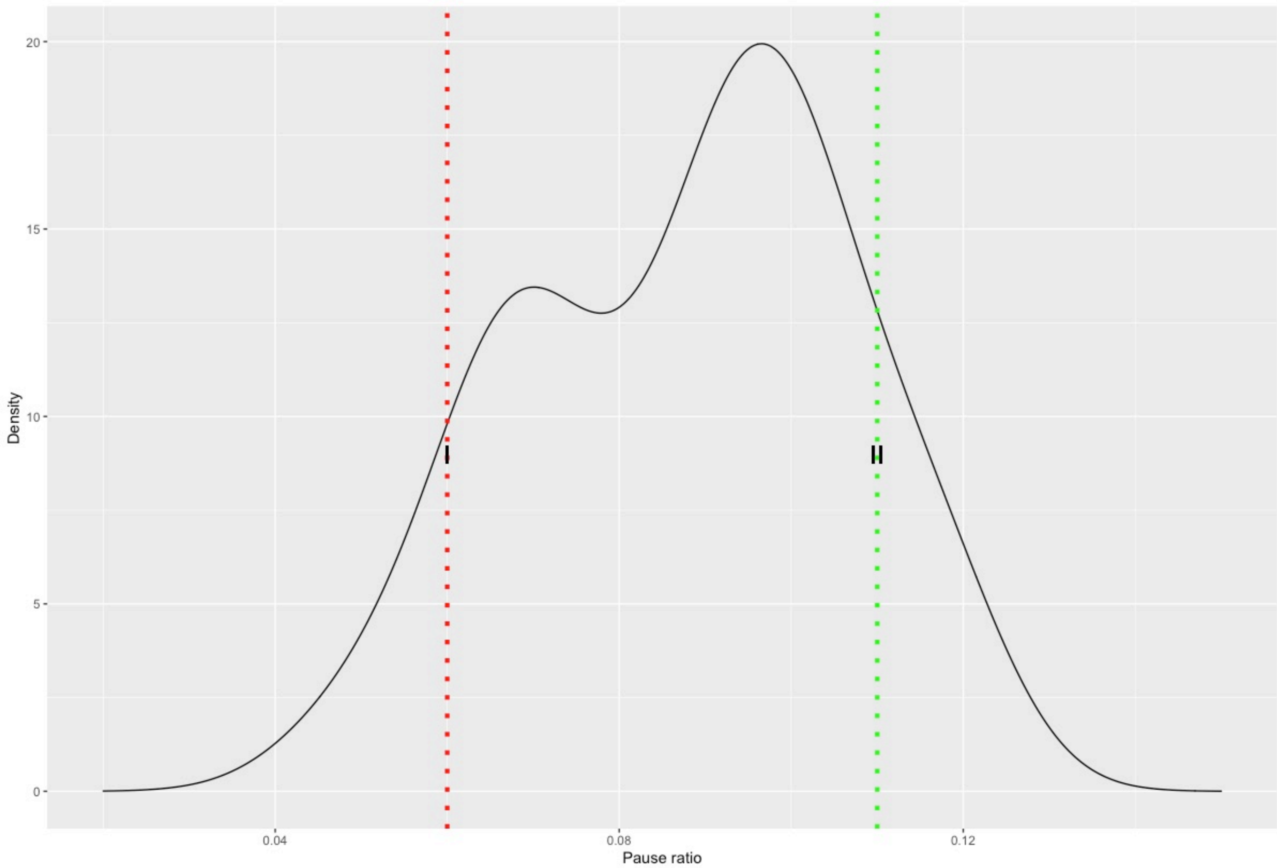

Figure 23. Split pause ratio distribution of the construction *do-chitatj*

The prefix-base constructions' polysemy facilitates the process of disambiguation but does not necessarily trigger it. It is often the case that speakers may have contrasting interpretations of the constructions that have one meaning per category/type slot, which results in split distributions of pause ratios in my data. Let us consider another example. Figure 23 shows the density plot of the pause ratios collected during my experiment for the prefix-base construction *do-chitatj* 'to finish reading'. As the allowing for variability constructional meaning 'to bring an action to an end or to a limit' suggests, this distribution may result from the fact that some of my experiment's participants activated the 'bringing to a limit' possibility of interpretation, while others interpreted the verb as the 'bringing to an end' instance. So in the pronunciation of the former, *do-chitatj* with its conventional meaning, was a member of the first population of pause ratios, while in the pronunciation of the latter, it shifted to the second population, which, as I have found, is characteristic for the constructions of the prepositional type encoding literal meanings (*time*).

Taking this into consideration, I can now account for the low accuracy of predicting the population of pause ratios by a prefix of the construction. It is appropriate to think back to the prefixes, for which the accuracy score was especially disappointing: *pere-*, *vy-*, *pred-*, *na-*, and *ob-* (see Figure 13). All these prefixes are outliers in their respective groups. The first two of them, while being of the unprepositional type, have developed a wide variety of literal and metaphorical meanings and thus tend to be aligned by speakers of Russian with non-prepositional and prepositional prefixes. According to the matrix in Table 19, most of their meanings' pause ratios are underestimated by my prediction algorithm and belong to the second or even third populations. At the same time, prefixes *pred-* and *na-*, on the one hand, and *ob-*, on the other, can be considered prototypical examples of, respectively, L[P] (literal meaning, prepositional type) and M[N] (metaphorical meaning, non-prepositional type) groups, which do not generally presuppose a great degree of prefix's detachment from its base. That is why, as the matrix in Table 19 shows, these constructions are routinely downgraded to the first population with the lowest mean pause ratio, aligning with unprepositional constructions.

3.4 Conclusion

I started this study with a hypothesis that in Russian pronunciation, there is a longer pause between prepositional prefixes and bases than between unprepositional prefixes and bases due to the compositional nature of the former and the non-compositional nature of the latter. Having shown that this is indeed the case, I, however, could not help noticing that the actual variability of pause ratios resists being reduced to just two homogeneous groups of values.

Without knowing how to account for this variability, I resorted to the Markov chain Monte Carlo algorithm, which successfully identified three underlying populations of values instead of the two I initially expected. In order to provide an explanation of how the actual data may be reasonably mapped onto the revealed distributions, I suggested taking into account that the 'prepositonality' of a prefix-base construction is a matter of degree and that this continuum can be subdivided two-dimensionally: first, along the axis of semantic category, and second, along the axis of compositional type. Thus, in lieu of the original dichotomous division, each of the 125 prefix-base constructions in my data was assigned to a specific slot in a 3-by-3 matrix matching one of the possible semantic categories (literal, metaphorical, and conventional) with one of the possible compositional types (prepositional, non-prepositional, and unprepositional).

A regression analysis that factored in these grouping predictors while controlling for the random variation of segmental and suprasegmental features of particular prefixed verbs revealed that with different compositional types of prefix-base constructions, different semantic categories are characterized by the largest pause ratio: unprepositional constructions give preference to conventional meanings, non-prepositional constructions to literal meanings, and prepositional constructions to metaphorical meanings.

This picture makes intuitive sense. Non-prepositional constructions do not have corresponding prepositions to encode the most basic, literal meanings, and that is why this semantic category is characterized by the largest pause ratio among them: the prefix tends to be further detached from the base in pronunciation to accentuate this prefix's preposition-like behavior. In contrast, prepositional constructions do have corresponding prepositions for their literal meanings and so tend to employ a greater pause ratio for flagging construction-specific, metaphorical meanings. As for the unprepositional constructions, since they do not relate to any prepositions at all, the conventional meaning is naturally the most distinguishable within this group.

Finally, I showed some evidence of audience design in the production of prefix-base constructions by Russian speakers. When confronted with a prefix-base construction that has multiple meanings for a single category/type slot, the speakers try to disambiguate those meanings by keeping one pause ratio as a reference level and shifting another to a lower (if possible) or higher population, depending on the preferences of a particular category/type slot. Thus, verbs with unprepositional prefixes may behave like verbs with prepositional prefixes and vice versa depending on how the speaker interprets the meaning of a particular construction and how explicit he or she wants to make this meaning for the hearer.

4 Inter-morpheme periods of silence in Russian prefixed verbs: a follow-up study

4.1 Introduction

The main results of the experiment reported in the previous chapter can be formulated as follows. First, there is a significant difference in the length of silent periods between prefix and base for verbs with prepositional (e.g., *v*-) and unprepositional (e.g., *pere*-) prefixes. Prefixes that have prepositional counterparts tend to be separated from their bases by a longer pause than prefixes that have none.

Second, with different compositional types of prefix-base constructions (prepositional, non-prepositional, and unprepositional), different meaning categories (literal, metaphorical, and conventional) are characterized by the longest silent period: 1) prepositional constructions have corresponding prepositions for their literal meanings and so tend to employ a greater pause ratio for flagging construction-specific, metaphorical and conventional meanings, 2) unprepositional constructions do not relate to any prepositions at all and thus represent the mirror image of prepositional constructions, flagging most basic literal meanings, 3) non-prepositional constructions fall somewhere in between. On the one hand, they have their corresponding prepositions and so align with prepositional constructions in that prefixes with construction-specific meanings tend to be flagged. On the other hand, their literal meanings do not coincide with the meanings encoded by respective prepositions, and so these constructions align with unprepositional constructions in that the prefix tends to be further detached from the base in pronunciation to accentuate this prefix's preposition-like behaviour.

Third, when confronted with a prefix-base construction that has multiple meanings for a single category/type slot, the speakers try to disambiguate those meanings by keeping one silent period as a reference level and shifting another to a lower (if possible) or higher level, depending on the preferences of a particular category/type slot.

The current follow-up study was designed to verify the results of the previous experiment and answer the following questions. Will the observed differences hold if one takes absolute rather than relative lengths into account? Will the observed differences hold if one controls for all phonetic differences in target verbs and considers only the variability that is left unexplained by these factors? Will the observed differences hold if one replaces real bases in target verbs with nonce base while retaining the prefixes? In other words, do the observed differences pertain to the participants' acquaintance with target verbs (their parsability, language frequency, etc.) or to the prefix-base constructional schemas?

4.2 Experimental design and data

To ensure comparability of two experimental settings, the follow-up study retained much of the previous experiment's design. Same 28 Russian verbal prefixes and same 125 construction-meaning pairings (91 meanings for prepositional prefixes and 34 meanings for unprepositional prefixes) came under investigation. Comparable number of participants was employed: each sentence was pronounced, on average, by 29 different native speakers.

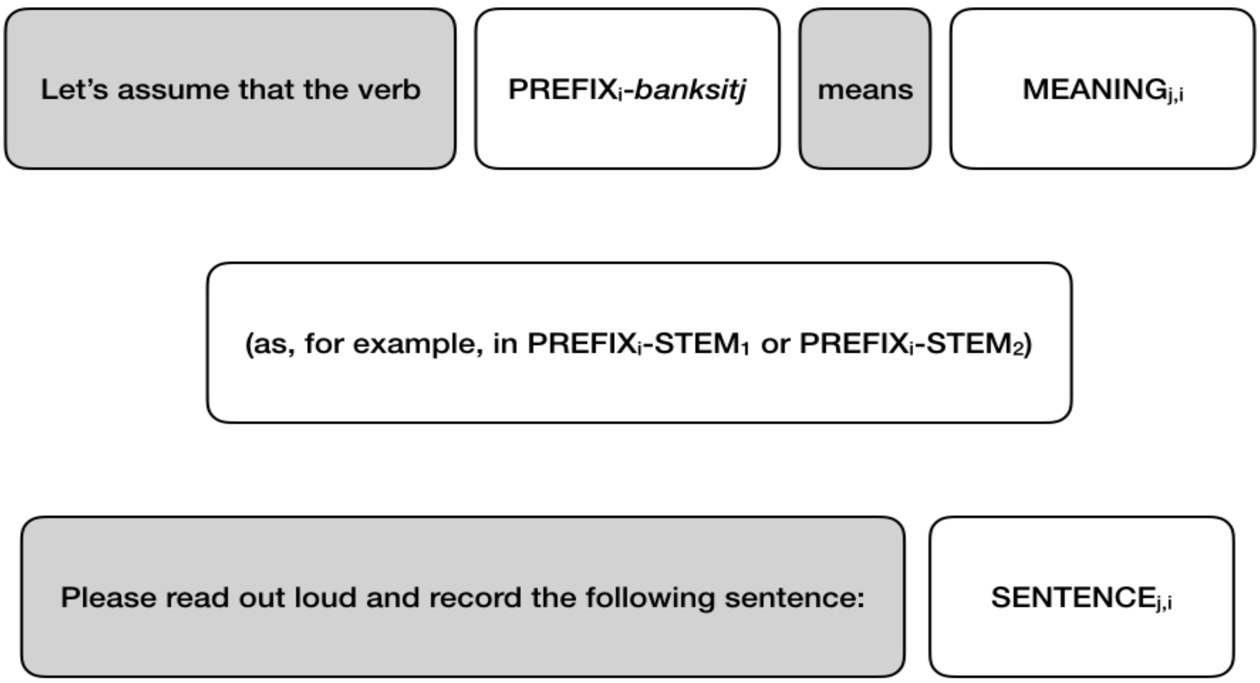

Figure 24. Instruction for the participants' template

In some important aspects, however, experimental design of the follow-up study differed from that of the initial experiment. While previously I illustrated each construction meaning with a sentence containing respective verb from the Russian National Corpus, now I used self-invented formulaic sentences: minimal context, necessary to correctly semanticise the verb (no longer than seven words), SVO order, ditransitive (where possible), same subject (proper noun), all verbs used in past tense. In each verb, an actual base was replaced with a nonce base -*banksi*-.

Instructions for the participants were also different, written according to the template in Figure 24, where $PREFIX_i \in \{PREFIX_1, \ldots, PREFIX_{28}\}$; $MEANING_{j,i}$ is one of the meanings attested for $PREFIX_i$ in Russian Grammar; $PREFIX_i$-$STEM_1$ and $PREFIX_i$-$STEM_2$ are real verbs provided as examples illustrating $MEANING_{j,i}$ in Russian Grammar; $SENTENCE_{j,i}$ contains the verb $PREFIX_i$-*banksitj* in the $MEANING_{j,i}$. Words $PREFIX_i$-$STEM_1$ and $PREFIX_i$-$STEM_2$ were selected in such a way that their initial phonemes differed in the manner of articulation and either of them could be potentially substituted for the verb $PREFIX_i$-*banksitj* in the $SENTENCE_{j,i}$ without rendering it senseless.

The following sentence provides an example of the instruction: 'Dopustim, chto glagol PROBANKSITJ oznachaet 'napravitj dejstvie skvozj chto-libo vnutrj' (kak, naprimer, **progryztj** ili **probitj**). Pozhalujsta, prochitajte na diktofon sledujushchuju frazu: *Petya probanksil sebe dorogu*'. In English: 'Let's assume that the verb PROBANKSITJ means 'to direct an action through something' (as, for example, in **gnaw through** or **break through**). Please read out loud and record the following sentence: *Petya probanksil his way through*'.

Having obtained the results (3,593 observations), I hand-segmented the acoustic waveforms of target verbs in Praat (Boersma and Weenink, 2020), manually coded visually identifiable periods of silence at the boundaries between prefix and base and measured their lengths in milliseconds. Since one of the goals of the follow-up study was to see whether the previously observed differences hold if one takes into account absolute rather than relative values, the target verbs from the initial experiment were measured in the same way (3,696 observations). Merging two samples gave me 7,289 observations in total.

After that, I fitted a linear regression model to the data to account for the part of variability in the lengths of the periods of silence that is induced by phonetic factors only. The following predictor variables were used: 1) the presence of vowel in the prefix (yes or no), 2) the type of the prefix's final phoneme (fricative, stop, vowel), 3) the type of the base's initial phoneme (affricate, approximant, fricative, nasal, stop, trill, vowel), 4) the index of the stressed syllable, and 5) the overall number of syllables. From the linear regression output, I got the residuals that now quantify the variability that cannot be attributed to phonetic differences in target verbs, but rather is indicative of the fact that those observations come from different populations. All analyses to follow were performed on the residuals.

4.3 Data modelling

First thing to take into account is that there is a lot of structure in the data (Figure 25). On the most basic level, one finds 7,289 individual observations. Each observation belongs to one of 125 unique sentences that together represent all possible construction-meaning pairings. Next, each sentence can be viewed as an instantiation of one of the 28 Russian verbal prefixes. Each prefix, then, can be subsumed under one of three meaning categories: literal, metaphorical, or conventional. Each of those meaning categories can be expressed by either of two possible construction types: prepositional or unprepositional. Finally, both construction types are attested for real and nonce verbs.

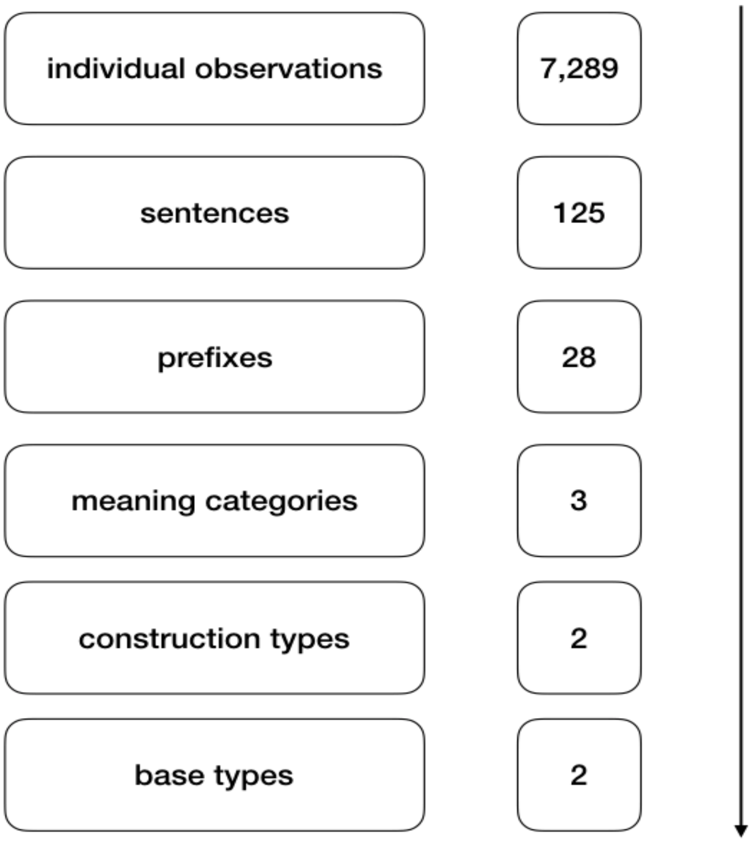

Figure 25. Structure of the experimental data

Given this multi-layered structure and my goal of finding out how interaction of different layers affects the separability of prefix and base in pronunciation, the most natural way to analyse the obtained residuals is hierarchical modelling. I used Markov chain Monte Carlo methods to build a hierarchical model of the data and create a posterior distribution of the parameters of interest. Hierarchical modelling was performed in a number of consecutive steps, in a bottom-up fashion, so as to replicate the assumed structure of the data.

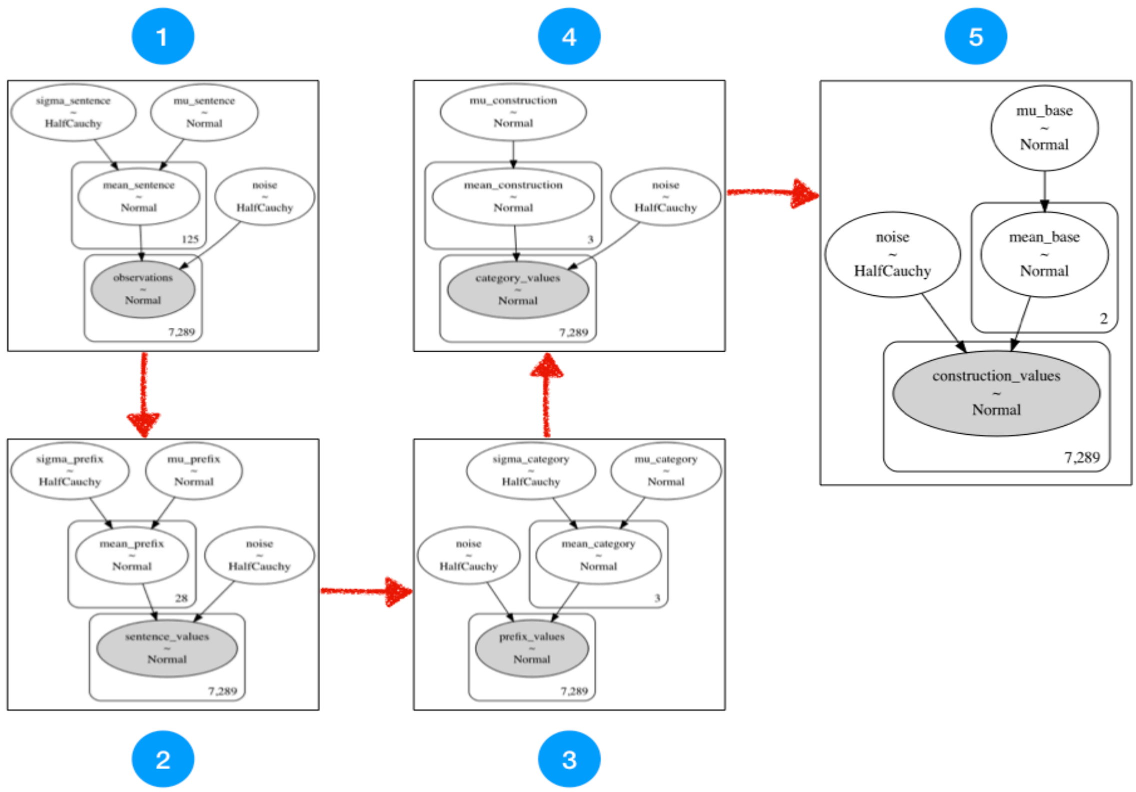

Figure 26. Structure of the hierarchical model

The algorithm is visualised in Figure 26. First, mean value of residuals for each sentence was obtained drawing on individual observations (observation → sentence); it was assumed that sentence values come from one normal distribution with hyperparameters *mu_sentence* and *sigma_sentence*. Second, values of individual observations were replaced with respective sentence mean values, after which mean value for each prefix was obtained drawing on them (sentence → prefix); it was assumed that prefix values come from one normal distribution with hyperparameters *mu_prefix* and *sigma_prefix*. Third, sentence mean values were replaced with respective prefix mean values, after which mean value for each meaning category was obtained drawing on them (prefix → category); it was assumed that category values come from one normal distribution with hyperparameters *mu_category* and *sigma_category*. Fourth, prefix mean values were replaced with respective category mean values, after which mean value for each construction type was obtained drawing on them (category → construction); it was assumed that construction values come from one normal distribution with hyperparameter *mu_construction* and $\sigma = 5$. Finally, category mean values were replaced with respective construction mean values, after which mean value for each type of base was obtained drawing on them (construction → base); it was assumed that base values come from one normal distribution with hyperparameter *mu_base* and $\sigma = 5$. Hyperparameters *mu_sentence*, *mu_prefix*, *mu_category*, *mu_construction*, and *mu_base* were all sampled from a Gaussian distribution ($\mu = 0$, $\sigma = 5$). Hyperparameters *sigma_sentence*, *sigma_prefix*, and *sigma_category* were all sampled from a Cauchy distribution ($\beta = 10$) truncated to only have nonzero probability density for values greater than or equal to the location of the peak.

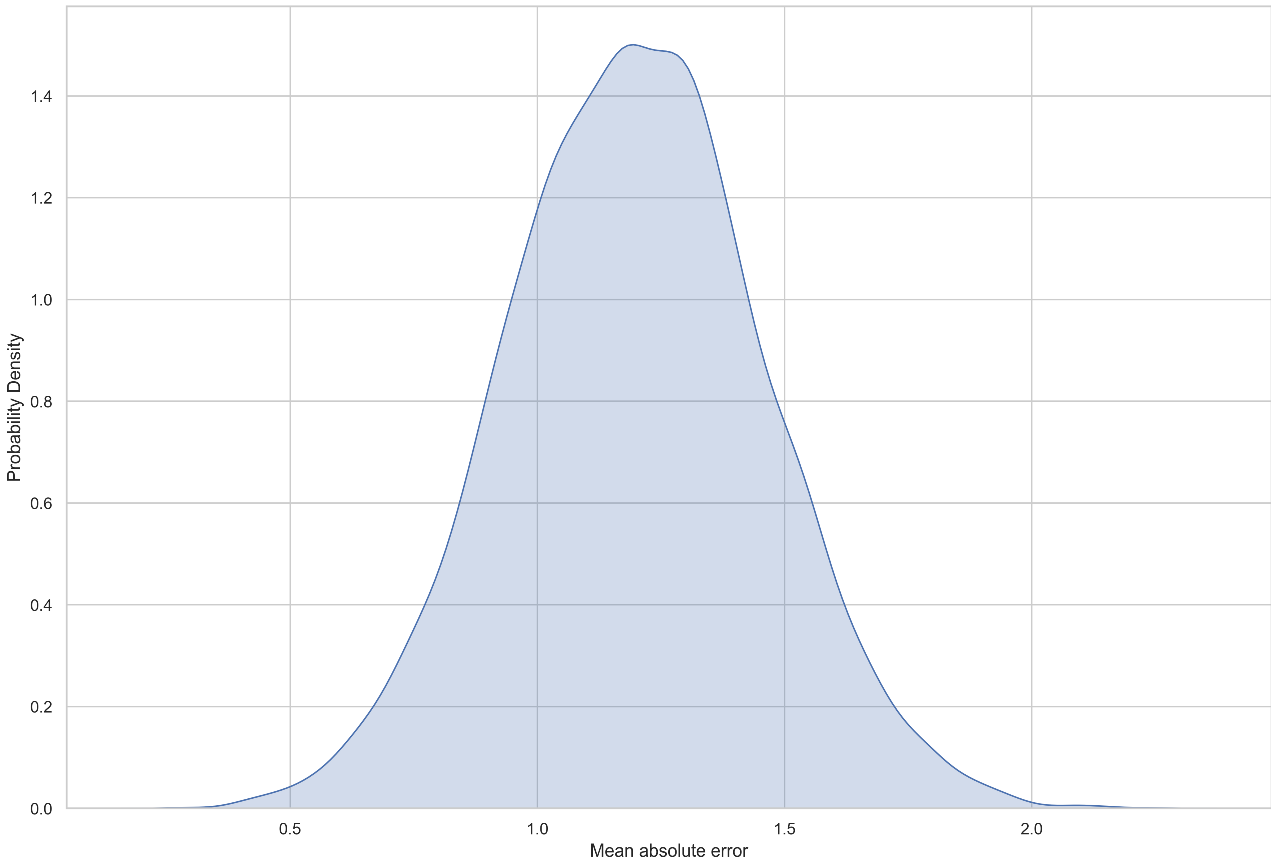

Figure 27. Probability density of the distribution of mean absolute errors

At the next stage, the individual observations were modelled as coming from a Gaussian distribution with mean equal to a linear combination of terms $\beta_{0,sentence} + \beta_{1,sentence} * \mu_{sentence} + \beta_{0,prefix} + \beta_{1,prefix} * \mu_{prefix} + \beta_{0,category} + \beta_{1,category} * \mu_{category} + \beta_{0,construction} + \beta_{1,construction} * \mu_{construction} + \beta_{0,base} + \beta_{1,base} * \mu_{base}$ and variance equal to linear combination of terms $\sigma_{sentence} + \sigma_{prefix} + \sigma_{category} + \sigma_{construction}$

+ σ_{base}. 5,000 samples were drawn in three different chains (each with initial 2,000 burn-in iterations), resulting in the total of 15,000 posterior distributions.

To check the accuracy of the model, I calculated the mean absolute error: $E_j = \Sigma \, || observed_i| - |simulated_i|| \, / \, N$, for N = 7289, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, 15000\}$. The probability density of the distribution of mean absolute errors is visualised in Figure 27. The mean of this distribution can be computed to lie within the interval of [1.19, 1.20] with 95% certainty.

Yet another posterior predictive check of the model is given in Figure 28. In the left panel of this figure, one can see values averaged across all samples from the posterior plotted against real observations. In the right panel, real observations are overlaid with individual values obtained from the sample characterised by least mean absolute error [$B = \mathrm{argmin}_E(E_1, \ldots, E_{15000})$]. It is clear that the model reproduces the data generating process fairly well.

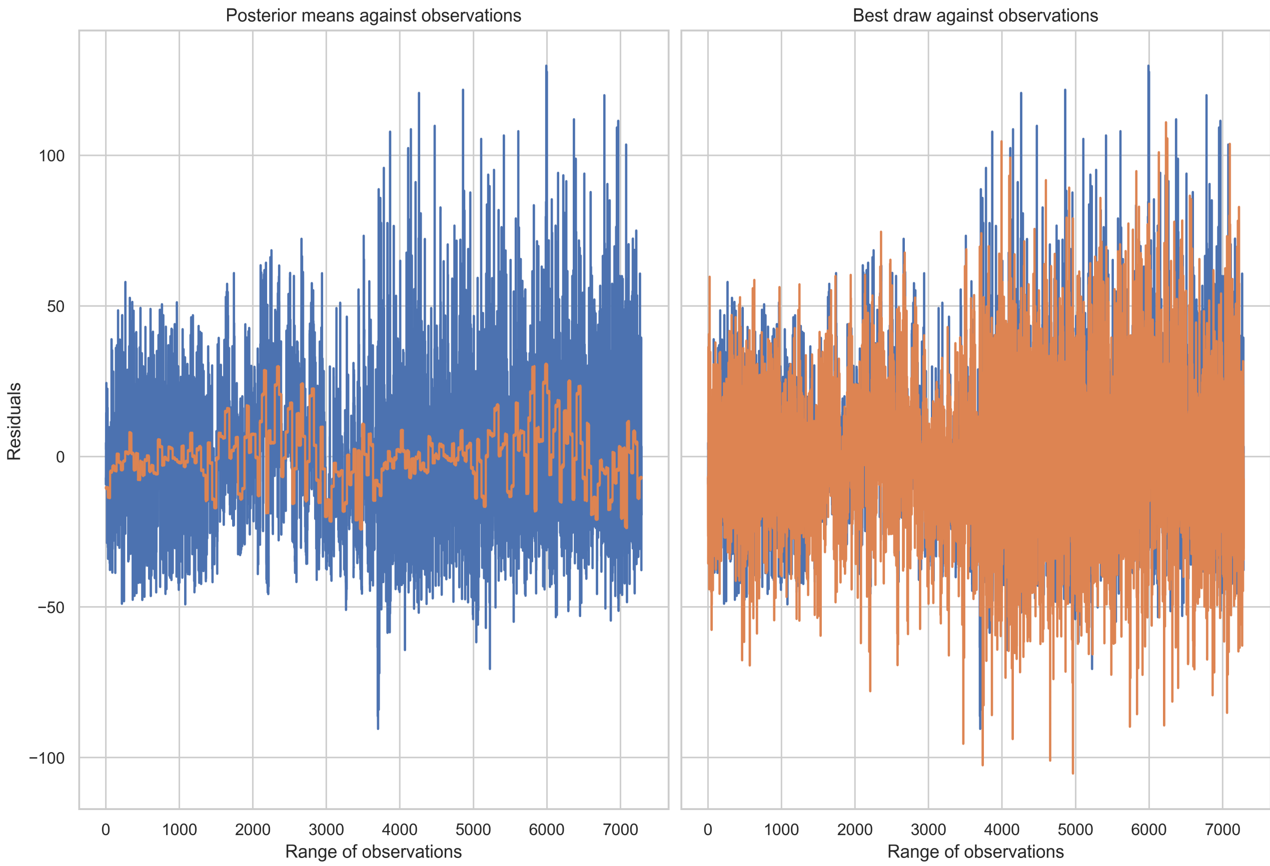

Figure 28. Values from the posterior plotted against real observations

4.4 Inferences from the model

With a reasonably accurate model at hand, it is now possible to execute all sorts of probabilistic queries, by simply selecting a subset of values that meet certain criteria from the posterior distribution with least mean absolute error and then dividing the length of this subset by the length of the sample. Specifically, I am interested in estimating quantity P(Residual > 0 | f_m, f_c, f_b), where {f_m, f_c, f_b} is the set of all possible values of the variables *meaning category*, *construction type*, and *base type*. This conditional probability is interesting because observing a positive residual means that the length of respective period of silence between prefix and base is greater than what can be explained away by purely phonetic factors.

The main obtained results are as follows. The probability of observing a positive residual with nonce verbs is greater than such probability for real verbs (Table 20). This is an anticipated result: participants found it easier to 'constructionalise' a verb, that is, to make a greater pause between prefix and base in pronunciation, when the prefix was familiar to them and the base was unknown.

Table 20. Probabilistic inference: base types

type of verb	P (Residual > 0)
real verbs	0.486
nonce verbs	0.503

Next, I found that, both for real and nonce verbs, three different construction types form a continuum: it is more likely to observe a positive residual with prepositional prefixes, less likely with non-prepositional prefixes, and even more unlikely with unprepositional prefixes (Table 21).

Table 21. Probabilistic inference: construction types

construction type	P (Residual > 0)	
	real verbs	nonce verbs
prepositional	0.525	0.527
non-prepositional	0.519	0.526
unprepositional	0.394	0.441

With meaning categories, trends are, again, identical for real and nonce verbs and yet somewhat ambiguous: in constructions with conventional meaning, positive residuals are most probable, constructions with literal and metaphorical meanings follow behind (Table 22).

Table 22. Probabilistic inference: meaning categories

meaning categories	P (Residual > 0)	
	real verbs	nonce verbs
literal	0.489	0.502
metaphorical	0.475	0.497
conventional	0.528	0.531

Table 23. Probabilistic inference: construction types and meaning categories

construction types	meaning categories	P (Residual > 0)	
		real verbs	nonce verbs
prepositional	literal	0.471	0.490
	metaphorical	0.590	0.573
	conventional	0.679	0.635
non-prepositional	literal	0.495	0.505
	metaphorical	0.505	0.517
	conventional	0.583	0.571
unprepositional	literal	0.524	0.526
	metaphorical	0.366	0.424
	conventional	0.330	0.388

The picture becomes more clear if one takes into account the interaction of construction types and meaning categories. On the one hand, prepositional and non-prepositional constructions align with each other and contrast with unprepositional constructions with regard to their positive residuals' probabilities for different meaning categories. The former trend may be described as follows: L < M < C, the latter one as follows: L > M > C. The picture is the same for real and nonce verbs. On the other hand, there is a clear difference between prepositional and non-prepositional constructions themselves. Positive residuals are more likely with literal constructions of non-prepositional type than with literal constructions of prepositional type. Conversely, positive residuals are more likely with metaphorical and conventional constructions of prepositional type than with metaphorical and conventional constructions of non-prepositional type. That is, again, equally true for real and nonce verbs (Table 23).

So far, I have been able to confirm two results from the previous study: 1) prefixes that have prepositional counterparts tend to be separated from their bases by a longer pause than prefixes that have none, 2) with different compositional types of prefix-base constructions (prepositional, non-prepositional, and unprepositional), different meaning categories (literal, metaphorical, and conventional) are flagged, i.e., characterized by the longest period of silence. One thing that still requires testing is whether, when confronted with a prefix-base construction that has multiple meanings for a single category/type slot, speakers try to disambiguate those meanings by keeping one silent period as a reference level and shifting another to a lower (if possible) or higher level, depending on the preferences of a particular prefix / meaning category slot.

One way to test this is as follows. I collected all residuals for prefixes that have one meaning per category (*place*, *direction*, *time*, *quality*, *quantity*, *state*, *relation*, *limit*) in Group 1 and all residuals that have more than one meaning per category in Group 2. Overall, there were 74 samples in Group 1 and 24 samples in Group 2. Next, for each prefix-category pairing, I checked whether its residuals come from Gaussian distribution by applying Shapiro–Wilk test. The *p*-values associated with produced test statistics were collected and then compared to each other by means of *t*-test for individual observations.

My hypothesis was that for prefixes in Group 1, the residuals would tend to be normally distributed while the residuals in Group 2 would tend to come from a bimodal distribution. Then, given that the null hypothesis of Shapiro–Wilk's test is that a sample $x_1, \ldots, x_n$ comes from a normally distributed population, the *p*-values associated with test statistics should be higher in Group 1 than in Group 2. This was found to be true both for real ($t = 1.98$, $p = 0.04$) and nonce ($t = 2.58$, $p = 0.01$) verbs in my data. Thus, the third finding of the initial real-verb experiment was also confirmed.

To provide an example, let's consider two cases. Unprepositional construction [*pere*- + BASE] can encode only one sense that can be subsumed under the meaning category of *place*: 'to place something between different objects or parts of one object by means of an action identified by the base' as in *pere-sypatj* 'sprinkle with something'. On the other hand, the same construction [*pere*- + BASE] can encode two senses that belong to the meaning category of *time*: 1) 'to extend an action identified by the base to a specific, usually necessary or predetermined period of time' as in *pere-zhdatj* 'wait till the end of something' and 2) 'to cease an action identified by the base, usually after a long or intensive performing of the action' as in *pere-hotetj* 'stop wanting'. The distributions of residuals for the observations in these two prefix / meaning category slots, which were obtained in the real-verb experiment, are plotted in Figure 29. The vertical dashed lines on the plot show the means of the samples: the green one for *pere-sypatj*, the red one for *pere-zhdatj*, the blue one for *pere-hotetj*.

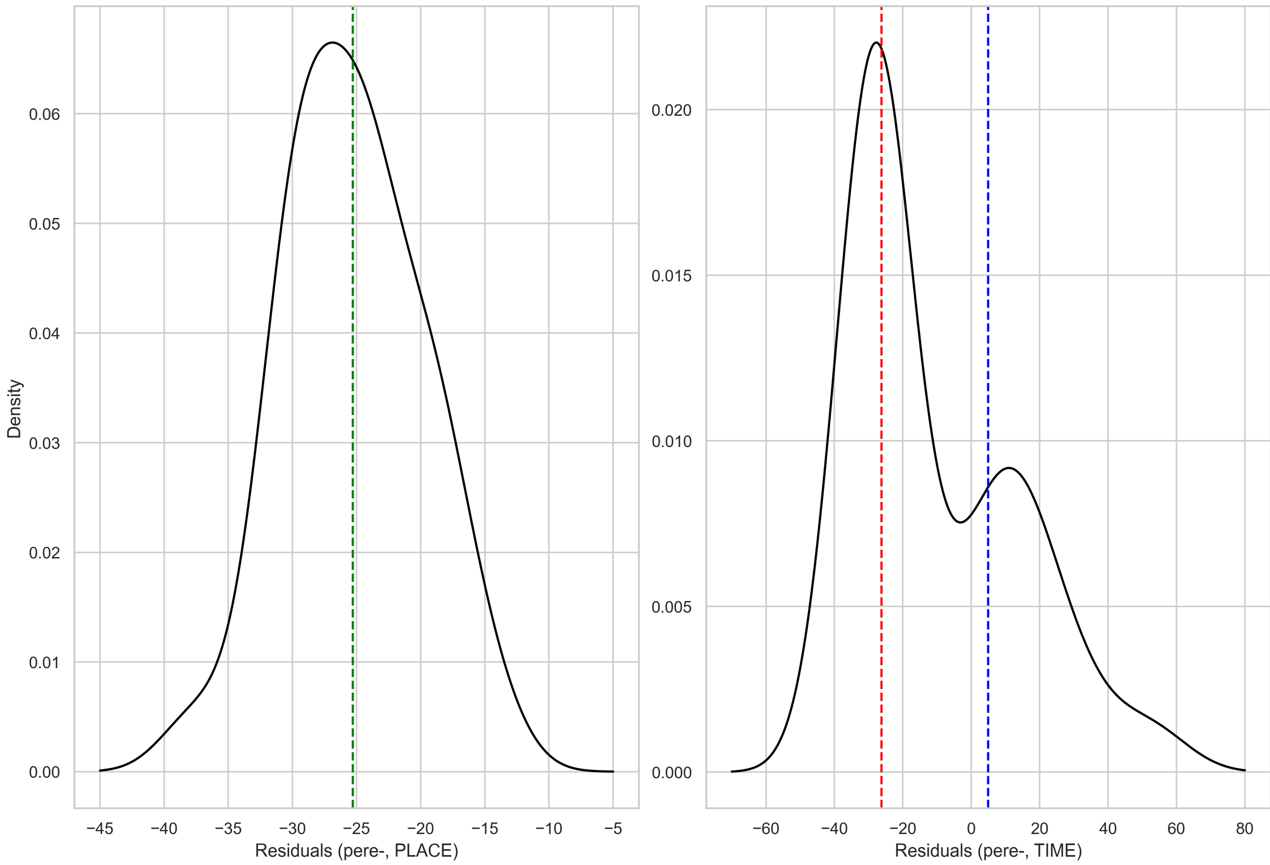

Figure 29. Distribution of residuals for the construction [*pere*- + BASE] encoding meaning categories of *place* (left-hand panel) and *time* (right-hand panel)

The distribution in the right panel of Figure 29 is clearly bimodal, in contrast with the distribution in the figure's left panel. One can conclude that participants of the experiment treated verbs *pere-zhdatj* and *pere-hotetj* not alike though the actual differences in their senses are very subtle. To better understand why this happened, it is instructive to compare the distributions of residuals obtained for [*pere*- + BASE] construction with *time* meaning in the real-verb and nonce-verb experiments (Figure 30).

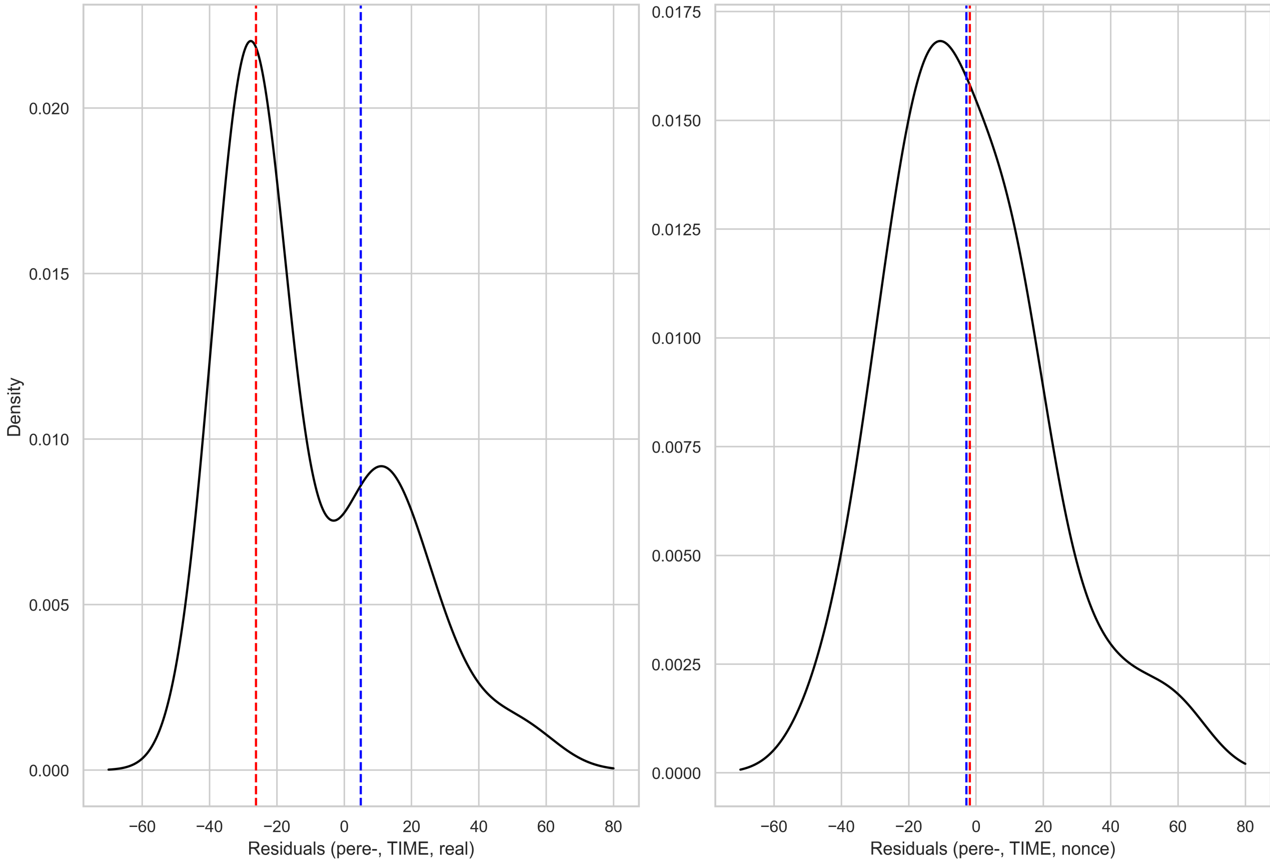

Figure 30. Distribution of residuals for the construction [*pere*- + BASE] encoding meaning category of *time* with real (left-hand panel) and nonce bases (right-hand panel)

Whereas with nonce-base stimuli, the joint distribution of residuals is centred around zero, showing that there is no specific preference in detaching prefix and base in pronunciation, with real-base stimuli, the distribution is split. What is really remarkable about this split is that neither density's peak is located at zero. This might have resulted from the fact that some of the participants, when encountering a prefix-base construction with multiple senses within a single category/type slot, tried to disambiguate them by way of desemanticizing (perfectivizing) one instance of use (*pere-zhdatj*) and flagging construction-specific meaning of another (*pere-hotetj*). Importantly, in this process, not only the length of the period of silence separating prefix and base of the latter instance became *longer* than can be accounted for by phonetic factors, but also the length of the period of silence in the former instance became *shorter* than expected.

It can be observed that among those prefixes that have one meaning per category/type slot and nevertheless reveal a bimodal distribution, literal, metaphorical and conventional semantic categories are presented in the following proportion for real verbs — 4 : 11 : 8 and in the following proportion for nonce verbs — 11 : 14 : 1. It is clear that with real verbs, participants were mostly uncertain about the most abstract category of limit. However, with nonce verbs, participants showed least unanimity when dealing with semantically most concrete constructions, those encoding literal, prepositional meanings ($\chi^2 = 14.55$, $p = 0.01$). Again, this is an anticipated result if we agree that in prefix-base verbal constructions prefixes activate general constructional meaning while bases provide necessary lexical specification.

5 How and (crucially) why to measure linguistic productivity

5.1 Introduction

The linguistic productivity of affixes has been an important topic of research for decades. It is most simply defined as '[t]he property of an affix to be used to coin new complex words' (Plag, 2018: 44). Plag, in his later work (2021), distinguishes between two possible approaches to linguistic productivity: it can be treated either categorically (qualitatively; cf. Bauer, 2001: 205) or continuously (quantitatively; cf. Bolinger, 1948: 18). Of these two approaches, as Plag remarks, the former is now mostly abandoned, and the latter is preferred. In what follows, I will adhere to the frequentist idea that totally unproductive and fully productive processes are end-points on a continuous scale, with infinitely many intermediate stages in between.

With this idea in mind, one needs a reliable way of measuring how productive a specific affix is. Many measures have been proposed in the literature so far; for example:

(i) the number of attested types (i.e., different words) with a given affix at a given point in time;

(ii) the ratio of the number of attested words with a given affix to the number of words that could, in principle, be formed with that affix (Aronoff, 1976);

(iii) the number of neologisms with a given affix at a given point in time (Plag, 2021);

(iv) Baayen's set of measures (Baayen and Lieber, 1991; Baayen, 1992, 1993, 1994, 2009; Baayen and Renouf, 1996):

> (iv.i) 'expanding productivity' (or 'hapax-conditioned degree of productivity')—the ratio of the number of hapax legomena with a given affix to the total number of hapax legomena in a given corpus;
>
> (iv.ii) 'potential productivity' (or 'category-conditioned degree of productivity')—the ratio of the number of hapax legomena with a given affix to the total number of tokens with that affix in a given corpus.

Of all these different measures, the one that has become the most well-known and widespread in the literature is the so-called 'potential productivity', which can be interpreted as follows: 'a large number of hapaxes lead to a high value of P, thus indicating a productive morphological process. Conversely, large numbers of high-frequency items lead <…> to a decrease of P, indicating low productivity' (Plag, 2021: 488). This measure has been used and continues to be referred to in multiple studies until the present day (Fernández-Domínguez, Díaz-Negrillo, and Štekauer, 2007; Plag and Baayen, 2009; Zirkel, 2010; Marzi and Ferro, 2014; Mendaza, 2015; Pierrehumbert and Granell, 2018, among others).

All this time, the idea of a 'category-conditioned degree of productivity' and the theories that grew out of it have not been seriously challenged (see, however, Bauer, 2001; Gaeta and Ricca, 2006; Pustylnikov and Schneider-Wiejowski, 2010). Gaeta and Ricca (2006) pointed out that the measure is ill-suited for the comparison of affixes with different token numbers since one will always overestimate the values of productivity for the less-frequent constructions. Unfortunately, the improvement the authors proposed (to compare the counts of hapaxes when equal numbers of tokens have been sampled for each affix) does not change the overall picture (Baayen, 2009: 905). However, as I will try to show, the method of assessing linguistic productivity by calculating the ratio of the number of hapax legomena with a given affix to the total number of tokens with that affix in a given corpus seems to be unreliable.

5.2 Hapax-based productivity measure and related issues

First, let us look at the table adopted from Plag (2021: 490), where the numbers that I am interested in are provided for some English suffixes: n_1^{aff} stands for the number of hapaxes, N^{aff} for the number of tokens and P for the 'potential productivity' measure, calculated as a simple ratio n_1^{aff} / N^{aff} (Table 24). Some problems with the 'potential productivity' measure are immediately clear. First, the number of the tokens' standard deviation (SD = 458,563) is 1,860 times higher than the number of the hapax legomena's standard deviation (SD = 246). Given the differences in both means and standard deviations, it is more reliable to compare the coefficients of variance of the two variables. The unbiased estimators (Sokal and Rohlf, 1995: 58) for a given sample size n = 7 will be CV(Tokens) = (1+1/28) * (458,563 / 283,545) = 4.45; CV(Hapaxes) = (1+1/28) * (246 / 403) = 1.67. Thus, the number of the tokens' coefficient of variation is 2.7 times higher than the number of the hapax legomena's coefficient of variation.

Why is this important? Taking this fact into account, one might reason as follows: the number of hapax legomena tends to be uniformly distributed, while the number of tokens varies greatly. With a fixed numerator and varying denominator, it is clear that the number of tokens will influence the measure of productivity greater than the number of hapax legomena, therefore affixes with the greater number of tokens will always be assessed as less productive. As stated by Hayya, Armstrong, and Gressis (1975), for a ratio of two approximately normally distributed random variables W = X/Y, as the coefficient of variation CV(X) approaches zero, X approaches a constant, and W approaches a normal variable that is proportional to 1/Y.

Table 24. Plag's data with productivity measures (some columns are omitted)

affix	N^{af} (tokens)	n_1^{aff} (hapaxes)	*P* (productivity)
-ion	1,369,116	524	0.00038
-ish	7,745	262	0.0338
-ist	98,823	354	0.0036
-ity	371,747	341	0.00092
-less	28,340	272	0.0096
-ness	106,957	943	0.0088
-wise	2,091	128	0.061

This is not the only peculiarity with the data in Table 24. It can be shown that (i) the number of hapax legomena is strongly positively correlated with the total number of tokens (ρ = 0.82, p = 0.02), so the more tokens, the more hapax legomena for a given affix; (ii) the number of tokens is significantly negatively correlated with productivity measure (ρ = -0.96, p < 0.001), which seems to support Baayen's idea; (iii) however, the number of hapax legomena is also negatively correlated with productivity measure (ρ = -0.75, p = 0.05), which looks counterintuitive at first glance.

These three observed correlations define a consistent picture if one casts the problem into the framework of causal models. A causal model has the same form as a probabilistic Bayesian network: it is a directed acyclic graph (DAG) over some random variables. The model asserts that each variable is governed by a causal mechanism that (stochastically) determines its value based on the values of its parents (Koller and Friedman, 2009: 1014). Now, let us consider Model A in Figure 31 (left-hand panel). The number of hapaxes may exert some influence on the productivity measure, but the number of tokens may also do so; besides, the number of hapaxes most likely depends on the number of tokens. This is problematic because by measuring the numbers of hapaxes and tokens separately, one does not consider the so-called 'back-door path' between the variables Hapaxes and Productivity introduced by the variable Tokens (red dotted line in Figure 31). More formally, for any two variables X and Y in the model, a back-door path is any path from X to Y that starts with an arrow pointing at X. To correctly identify causal mechanisms, one needs to deconfound X and Y by blocking every back-door path (because such paths allow spurious, non-causal correlation between X and Y) (Pearl, 2009; Pearl, Glymour, and Jewell, 2016).

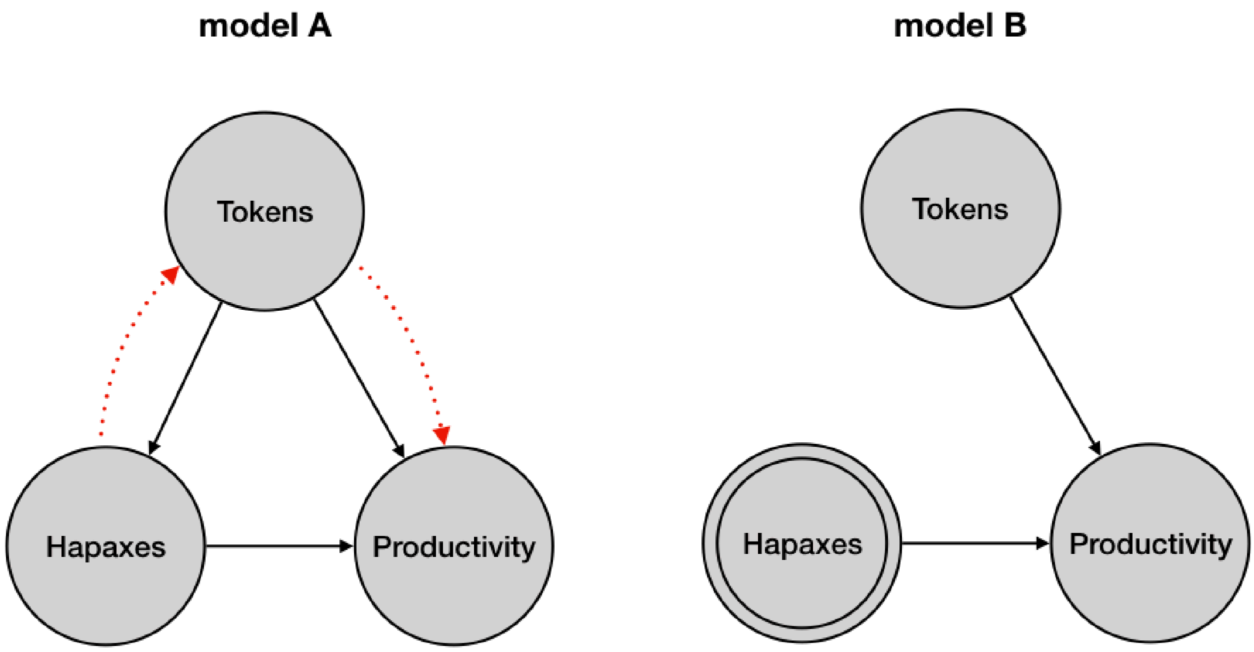

Figure 31. Original (A) and mutilated (B) networks of 'potential productivity' measure

This problem may be solved within the framework of causal networks by the so-called *do*-operator. This implies creating a mutilated network by erasing the edge coming from Tokens to Hapaxes and answering not the query of the form P(Productivity | Hapaxes) but rather P(Productivity | *do*[Hapaxes]). Basically, we need to isolate the component of the correlation that is due to the causal effect of one variable on another. It has been shown (Koller and Friedman, 2009) that for models such as B (Figure 31, right-hand side), queries of the form P(Productivity | *do*[Hapaxes]) are analogous to queries

$$\sum_{\text{Tokens}} P(\text{Productivity} \mid \text{Hapaxes}, \text{Tokens}) P(\text{Tokens}),$$

the answers to which can be inferred from the data.

To perform inference, I used Plag's data in binarised form. All continuous values were replaced with either 0 or 1, depending on whether they fell below or above the mean, column-wise. Given these data and the mutilated network in Model B, it is now straightforward to calculate the probabilities of interest; namely, (1) the probability of productivity measure being high if we intervene and make the number of hapaxes low, and (2) the probability of productivity measure being low if we intervene and make the number of hapaxes low.

The first probability was estimated as P(Productivity = high | Hapaxes = low, Tokens = high) * P(Tokens = high) + P(Productivity = high | Hapaxes = low, Tokens = low) * P(Tokens = low) = 0 * 0.572 + 1 * 0.428 = 0.428. The second probability is, obviously, 1 - 0.428 = 0.572. It is clear that the obtained probabilities are equal to the probabilities of the number of tokens being low and the number of tokens being high, respectively. This means that the number of hapaxes plays no role whatsoever in the productivity measure, which is thus totally dependent on the number of tokens, importantly, in a reciprocal manner: the lower the number of tokens, the higher the productivity value, and vice versa.

Plag's exemplary data contain only seven affixes. However, essentially the same features may be observed with a much larger dataset from Hay and Baayen's paper (2002: 233–235), in which they presented their calculations for 80 English prefixes and suffixes. The only problem with these data is that, while providing the numbers of hapaxes and productivity measures, the authors do not report the numbers of tokens. Nevertheless, one can easily reconstruct those numbers by drawing on two values that Hay and Baayen provide: (1) token-P—'the summed frequency of the words which fall above the parsing line' and (2) token-PR—'the proportion of tokens which fall above the parsing line'. Applying the simple formula *tokens = tokens-P / tokens-PR*, one gets the estimated total number of tokens for each affix (Appendix 2).

In these data, I again found a strong negative correlation between Tokens and Productivity variables (ρ = -0.73, $p < 0.001$), a moderate positive correlation between Hapaxes and Tokens (ρ = 0.45, $p < 0.001$), and an insignificant correlation between Hapaxes and Productivity (ρ = 0.17, p = 0.11).

In essence, all of the above suggests that the causal model, implied by Baayen's way of measuring linguistic productivity (Figure 32, left-hand panel), is inaccurate and should be replaced with the model visualised in Figure 32 (right-hand panel). I compared these models using the Bayesian scoring criterion (Neapolitan and Jiang, 2007: 445–447) in order to determine and select the DAG pattern with maximum probability conditional on the data. It was found that P (implied model | data) = 0.44 and P (real model | data) = 0.56, which means that, given the data at hand, it is more likely that when we condition on Tokens, Hapaxes and Productivity become independent.

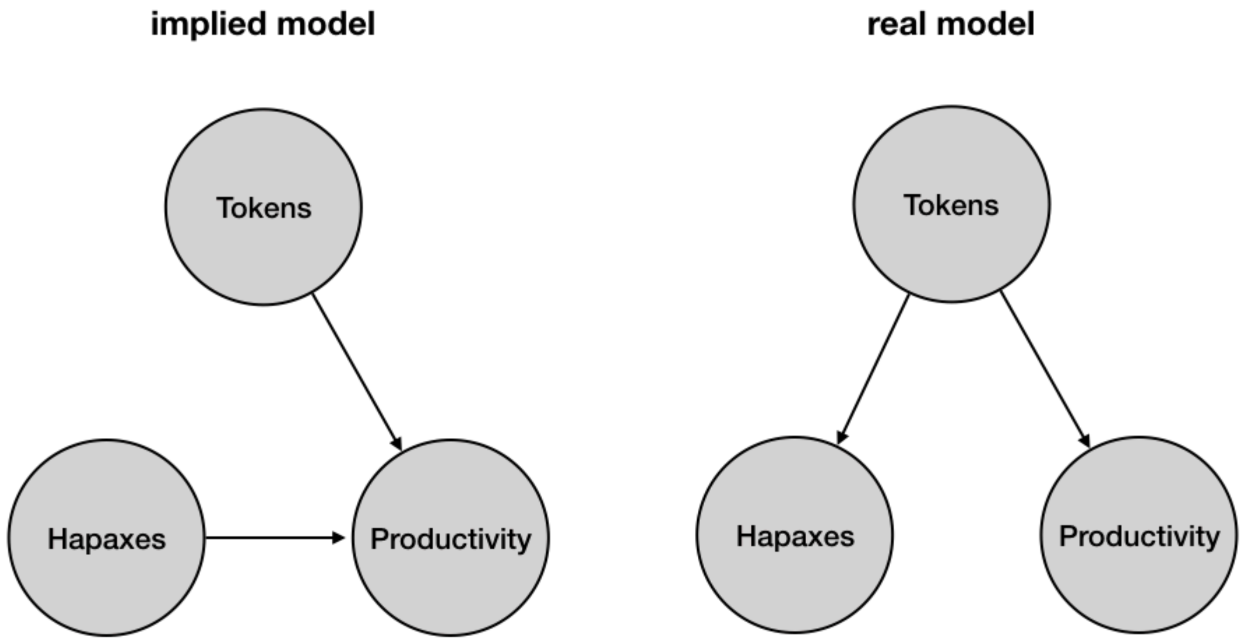

Figure 32. Two probabilistic models of ‘potential productivity’ measure

Given the model and Baayen’s formula for calculating productivity, the nature of the relation between the productivity measure and the number of tokens becomes clear. Since the number of tokens is placed in the denominator, the measure of productivity will always be lower for affixes with high token frequency and higher for affixes with low token frequency.

The fact that the ‘hapax-based’ productivity measure estimates the reciprocal of the total number of tokens for each affix can be verified by means of Markov chain Monte Carlo hierarchical modelling. Drawing on the data from Plag’s work, let us first assume that the number of tokens is treated as given and the number of hapaxes as unknown, something that I want to simulate.

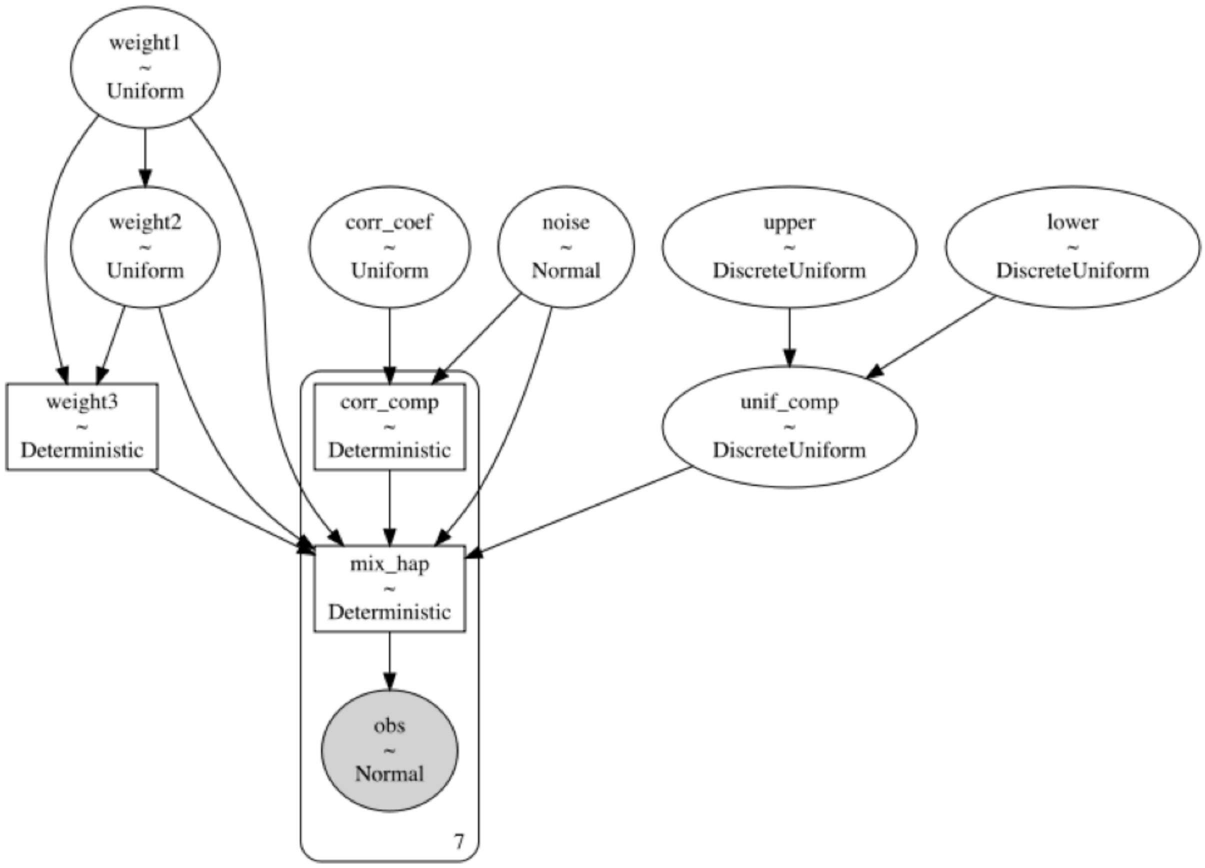

Figure 33. Hierarchical model for simulating the number of hapaxes in Plag's (2021) data

Let us also assume that we believe the population of hapaxes for this set of affixes to be a mixture (labelled *mix_hap* in Figure 33) of three different components: (1) the first one is a uniform distribution (labelled *unif_comp*) that can take any value within certain bounds (labelled *upper* and *lower*) with equal likelihood, (2) the second one is a distribution whose parameters are defined by its positive correlation with the total number of tokens (labelled *corr_coeff*), and finally, (3) the third one is a Gaussian distribution of random noise that can account for any affix-specific preferences in the number of hapaxes (labelled *noise*). Components (2) and (3) together constitute one mixture distribution labelled *corr_comp*, which accounts for the overall non-uniformity of hapaxes' values. To identify the exact proportion of each distribution in *mix_hap*, I programmed the simulated population as a weighted linear combination of all three components, with *weight1* attached to the first one, *weight2* to the second, and *weight3* to the third.

As for the hyperparameters, the prior on the lower bound of the uniform distribution was set uniformly in the range [0, 500] and the prior on the upper bound in the range [500, 1000]. The prior on the correlation coefficient was set uniformly in the range [0.0, 0.7]. The prior on the random noise was set to $\mu = 0$, $\sigma = 100{,}000$. The priors for the weights were selected in the following manner: *weight1* — uniformly distributed in the range [0, 1], *weight2* — uniformly distributed in the range [0, 1-*weight1*], and *weight3* — deterministically equal to 1-*weight1*-*weight2*.

Intuitively, if one believes that the number of hapaxes reflects in any way the linguistic productivity of an affix and is independent of its total number of tokens, one would expect the value of *weight3* to be the highest. However, after sampling three chains for 1,000 burn-in and 2,000 draw iterations, I obtained the following estimates: *weight1* = 0.973, *weight2* = 0.015, and *weight3* = 0.012. This clearly shows that the first component dominates the mixture distribution; that is, the number of hapaxes is uniformly distributed, approaching a constant in Baayen's equation.

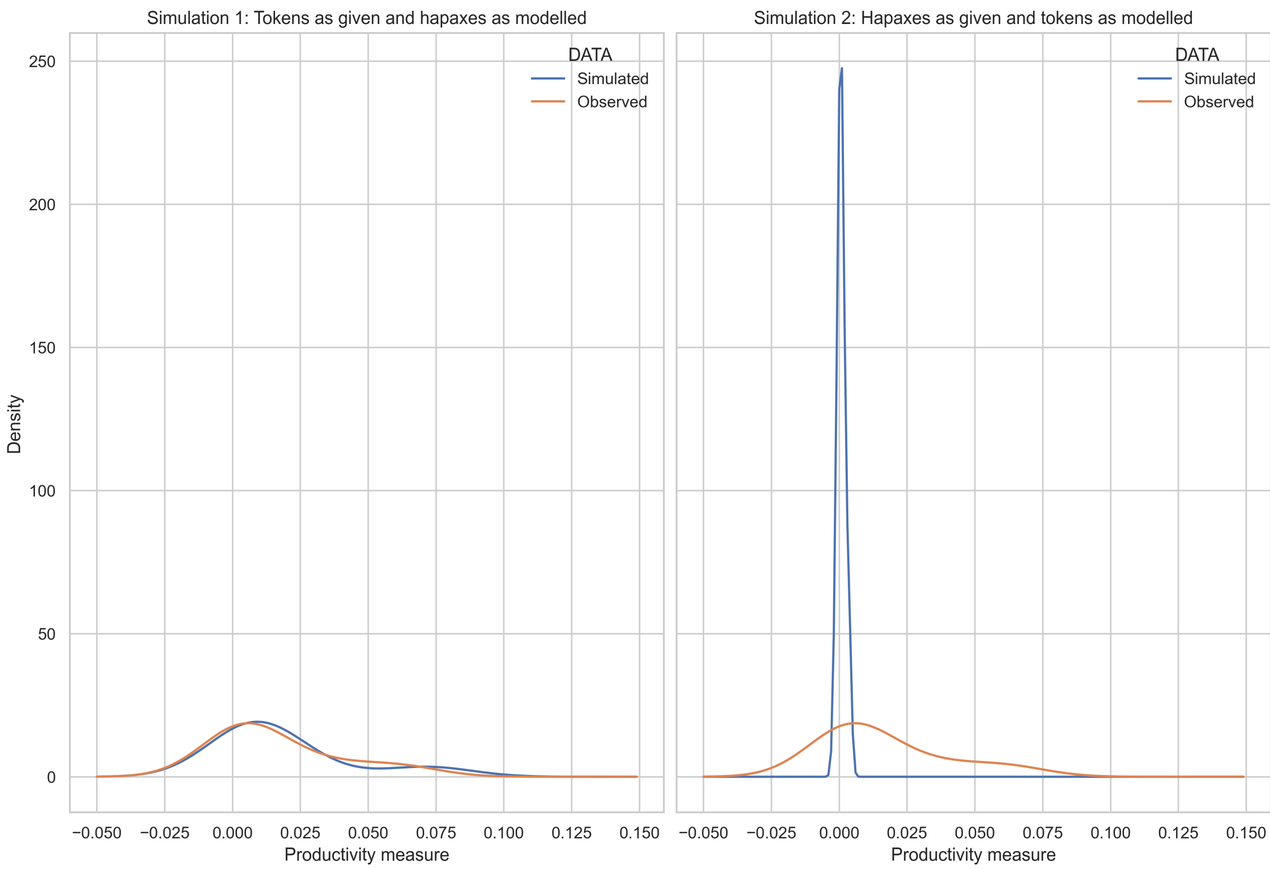

Figure 34. Density of observed productivity measures from Plag's example plotted against the densities of productivity measures obtained from two MCMC simulations

But how good is this model? To check it, I sampled from the posterior 6,000 linguistic productivity values, calculated according to Baayen's formula. The densities[4] of the simulated and observed values are plotted against each other in Figure 34 (left-hand panel) and constitute a very close match. Interestingly, although the model was trained on Plag's data, it approximates the distribution of productivity values from Hay and Baayen's article (2002) just as well.

The only affix whose value was found to be clearly outside the expected range is suffix *-like* with its 270 hapaxes for 709 tokens. Obviously, *-like* is an outlier in the dataset; its productivity value is estimated at 0.381, while the respective value of the second most productive affix, *super-*, is 0.084 (i.e., 4.5 times smaller). One possible explanation of this fact is the special status of *-like*, among other affixes: it has been argued in the literature that the semantic and structural properties of formations such as *rock-like* can be 'straightforwardly accounted for without postulating another derivational suffix under the simplest assumption that they are compound adjectives' (Dalton-Puffer and Plag, 2000: 238).

It is enlightening to compare the obtained simulation results with results from the model of the same architecture but with different parametrisation, where the number of hapaxes is treated as given and the number of tokens as unknown. The tweaks made to the priors include (1) the prior on the lower bound of the uniform distribution in the range [0, 3000], (2) the prior on the upper bound in the range [3000, 1500000], and (3) the prior on the random noise with $\mu = 0$ and $\sigma = 1{,}000{,}000$; other parameters remained unchanged. First, the estimates of weights now look completely different: *weight1* = 0.503, *weight2* = 0.249, and *weight3* = 0.248. This suggests that, unlike the case with the number of hapaxes, the number of tokens is far less uniformly distributed. Second, an attempt to predict the measures of linguistic productivity based on the samples from the posterior essentially fails, as can be observed in Figure 34 (right-hand panel).

Calculating the productivity measures of different affixes is not a matter of linguistic bookkeeping. Rather, we are interested in these measures because we want to understand how morphological productivity works (i.e., how derivational patterns spread). Of special importance here is the question of how the type and token frequency of linguistic items contribute to derivational patterns' self-propagation. From this standpoint, having a measure of linguistic productivity heavily biased by the number of tokens is a somewhat unwelcome premise.

I would like to suggest a potentially more accurate way of measuring affixes' productivity, a way that would objectively assess '[t]he property of an affix to be used to coin new complex

[4] The estimated density curves in Figure 34 extend to negative values that do not make sense in our case. This happens because the smoothing algorithm uses a Gaussian kernel (Waskom, 2021).

words' (Plag, 2018: 44) and would not be directly dependent upon tokens' frequency. Specifically, I suggest that linguistic productivity may be viewed as the probability of an affix to combine with a random base. The advantages of this approach include the following: (1) token frequency does not dominate the productivity measure but naturally influences the sampling of bases; (2) we are not just counting attested word types with an affix but rather simulating the construction of these types and then checking whether they are attested in the corpus; and (3) a corpus-based approach and randomised design assure that true neologisms and words coined long ago have equal chances to be selected.

Currently, the procedure has only been tested on prefixes, but its basic principles readily extend to suffixes.

5.3 Introducing a new algorithm for measuring linguistic productivity

The process of obtaining a linguistic productivity measure for a specific prefix consists of two parts. The first part runs as follows:

1. For each prefix P_i:
    - 1.1. A random sample of 100 content words $V = \{V_1, \ldots, V_{100}\}$ is drawn from a corpus of modern language.
        - 1.1.1. Each word from the sample is automatically checked for whether it contains any prefix, and if so, it is stripped of it, which results in the set of bases $B = \{B_1, \ldots, B_{100}\}$.
        - 1.1.2. The prefix P_i is automatically combined with each base $B_j \in B$, which results in the set of P_i-prefixed words $PB = \{PB_1, \ldots, PB_{100}\}$
        - 1.1.3. The frequency of each P_i-prefixed word $PB_j \in PB$ is checked in the same corpus and recorded, which results in the set of values $F = \{F_1, \ldots, F_{100}\}$.

Upon completing the first part of the algorithm, one could get a naive estimation of the linguistic productivity measure that is based on the observed data. For example, one could obtain from F a subset of values FNZ, such that each value $FNZ_j \in F$, $FNZ_j \neq 0$, $j \leq 100$, and then calculate productivity of the prefix P_i as simple relative frequency n(FNZ) / n(F). However, I am interested in the probabilistic assessment of linguistic productivity. One way to estimate it is to try to predict what this value will be for the 101st base; that is, for the first base coming out of the sample. This constitutes the second part of the algorithm.

To implement it, I constructed a dynamic Bayesian network consisting of (1) a two-node DAG, where node X_0 represents an observation of a prefix-base combination at time *t*, and node X_1 represents an observation of a prefix-base combination at time $t + 1$, as well as (2) a joint probability distribution of the following form (Neapolitan and Jiang, 2007):

$$P_0(\mathrm{x}[0])\prod_{\mathrm{t}=0}^{\mathrm{T}-1} P_{\rightarrow}(\mathrm{x}[\mathrm{t}+1]\mathrm{x}[\mathrm{t}])$$

The cardinality of each node is equal to three, where 0 stands for no occurrence of a particular prefix-base combination in the corpus (indicating that the prefix does not combine with this base); 1 stands for a low-frequency occurrence (a prefix-base combination was considered of low frequency if its number of hits in the corpus was lower than the 0.5 quantile of the previously obtained results in the sample); and 2 stands for a high-frequency occurrence (number of hits in the corpus greater than or equal to the 0.5 quantile of the previously obtained results).

The first part of the joint probability distribution, representing the initial state, was parametrised as follows: $P(X_0 = 0) = 0.4$, $P(X_0 = 1) = 0.4$, $P(X_0 = 2) = 0.2$. The second part, which contains transition probabilities for estimating unknown states given some known observations, was parametrised as follows (Table 25):

Table 25. Transition probabilities for the dynamic Bayesian network

	$X_0 = 0$	$X_0 = 1$	$X_0 = 2$
$X_1 = 0$	0.7	0.2	0.1
$X_1 = 1$	0.2	0.7	0.1
$X_1 = 2$	0.1	0.2	0.7

These probabilities mirror my prior beliefs about the distribution of lexical items; however, they were used only once to predict the outcome of the second prefix-base combination in the sample. At each subsequent step, both initial and transition probabilities were updated based on the observed evidence. Thus, the value of $P(X_1[t+1] = 1) + P(X_1[t+1] = 2)$ (or, equivalently, $1 - P(X_1[t+1] = 0)$) evaluated at time $t = 100$ constitutes the true value of linguistic productivity.

5.4 New productivity measure and its insights

Using the algorithm described above and the English internet corpus from 2018 provided by Sketch Engine (*ententen18_tt31*; 21,926,740,748 words), I obtained productivity values for all English prefixes in Hay and Baayen's (2002) data. Hyphenated and non-hyphenated variants were calculated separately and then summed up. It is instructive to compare, for all prefixes, the dynamics of the probabilities $P(X_1[t+1] = 0 \mid X_0[t])$ (blue lines of all subplots in Figure 35), $P(X_1[t+1] = 1 \mid X_0[t])$ (orange lines), and $P(X_1[t+1] = 2 \mid X_0[t])$ (green lines), for $t \in \{1, \ldots, 100\}$.

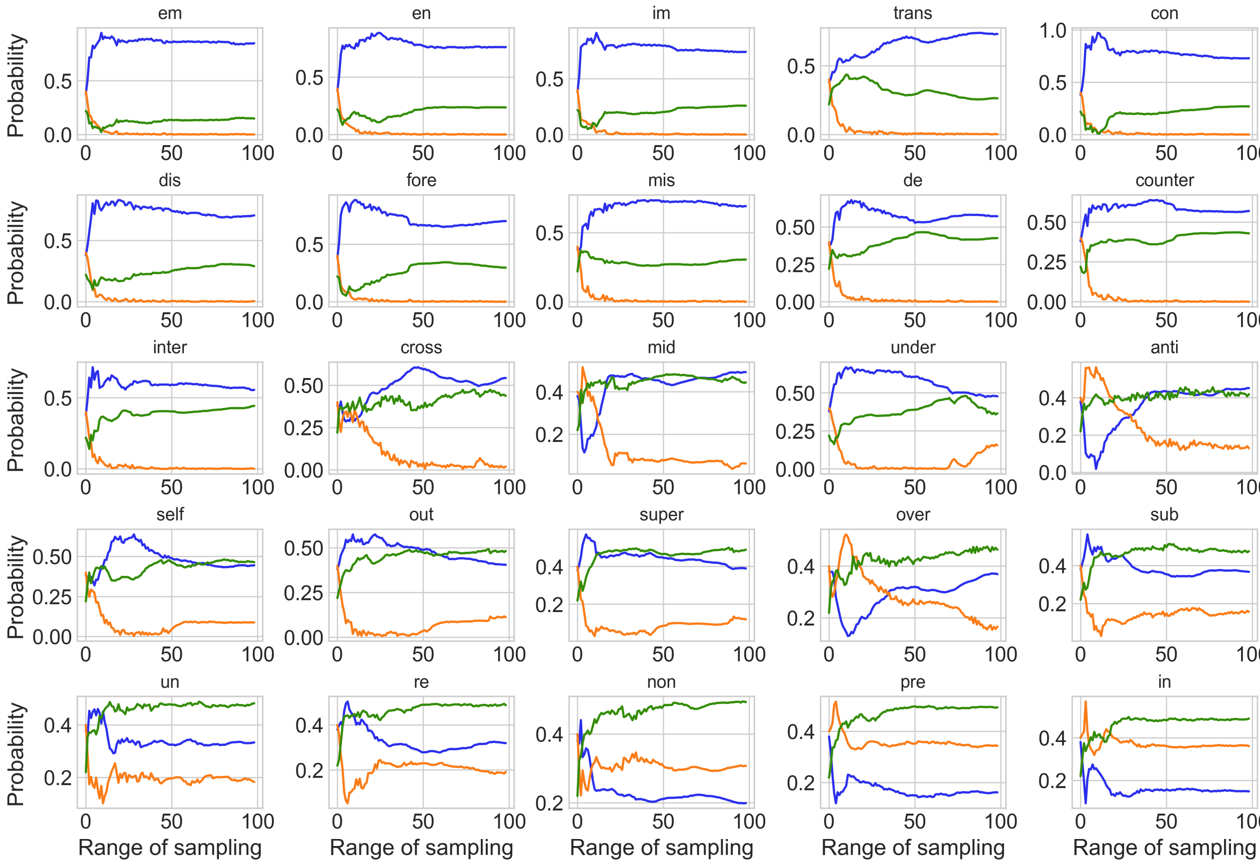

Figure 35. Probabilities of different values of $X_1[t+1]$ evaluated at times $t \in \{1, \ldots, 100\}$ for the English prefixes in Hay and Baayen's (2002) data

First, it is clear that all three probabilities, after initial uncertainty, converge towards the end of the range of sampling to some stationary distribution, as the variance of their values approaches zero. Thus, the aforementioned way of calculating linguistic productivity as the probability of a given prefix to combine with the first base outside of a sample of 100 random bases seems justified.

Second, if one takes into account that the prefixes in Figure 35 are arranged, left-to-right, top-to-bottom, in order of ascending productivity measure, one interesting phenomenon becomes evident that sheds some new light on the long-standing debate on what is more important for linguistic productivity: the number of types or number of tokens. Concerning the final evaluations of the probabilities, all prefixes can be subdivided into three large groups, depending on the hierarchical order of these evaluations (Table 26). The first group encompasses prefixes from *em-* to *anti-*, and the probabilities here are arranged in the following hierarchy: $P(X_1[t+1] = 0 \mid X_0[t]) > P(X_1[t+1] = 2 \mid X_0[t]) > P(X_1[t+1] = 1 \mid X_0[t])$. In the second group, one finds prefixes from *self-* to *re-*, with the probabilities arranged in this way: $P(X_1[t+1] = 2 \mid X_0[t]) > P(X_1[t+1] = 0 \mid X_0[t]) > P(X_1[t+1] = 1 \mid X_0[t])$. Finally, the prefixes *non-*, *pre-*, and *in-* belong to the last group: $P(X_1[t+1] = 2 \mid X_0[t]) > P(X_1[t+1] = 1 \mid X_0[t]) > P(X_1[t+1] = 0 \mid X_0[t])$.

What is more interesting is that these categorical differences emerge as manifestations of an inherently gradient structure. From Table 26, where the prefixes are arranged in order of ascending productivity, it can be seen that for the first group (*0_2_1*), the differences in probabilities $P(X_1[t+1] = 0 \mid X_0[t])$ and $P(X_1[t+1] = 2 \mid X_0[t])$ continuously decrease, while the differences in probabilities $P(X_1[t+1] = 2 \mid X_0[t])$ and $P(X_1[t+1] = 1 \mid X_0[t])$ continuously increase (with some minor fluctuations). Visually, in the succession of subplots in Figure 35, this process can be described in terms of a green curve climbing higher and higher, with the other two curves held constant until finally, it changes places with a blue one, thus opening up the second group of prefixes.

For this second group (*2_0_1*), a similar mechanism of change can be observed, though with different contrasts. The differences in probabilities $P(X_1[t+1] = 2 \mid X_0[t])$ and $P(X_1[t+1] = 0 \mid X_0[t])$ become bigger, while the differences in probabilities $P(X_1[t+1] = 0 \mid X_0[t])$ and $P(X_1[t+1] = 1 \mid X_0[t])$ become smaller. Again, across the respective subplots of Figure 35, this process can be roughly described as that of a blue curve falling down and swapping near the bottom with an orange one.

The third group (*2_1_0*), though smallest, is of the same gradient nature. The gap between probabilities $P(X_1[t+1] = 2 \mid X_0[t])$ and $P(X_1[t+1] = 1 \mid X_0[t])$ successively narrows, while the gap between probabilities $P(X_1[t+1] = 1 \mid X_0[t])$ and $P(X_1[t+1] = 0 \mid X_0[t])$ widens. For simplicity, one can visualise an orange curve in Figure 35 approaching a green one at the top of the plot.

Table 26. English prefixes' productivity measures arranged in ascending order

group	prefix	product.	$X_1 = 0$	$X_1 = 1$	$X_1 = 2$	contr._1	diff._1	contr._2	diff._2
0_2_1	*em-*	0.150	0.850	0.001	0.150	0_2	0.700	2_1	0.149
0_2_1	*en-*	0.237	0.763	0.001	0.236	0_2	0.526	2_1	0.236
0_2_1	*im-*	0.259	0.741	0.000	0.258	0_2	0.483	2_1	0.258
0_2_1	*trans-*	0.268	0.732	0.004	0.264	0_2	0.467	2_1	0.261
0_2_1	*con-*	0.270	0.730	0.001	0.269	0_2	0.461	2_1	0.269
0_2_1	*dis-*	0.296	0.704	0.005	0.291	0_2	0.413	2_1	0.286
0_2_1	*fore-*	0.298	0.702	0.001	0.297	0_2	0.406	2_1	0.296
0_2_1	*mis-*	0.307	0.693	0.000	0.306	0_2	0.387	2_1	0.306
0_2_1	*de-*	0.427	0.573	0.001	0.426	0_2	0.147	2_1	0.426
0_2_1	*counter-*	0.430	0.570	0.000	0.430	0_2	0.140	2_1	0.430
0_2_1	*inter-*	0.444	0.556	0.001	0.443	0_2	0.113	2_1	0.442
0_2_1	*cross-*	0.457	0.543	0.019	0.438	0_2	0.105	2_1	0.419
0_2_1	*mid-*	0.507	0.493	0.062	0.445	0_2	0.048	2_1	0.383
0_2_1	*under-*	0.522	0.478	0.156	0.366	0_2	0.112	2_1	0.210
0_2_1	*anti-*	0.548	0.452	0.129	0.419	0_2	0.033	2_1	0.290
2_0_1	*self-*	0.554	0.446	0.089	0.465	2_0	0.019	0_1	0.357
2_0_1	*out-*	0.595	0.405	0.112	0.483	2_0	0.078	0_1	0.293
2_0_1	*super-*	0.610	0.390	0.121	0.489	2_0	0.099	0_1	0.269
2_0_1	*over-*	0.631	0.369	0.167	0.464	2_0	0.095	0_1	0.202
2_0_1	*sub-*	0.633	0.367	0.156	0.476	2_0	0.109	0_1	0.211
2_0_1	*un-*	0.666	0.334	0.183	0.483	2_0	0.149	0_1	0.151
2_0_1	*re-*	0.680	0.320	0.192	0.488	2_0	0.169	0_1	0.127
2_1_0	*non-*	0.801	0.199	0.308	0.493	2_1	0.185	1_0	0.108
2_1_0	*pre-*	0.838	0.162	0.344	0.494	2_1	0.150	1_0	0.182
2_1_0	*in-*	0.851	0.149	0.361	0.490	2_1	0.129	1_0	0.213

To understand what all of the above tells us about the relation of linguistic productivity to the number of types and tokens, we need to recall exactly what each type of probabilistic hierarchy signifies. The arrangement of probabilities for each group may be interpreted as follows. Group *0_2_1*: a small number of types with a few occasional high-frequency tokens. Group *2_0_1*: a substantial number of high-frequency tokens but a very limited number of types overall. Group *2_1_0*: many types and many tokens. The gradient nature of the transitions from group to group provides some evidence as to how the linguistic productivity of affixes grows. Burgeoning linguistic productivity manifests in an increasing number of types. However, this process unfolds in two stages: first comes the increase in high-frequency items, and only then follows the increase in low-frequency items.

In order to make sure that these patterns are not language-specific, I used the Russian internet corpus from 2011 provided by Sketch Engine (*rutenten11_8*; 14,553,856,113 words) to repeat the whole procedure of measuring linguistic productivity with 27 Russian verbal prefixes. The results are provided in Figure 36 and Table 27. The similarity between the two languages is quite remarkable: one can easily identify in the Russian data the same tripartite division of affixes with the same fuzzy boundaries between the groups as in English.

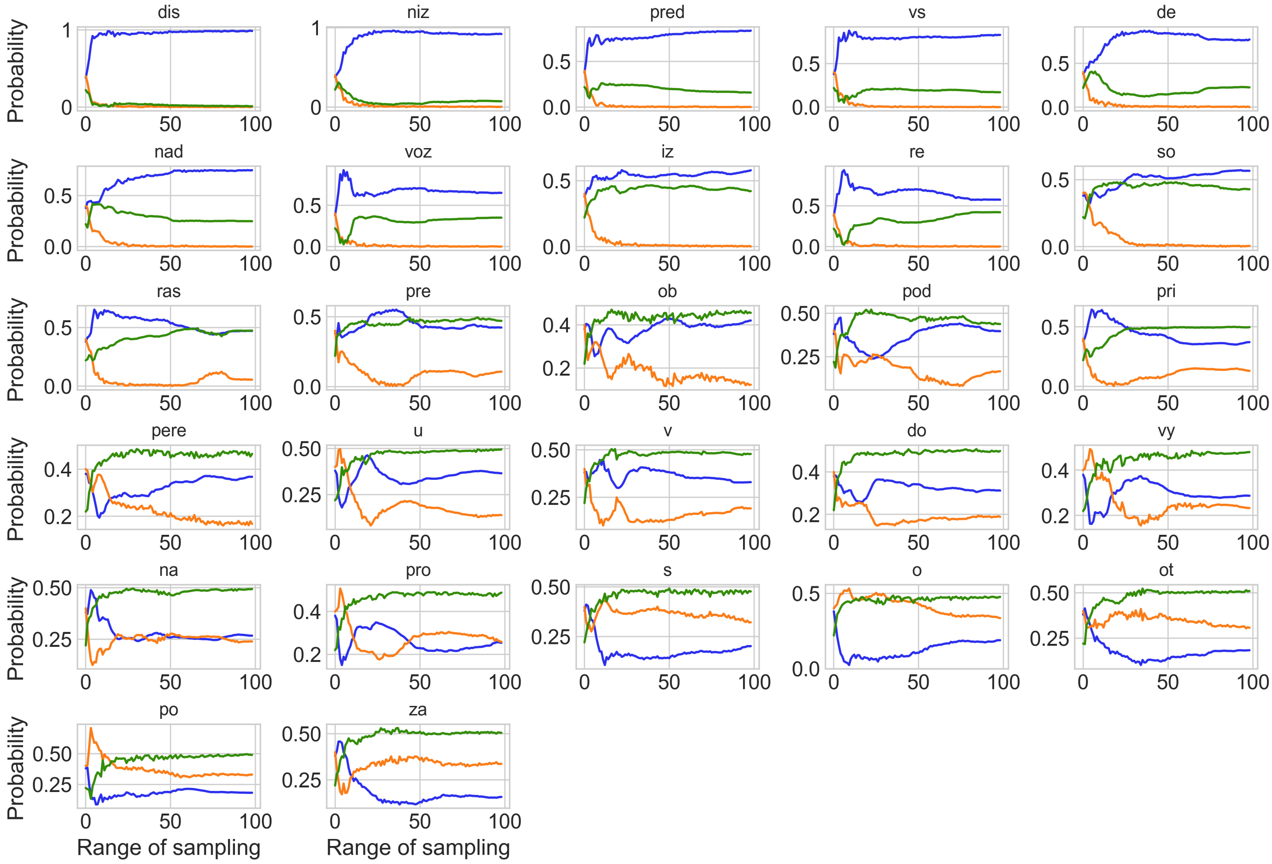

Figure 36. Probabilities of different values of $X_1[t+1]$ evaluated at times $t \in \{1, \ldots, 100\}$ for the Russian prefixes

Table 27. Russian prefixes' productivity measures arranged in ascending order

group	prefix	product.	$X_1 = 0$	$X_1 = 1$	$X_1 = 2$	contr._1	diff._1	contr._2	diff._2
0_2_1	*dis*	0.015	0.985	0.002	0.013	0_2	0.972	2_1	0.011
0_2_1	*niz*	0.080	0.920	0.006	0.074	0_2	0.845	2_1	0.069
0_2_1	*pred*	0.161	0.839	0.001	0.160	0_2	0.679	2_1	0.158
0_2_1	*vs*	0.171	0.829	0.000	0.171	0_2	0.658	2_1	0.171
0_2_1	*de*	0.227	0.773	0.001	0.226	0_2	0.547	2_1	0.225
0_2_1	*nad*	0.250	0.750	0.001	0.249	0_2	0.501	2_1	0.249
0_2_1	*voz*	0.350	0.650	0.001	0.350	0_2	0.300	2_1	0.349
0_2_1	*iz*	0.422	0.578	0.001	0.421	0_2	0.157	2_1	0.420
0_2_1	*re*	0.424	0.576	0.003	0.421	0_2	0.155	2_1	0.419
0_2_1	*so*	0.435	0.565	0.007	0.428	0_2	0.137	2_1	0.421
2_0_1	*ras*	0.529	0.471	0.055	0.474	2_0	0.003	0_1	0.416
2_0_1	*pre*	0.577	0.423	0.107	0.470	2_0	0.048	0_1	0.315
2_0_1	*ob*	0.579	0.421	0.123	0.456	2_0	0.036	0_1	0.297
2_0_1	*pod*	0.604	0.396	0.166	0.438	2_0	0.041	0_1	0.231
2_0_1	*pri*	0.628	0.372	0.131	0.497	2_0	0.125	0_1	0.241
2_0_1	*pere*	0.633	0.367	0.168	0.465	2_0	0.097	0_1	0.200
2_0_1	*u*	0.634	0.366	0.139	0.495	2_0	0.129	0_1	0.227
2_0_1	*v*	0.670	0.330	0.192	0.478	2_0	0.148	0_1	0.138
2_0_1	*do*	0.688	0.312	0.189	0.499	2_0	0.187	0_1	0.124
2_0_1	*vy*	0.713	0.287	0.233	0.480	2_0	0.193	0_1	0.054
2_0_1	*na*	0.732	0.268	0.239	0.493	2_0	0.225	0_1	0.028
2_1_0	*pro*	0.746	0.254	0.258	0.487	2_1	0.229	1_0	0.004
2_1_0	*s*	0.799	0.201	0.322	0.477	2_1	0.155	1_0	0.120
2_1_0	*o*	0.811	0.189	0.335	0.476	2_1	0.141	1_0	0.147
2_1_0	*ot*	0.817	0.183	0.307	0.510	2_1	0.203	1_0	0.124
2_1_0	*po*	0.818	0.182	0.329	0.489	2_1	0.160	1_0	0.148
2_1_0	*za*	0.842	0.158	0.337	0.505	2_1	0.168	1_0	0.179

5.5 Comparing two productivity measures

Let us now compare my results (blue line in Figure 37) plotted against Hay and Baayen's productivity measures, multiplied by 10 for comparability (orange line). Substantial differences between these two types of assessment are observable, and some inconsistencies are striking. I will leave it to everyone's judgement to decide, for example, how justified it is to place *non-*, based on Baayen's measure, at a so much higher productivity level compared to *un-* (0.07 vs. 0.005, i.e., 14 times greater). My data show that *un-* is extremely productive in modern English; it can combine with almost any base, whether verbal or nominal; to name some examples: (i) *It's such an **un-Michigan** thing to do*; (ii) *Fresh off the trail-blazing heels of SXSW comes Skillshare, the '**un-university**' of online universities*; (iii) *But you can't **un-shoot** a person*.

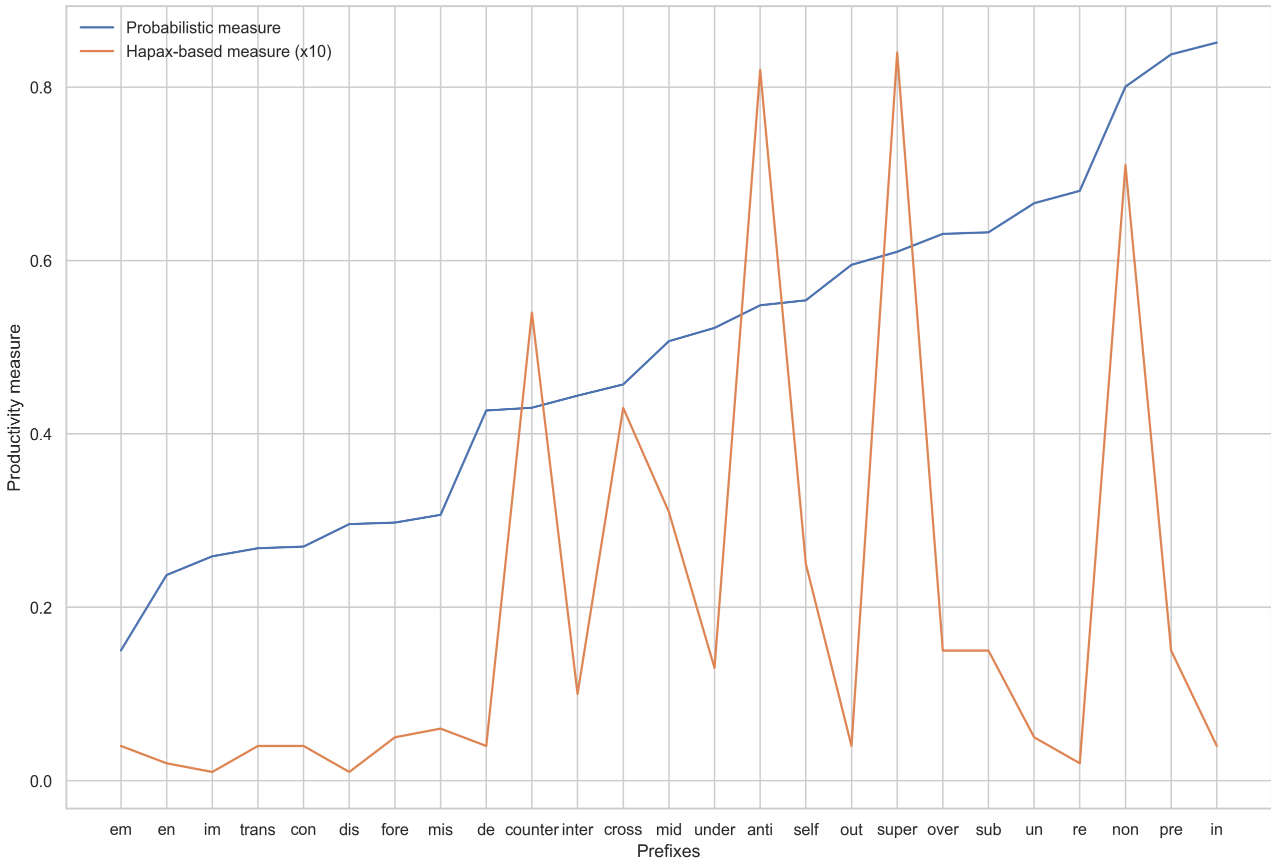

Figure 37. English prefixes' productivity measures calculated as probabilities of combining with a random base and as ratios of the number of hapaxes to the number of tokens

The top status of *in-*, based on my measure, may look surprising. However, one should take into account that there are two homonymous prefixes: a negative *in-*, as in *inaccurate*, and a prepositional *in-*, as in the following examples: (i) *This precludes **in-contact** operation*; (ii) <…>

*this was a great **inlook** at movie industry*; (iii) <…> ***inmouth** or eye exposure occurs* <…>; (iv) *Practice the wrong technique and it will be locked in and be hard to **in-do**;* (v) *My favorite **in-major** class was Software Engineering.* While the former *in-* is only of very limited productivity in modern English, the latter one knows almost no bounds. Unfortunately, right now, it seems impossible for the proposed algorithm to distinguish between the two, so formal equivalence inevitably leads to overlap in the results.

To finally contrast the two measures of linguistic productivity, I plotted both against the lines showing the number of types and number of tokens for each prefix (Figure 38). These values were obtained from the random samples of 100 bases that I used to calculate productivity measures. The results are conspicuous. While my measure is (unsurprisingly) perfectly correlated with the number of types, Baayen's measure appears as a mirror image of the number of tokens; the peaks of the former almost perfectly coincide with the valleys of the latter, and vice versa.

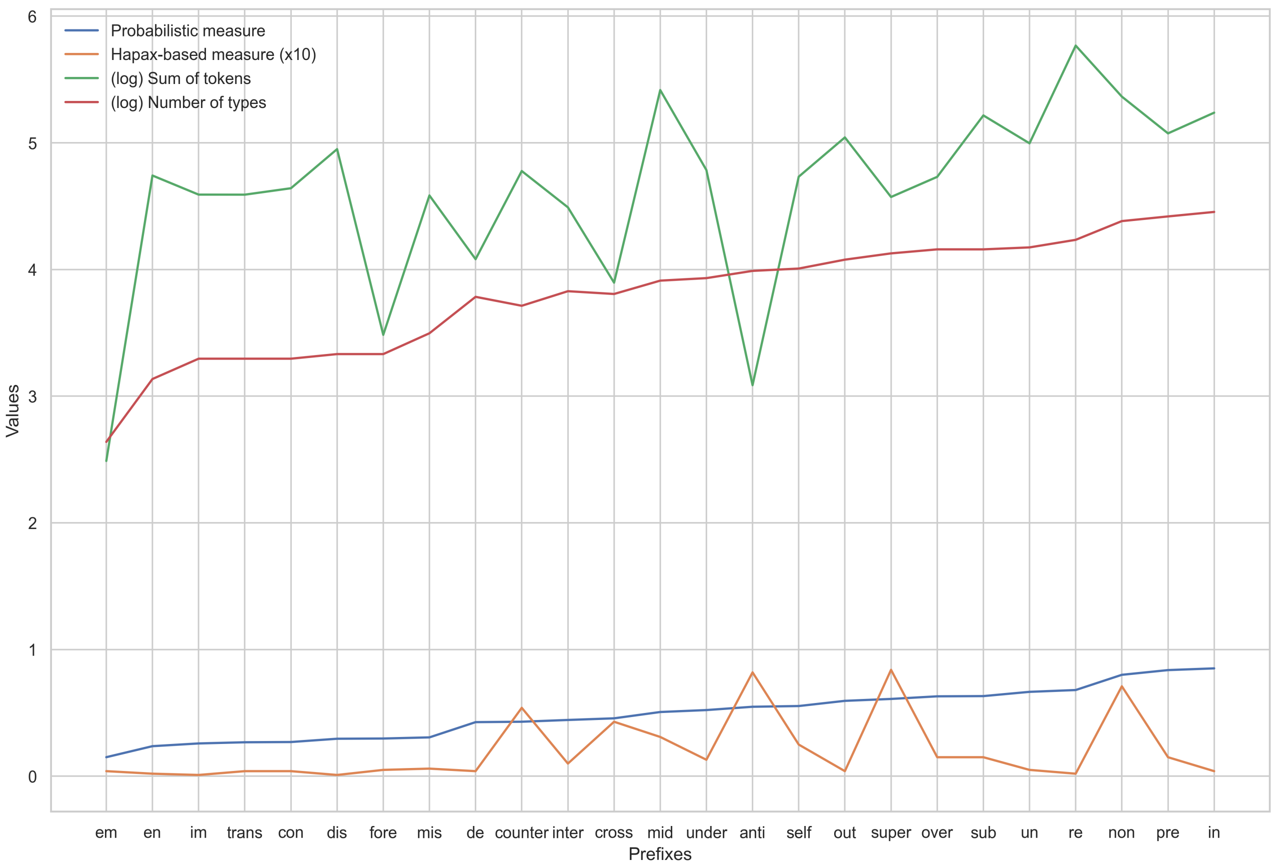

Figure 38. English prefixes' productivity measures plotted against the numbers of types and tokens

Having obtained the numbers, I wanted to check whether Hay and Baayen's findings would hold when tested against them. To do so, I performed a correlation analysis of English prefixes' productivity values and these prefixes' total sums of tokens. Given Hay and Baayen's results, one would expect to find a negative correlation (cf. 'the more often you encounter an affix, <…> the less productive that affix is likely to be' (Hay and Baayen, 2002: 219)). However, what can be seen in Figure 39 is a sufficiently strong positive trend ($\rho = 0.6$, $p = 0.001$): the more productive a prefix, the higher the token frequency of all words with this prefix. Noteworthy, a very similar positive correlation between prefixes' productivity values and these prefixes' total sums of tokens was observed with the Russian data: $\rho = 0.57$, $p < 0.001$.

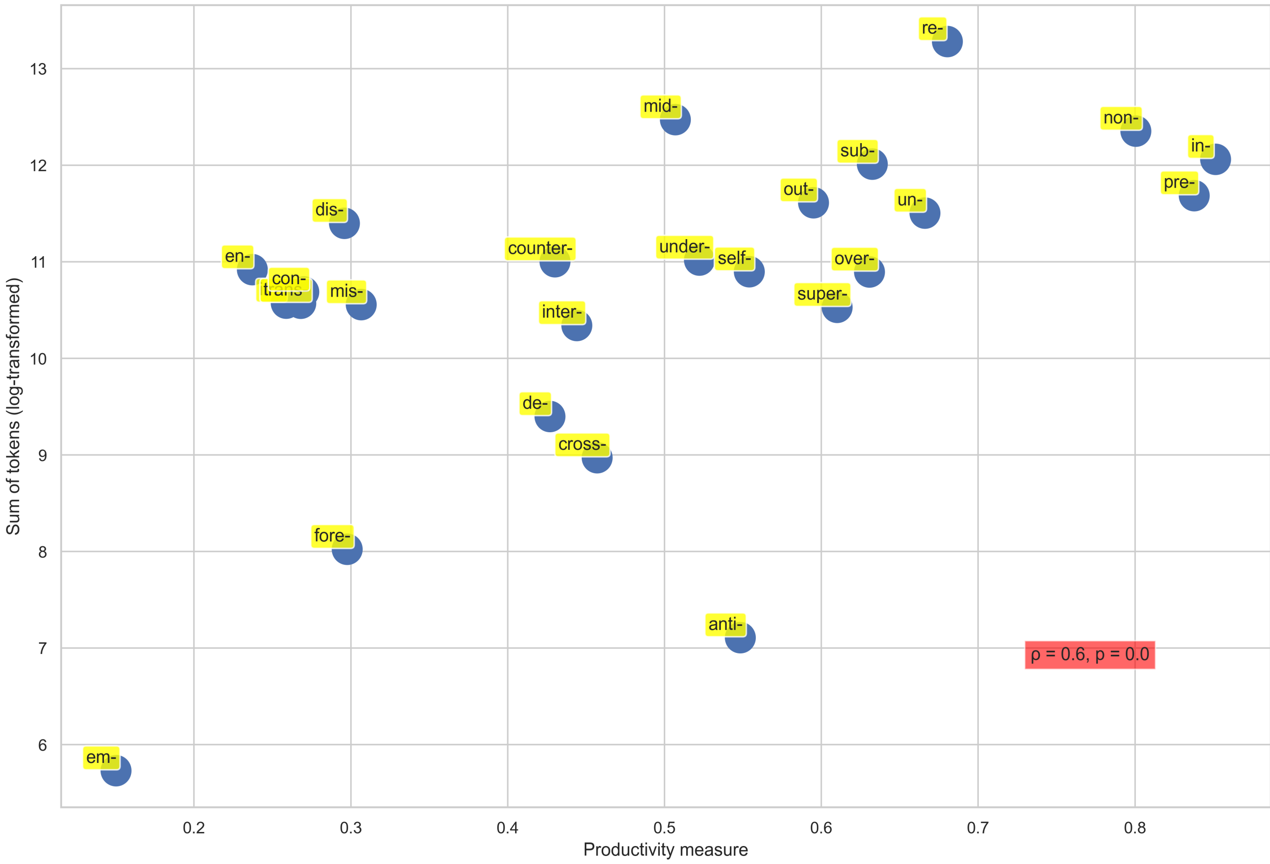

Figure 39. English prefixes' productivity measures correlated with prefixes' sums of tokens

5.6 Conclusion

It is a long-established view that high token frequency represents a sort of stumbling block for affixes' linguistic productivity. It has been argued that affixes encountered in many frequent items become less parsable and, by that, lose their ability to combine with new bases, cf.: 'The less useful an affix is <…>, the more productive it is likely to be' (Hay and Baayen, 2002: 222). However, based on my findings, the picture appears to be more complicated: high-frequency derivations with an affix, once they are accumulated in a certain amount of types, do not block the emergence of new low-frequency coinages but rather facilitate them, serving as pathbreakers for neologisms.

In this chapter, I tried to show that the unexpected relationship between affixes' frequency and productivity that has been observed in the literature is, most likely, an artefact of the established way of measuring linguistic items' productivity. Very simply, if you have an equation $productivity_i = c / T_i$, where c approaches a constant and T_i is the total number of derivations with an affix i, it is hard to expect anything other than a negative correlation between $productivity_i$ and T_i.

To provide a way out of this circular reasoning, I suggest that linguistic productivity is best viewed as the probability of an affix to combine with a random base. Using the internet corpus of English from 2018, I evaluated the linguistic productivity of 25 English prefixes: *anti*-, *con*-, *counter*-, *cross*-, *de*-, *dis*-, *em*-, *en*-, *fore*-, *im*-, *in*-, *inter*-, *mid*-, *mis*-, *non*-, *out*-, *over*-, *pre*-, *re*-, *self*-, *sub*-, *super*-, *trans*-, *un*-, *under*-. For each prefix, three probabilities were obtained: (1) P(X = 0), the probability of no occurrence of the combination of this prefix with a random base in the corpus; (2) P(X = 1), the probability that the combination of this prefix with a random base will be of low frequency; and (3) P(X = 2), the probability that the combination of this prefix with a random base will be frequent.

The true measure of linguistic productivity was estimated in two steps. First, the initial and transition probability distributions for a two-time-slice dynamic Bayesian network were learned on a random sample of 100 random bases obtained from the corpus. Second, the value of P(X = 1) + P(X = 2) was calculated for the 101st random base, given the last base in the sample. I found that, based on the evaluations of these probabilities, all prefixes, when arranged in order of ascending productivity, could be subdivided into three groups. The first group encompasses prefixes with the probabilities hierarchically arranged as P(X = 0) > P(X = 2) > P(X = 1). In the second group, one finds prefixes where the probabilities are aligned in this way: P(X = 2) > P(X = 0) > P(X = 1). Finally, the prefixes that belong to the last group reveal the following pattern: P(X = 2) > P(X = 1) > P(X = 0).

Interestingly, these categorical differences were found to emerge as manifestations of an inherently gradient structure. Thus, within the first group, the differences between probabilities P(X = 0) and P(X = 2) continuously decrease, while the differences between probabilities P(X = 2) and P(X = 1) continuously increase. Within the second group, a similar mechanism of change can be observed, though with different contrasts. The differences between probabilities P(X = 2) and P(X = 0) become larger, while the differences between probabilities P(X = 0) and P(X = 1) become smaller. Finally, within the third group, the gap between probabilities P(X = 2) and P(X = 1) successively narrows, while the gap between probabilities P(X = 1) and P(X = 0) widens.

All of the above raises an interesting question of how derivational patterns spread. As Haspelmath noted, ‘what is really remarkable about morphology is that morphological rules may <…> be unproductive’ (Haspelmath, 2002: 40). As an example of an unproductive derivational rule in English, he mentioned the suffix *-al*, the list of action nouns formed with which is fixed and cannot be extended, so that no words like **repairal*, **ignoral*, and **amusal* are possible. No less remarkable, however, is the fact that the productivity of even fully productive English affixes is not without its limits. Thus, it is not clear why, for example, given the high frequency of the verb *give* and the high productivity of the prefix *re-*, the derivation *re-give* is extremely unpopular, with 0 hits per million tokens in both COCA and *ententen18_tt31*. It is also unclear why, given that the verb *evolve* is more frequent than the verb *regulate*, only *dis-regulate* is actually attested in COCA and *ententen18_tt31*, although there seems to be nothing conceptually improbable or semantically incompatible in the possible combination *dis-evolve*.

One might conclude that, even for very productive affixes, there is no simple linear relation between base and derivation frequency. Rather, it is high-frequency items with a certain affix that play a pivotal role in the self-propagating of respective derivational patterns and the structuring of its output, with less-frequent members being grouped around more prominent ones. I believe that this ‘clustering hypothesis’ deserves further investigation.

6 How derivational patterns propagate themselves through the discourse

6.1 Introduction

Of all the different measures of linguistic (morphological) productivity, the one that has become the most well-known and widespread in the literature is the so-called ‘potential productivity’ (Baayen and Lieber, 1991; Baayen, 1992, 1993, 1994, 2009; Baayen and Renouf, 1996). In potential productivity, ‘a large number of hapaxes lead to a high value of P, thus indicating a productive morphological process. Conversely, large numbers of high-frequency items lead (…) to a decrease of P, indicating low productivity’ (Plag, 2021: 488). This measure has been used in multiple studies and continues to be used today (Fernández-Domínguez, Díaz-Negrillo, and Štekauer, 2007; Plag and Baayen, 2009; Zirkel, 2010; Marzi and Ferro, 2014; Mendaza, 2015; Pierrehumbert and Granell, 2018; De Smet, 2020, among others).

This account is aligned with the long-established view that a high token frequency represents a sort of stumbling block for affixes’ linguistic productivity. It has been noted many times that productivity increases with a pattern’s type frequency (Plag, 2003; Barðdal, 2006; Goldberg, 2006; Stefanowitsch, 2008; Schmid, 2017) because language users tend to form schematic representations of (constructionalise) certain patterns when they encounter many varying instantiations of them. Alternatively, high-frequency items are less likely to be conceived of as parts of constructions and do not tend to contribute to productive schema formation (Guillaume, 1973; Baayen and Lieber, 1991; Moder, 1992; Bybee, 1985, 1995, 2007). Researchers have argued that affixes that are encountered in many frequent items become less parsable and, thus, lose their abilities to combine with new bases. In other words, ‘the less useful an affix is (…), the more productive it is likely to be’ (Hay and Baayen, 2002: 222).

As summarised by Plag, ‘unproductive morphological categories will be characterised by a preponderance of words with rather high frequencies and by a small number of words with low frequencies. With regard to productive processes, one expects the opposite, namely large numbers of low-frequency words and small numbers of high-frequency words’ (Plag, 2018: 54).

This view of linguistic productivity, though logical and intuitively appealing, is neither unproblematic nor free of controversy. There are, as I see it, several major issues. First, tentative evidence appears to indicate that, although type frequency might be the main determinant of productivity, its interplay with token frequency is more complex than generally assumed. Specifically, De Smet reported that a high token frequency actually promotes productivity, provided it is combined with a high type frequency (De Smet, 2020).

Second, using the potential productivity hapax-based measure to demonstrate that high token frequency impedes productivity can be regarded as a self-fulfilling prophecy. It has been pointed out in the literature that this measure is ill-suited for the comparison of affixes with different token numbers (Gaeta and Ricca, 2006; see also Bauer, 2001; Pustylnikov and Schneider-Wiejowski, 2010). Calculating the ratio of the number of hapax legomena with a given affix to the total number of tokens with that affix is likely to result in overestimating the value of productivity for less-frequent constructions. Hence, it would be useful to explore whether the proposed relationship between type and token frequency will hold under the application of a measure of linguistic productivity that is not directly dependent upon one of the quantities.

Third, there are numerous indicators that affixes' recognisability and applicability are not linearly related. As Haspelmath noted, 'what is really remarkable about morphology is that morphological rules may (…) be unproductive' (Haspelmath, 2002: 40). Haspelmath mentioned the suffix *-al* as an example of an unproductive derivational rule in English. The list of action nouns formed with this suffix is fixed and cannot be extended, so words like **repairal*, **ignoral*, **amusal* are not possible.

No less remarkable, however, is the fact that the productivity of even fully productive English affixes is not without limits. It is not clear why, for example, given the high frequency of the verb *give* and the high productivity of the prefix *re-*, the derivation *re-give* is so extremely unpopular, with zero hits per million tokens both in COCA and the English Web 2018 corpus (*ententen18_tt31*, Sketch Engine; 21,926,740,748 words; Jakubíček et al., 2013). It is also not clear why, given that the verb *evolve* is more frequent than the verb *regulate* (37.56 i.p.m. vs. 31.89 i.p.m. in *ententen18_tt31*), only *dis-regulate* is actually attested in COCA and *ententen18_tt31* although there seems to be nothing conceptually improbable or semantically incompatible in a possible combination *dis-evolve*.

These observations are difficult to reconcile with the traditional account that links parsability and productivity. Hay, for example, assumed that more frequent bases tend to be associated with more frequent derivations and that '[the frequency with which] a (transparent) derived form is deployed in speech is likely to be a partial function of the frequency of the form upon which it is based' (Hay, 2001: 1051). However, this simple parse-and-paste model may not be an accurate description of how derivational patterns spread. Specifically, it seems that even for very productive affixes, there is no simple linear relation between base and derivation frequency.

Finally, and perhaps most importantly, all of the studies of linguistic productivity with which I am familiar work under the simplifying assumption that it only matters how many times language users encounter complex words they need to form schematic representations of certain patterns. The contexts in which users encounter these words, their discourse co-occurrence preferences are completely disregarded. Still, it is known from the history of the morphological integration of French and Latin loanwords in Middle English that the recognisability of an affix can be greatly facilitated if the complex words that contain it collocate with their own bases:

> The integration of French loanwords took place in several stages. The precondition for their utilization was the existence of a sufficient number of loanwords with and without a given affix. If the number of such pairs was great enough, even native speakers without any knowledge of French were able to analyse the affixed partners as having a meaningful affix which later on could become productive in combination with native or with borrowed words… (Dietz, 2015: 1921).

All these observations, as well as the probabilistic estimation of the linguistic productivity of 25 English prefixes presented in the previous chapter, seem to suggest that token frequency as such, contrary to common beliefs, cannot be considered a stumbling block for derivational patterns. The observed dependence of the emergence of low-frequency derivations on the existence of numerous high-frequency derivations with the same affix requires clarification. I believe that this phenomenon can be explained as follows.

High-frequency derivations, once they are accumulated in a certain number of types, do not block the emergence of new low-frequency coinages but rather facilitate them, serving as pathbreakers for neologisms. This happens because derivational patterns propagate themselves in a 'contagious' fashion via discourse co-occurrences of frequent derivations with their bases. These co-occurrences, if sufficiently numerous, lead to better recognition of the pattern (derivational element) and its subsequent applicability to new bases that tend to appear in the same contexts. Thus, high-frequency items with a certain affix play a pivotal role in the self-propagating of respective derivational pattern and the structuring of its output, with less-frequent members being clustered around more prominent ones.

The greater the number of frequently used words with a certain affix, the higher the chances that some of them will collocate with their own bases. The more persistent these co-occurrences are, the more likely it is that the respective affix will become recognisable, parsable, and applicable, that is, productive. If so, then every instance of such a discourse-conditioned pattern's invigoration is a short-term memory process (Divjak, 2022; Schwieter and Wen, 2022), and the range of applicability of the temporarily refreshed pattern should be limited to the nearest context. Hence, one would expect to see many low-frequency coinages clustered around those high-frequency anchors with which they typically collocate. In the remainder of this chapter, I provide simulation-based and corpus evidence supporting this claim.

6.2 Data and hypothesis

The data for the study was obtained as follows. First, 995 random content words (nouns, verbs, and adjectives) without prefixes were sampled from *ententen18_tt31*, an internet corpus of English from 2018 containing more than 20 billion words. The raw frequencies of these bases, ranging from 48,421,599 to 54 tokens, were recorded. Second, each of 25 English prefixes on my list (*anti*-, *con*-, *counter*-, *cross*-, *de*-, *dis*-, *em*-, *en*-, *fore*-, *im*-, *in*-, *inter*-, *mid*-, *mis*-, *non*-, *out*-, *over*-, *pre*-, *re*-, *self*-, *sub*-, *super*-, *trans*-, *un*-, *under*-) was coupled with each of those 995 bases, so that the bases remained the same for all prefixes. The raw frequencies of all constructed derivations were then queried in the same corpus. The resulting dataset was arranged in descending order of base frequencies.

The frequency distributions of the derivations with each prefix are visualised in Figure 40. It is clear from the graph that the derivation frequency is not proportional to the rank of the respective base for any prefix. This impression can be confirmed by plotting both distributions on a log-log graph, with the axes being *log(derivation frequency)* and *log(base frequency rank)*.

For better readability, only every 30th of all base-derivation pairs actually attested in the corpus is displayed in Figure 41. The upper black line connects points that indicate base frequencies, the lower black line — those that indicate derivation frequencies, the coloured lines in between map derivations onto their bases. A remarkable criss-cross pattern can be observed for all 25 prefixes: higher-ranked bases tend to result in lower-ranked derivations while lower-ranked bases tend to result in higher-ranked derivations.

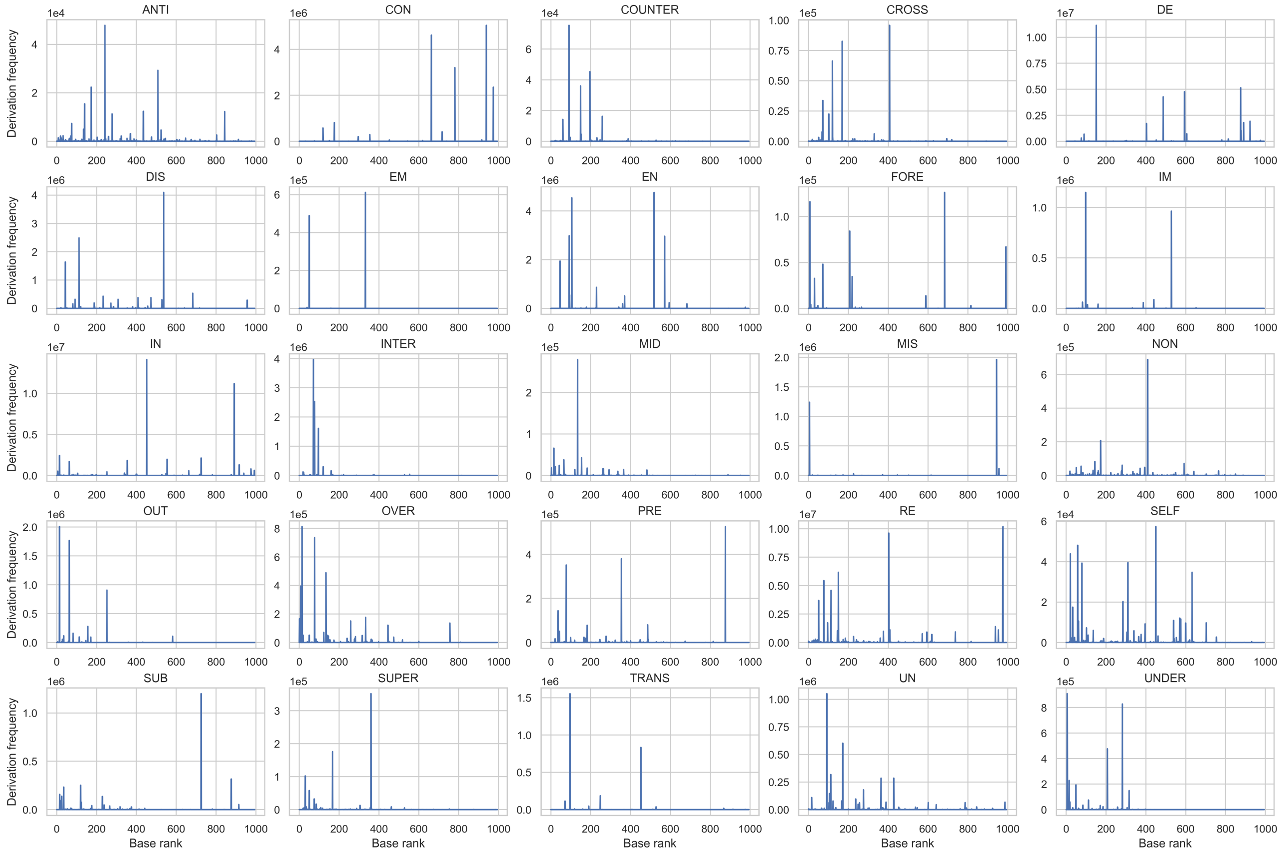

Figure 40. Frequency distributions of derivations with 25 English prefix

Figures 40 and 41 suggest that the frequencies of bases and derivations are, indeed, not linearly related and that the latter, when arranged in order of descending frequency of their bases, reveal a multiple-cluster pattern. However, the reliable identification of the clusters of derivations is problematic. First, the mean frequency of items in different clusters of the same prefix may vary as one moves down the line from more-frequent to less-frequent bases. Second, there may be high-frequency derivations that stand alone and do not sprout any clusters. Third, it is unclear how many non-attested (zero-frequency) derivations are tolerable between members of a cluster before it is no longer considered a cluster.

One possible simple solution would be to replace the actual frequency value of each derivation D_i by the sum of the frequency values of all other derivations weighted in accordance with their remoteness from this particular derivation (Table 28). Each weight, then, could be an exponent of a negative number of ranks separating D_i and D_j, for $i \neq j$.

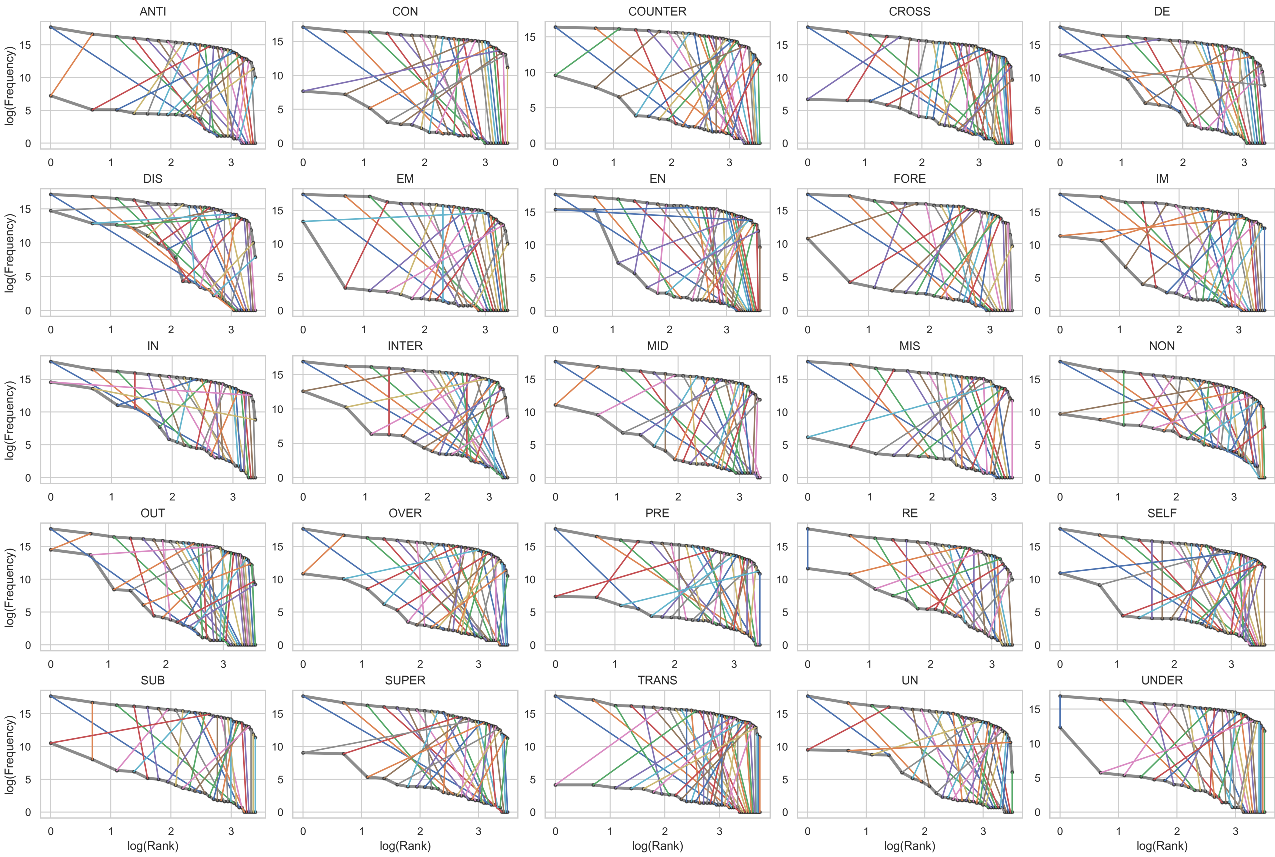

Figure 41. Log(frequency)-log(rank) graph of derivations and bases

Table 28. Template for calculating weighted frequencies in the process of cluster identification

observations	D_{i-3}	D_{i-2}	D_{i-1}	D_i	D_{i+1}	D_{i+2}	D_{i+3}
raw frequency	235	51151	0	27	2	1	44912
rank distance	-3	-2	-1	0	-1	-2	-3
weight	0.04	0.13	0.36	0	0.36	0.13	0.04
weighted frequency	11.69	6649.63	0	0	0.73	0.13	2236.03
VALUE OF INTEREST	…	…	…	8898.21	…	…	…

Advantages of this approach include the following: 1) weights are exponentially decaying as the rank distance between D_i and D_j is growing; thus, the contribution of remote high-frequency values in calculating the weighted frequency of D_i approaches zero; 2) the weighted frequency of stand-alone high-frequency derivations approaches zero as their actual values are thrown out of the equation and only the frequencies of their close neighbours are considered; 3) truly clustered derivation frequencies, regardless of the magnitude of their actual values, when weighted, reveal a characteristic violin-shape distribution that makes them visually identifiable.

What can be said about the groups of derivations that look like clusters[5]? First, they tend to appear in the upper half of base frequency distributions but are also encountered in the lower half. Second, clusters of derivations with identical bases and different prefixes only partially overlap or do not overlap at all. Third, between the clusters of derivations with the same prefix, when they are situated near each other, there tend to appear transitional zones, which results in the emergence of super-clusters. Fourth, for all prefixes, there are multiple cases of high-frequency derivations that do not form any clusters around them. Notably, these stand-alone derivations are generally semantically opaque items, or terminological units, or (sometimes) erroneous entries — where initial clusters of letters simply coincide in form with respective prefixes.

Finally, clustered derivations with the same prefix rarely show signs of semantic similarity or topical relatedness. Consider derivations with the prefix *mid-*. Most frequent of them are definitely subsumed within one category expressing the notion of the middle of a period of time: *mid-night*, *mid-week*, *mid-life*, *mid-year*, *mid-season*, *mid-September*, *mid-October*, *mid-winter*, *mid-December*, *mid-afternoon*, *mid-February*, *mid-month*, and so on. However, clustered around them are lexical items which often cannot easily be related to the idea of time (Table 29).

Table 29. Derivations with *mid-*: some frequent words and their cluster members

mid-stage (1359)		*mid-september (16232)*		*mid-winter (14429)*	
base rank	derivation and its raw frequency	base rank	derivation and its raw frequency	base rank	derivation and its raw frequency
207	*mid-ground* (680)	249	*mid-clean* (12)	360	*mid-vision* (4)
208	*mid-situation* (1)	250	*mid-box* (20)	361	*mid-shoot* (64)
209	*mid-software* (0)	251	*mid-variety* (29)	362	*mid-chemical* (1)
210	*mid-item* (0)	252	*mid-door* (12)	363	*mid-hall* (22)
211	*mid-web* (21)	253	*mid-host (0)*	364	*mid-limited* (1)
212	*mid-additional* (0)	254	*mid-river* (286)	366	*mid-database* (0)
214	*mid-maintain* (0)	255	*mid-deep* (12)	367	*mid-regional* (63)
215	*mid-update* (11)	256	*mid-earth* (50)	368	*mid-edit* (17)

[5] Table with weighted frequencies and cluster labels for all 25 prefixes is available at: https://docs.google.com/spreadsheets/d/12R2k4eYj-LlJns8yebtjSw2ahFUGsebj/edit?usp=sharing&ouid=115071188178573577784&rtpof=true&sd=true.

216	*mid-color* (12)	257	*mid-avoid* (0)	369	*mid-scientific* (0)
217	*mid-track* (131)	258	*mid-weight* (1638)	370	*mid-roll* (299)
218	*mid-girl* (0)	259	*mid-obtain* (0)	371	*mid-stick* (22)
219	*mid-father* (0)	260	*mid-european* (139)	372	*mid-quote* (17)
220	*mid-photo* (24)	261	*mid-october* (15976)	373	*mid-surprise* (1)
		262	*mid-career* (10349)	374	*mid-manner* (0)
		263	*mid-unique* (0)	375	*mid-net* (7)
		265	*mid-equipment* (0)	376	*mid-presentation* (31)
		266	*mid-station* (274)	377	*mid-surround* (2)

Among the prefix-base combinations in Table 29, one can easily distinguish those that are well-established (*mid-career*), those that are semantically incongruous and hence unattested (*mid-father*), and those that are created on the fly — context-driven neologisms that are hard to semanticise outside this particular context (*Chapters and pages should end* ***mid-situation*** *to increase the reader's desire to keep reading* or *It's got macros and regular expressions and automatic backup files in case you do something stupid in* ***mid-edit***).

Given all of the above, there are likely just two ways to account for the observed clustering preference of derivations. The first is to assume that no clustering preference really exists and that the cluster-like structure visualised in Figure 40 is just some artefact of the sampling procedure. Alternatively, one can hypothesise that English derivational patterns propagate themselves due to the discourse co-occurrences of frequent derivations with their bases. Specifically, the hypothesis that I intend to test is formulated as follows: the more connected a group of bases is (i.e., the more frequently they show up in the same contexts), the higher the probability that they will give rise to respective prefix derivations, provided that at least one derivation is strongly connected in discourse to at least one of the bases.

This hypothesis would imply, for example, that the frequent derivations with *mid-* in Table 29 belong to different clusters because they collocate with different bases mentioned in the table. As one can see from the network in Figure 42, where the edges connect collocating words and the positions of the nodes are calculated using the Fruchterman-Reingold force-directed algorithm, all the bases (unlabelled nodes) are clustered closely together, and the most frequent derivations form the exact three groups that I expected to find. The Fruchterman-Reingold algorithm pushes and pulls nodes apart as if they were connected via springs so that vertices that are adjacent to each other (have similar connections) are shown near each other, whereas vertices that are not adjacent are placed far apart (Fruchterman and Reingold, 1991). Hence, the layout of the network in Figure 42 supports my intuition about the different co-occurrence preferences of the frequent derivations with *mid-* in Table 29.

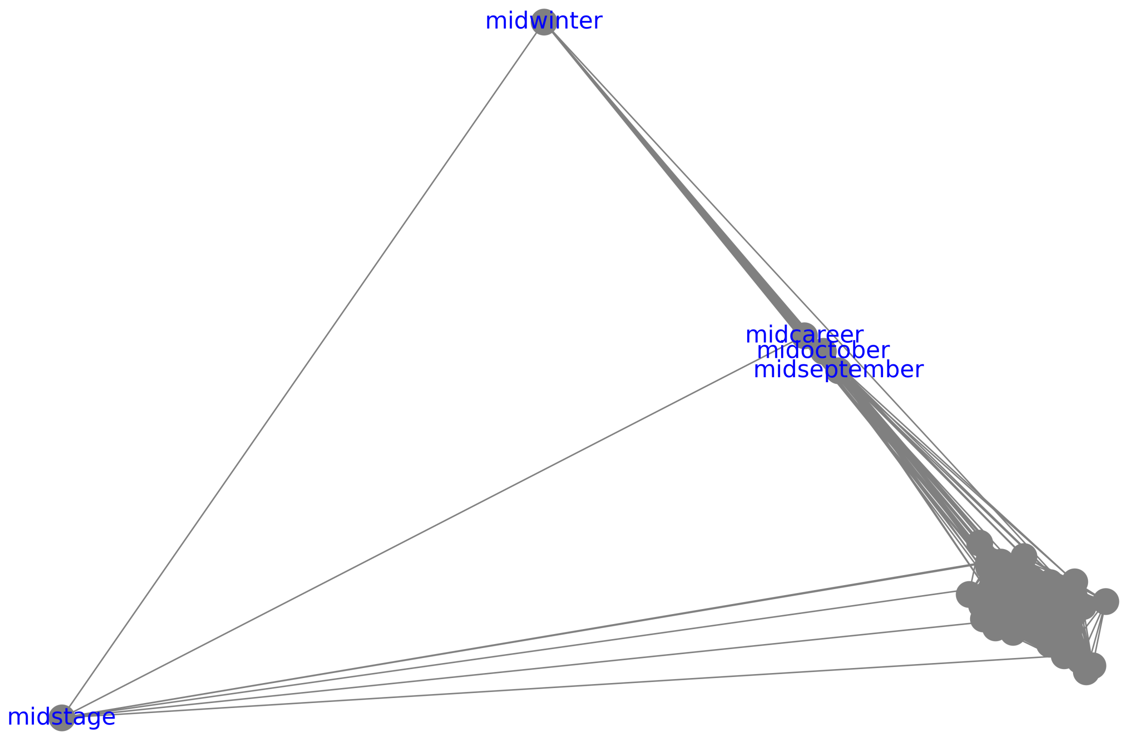

Figure 42. Network of collocating bases and derivations from Table 29

As another example, the aforementioned puzzle of why, although the verb *evolve* is more frequent than the verb *regulate*, only *dis-regulate* is actually attested can be explained by the fact that *evolve* does not belong to any cluster of derivations with the prefix *dis-*, while *regulate* is a member of the group formed around *dis-continuous*. Here, 'belonging to a cluster' means appearing in the same contexts, collocating frequently both with the central derivation and its base. From Figure 43, it is evident that *regulate* is indeed more strongly associated with both *continuous* and *dis-continuous* than with *evolve* (edges are labelled with the collocational strengths of the respective pairs of words).

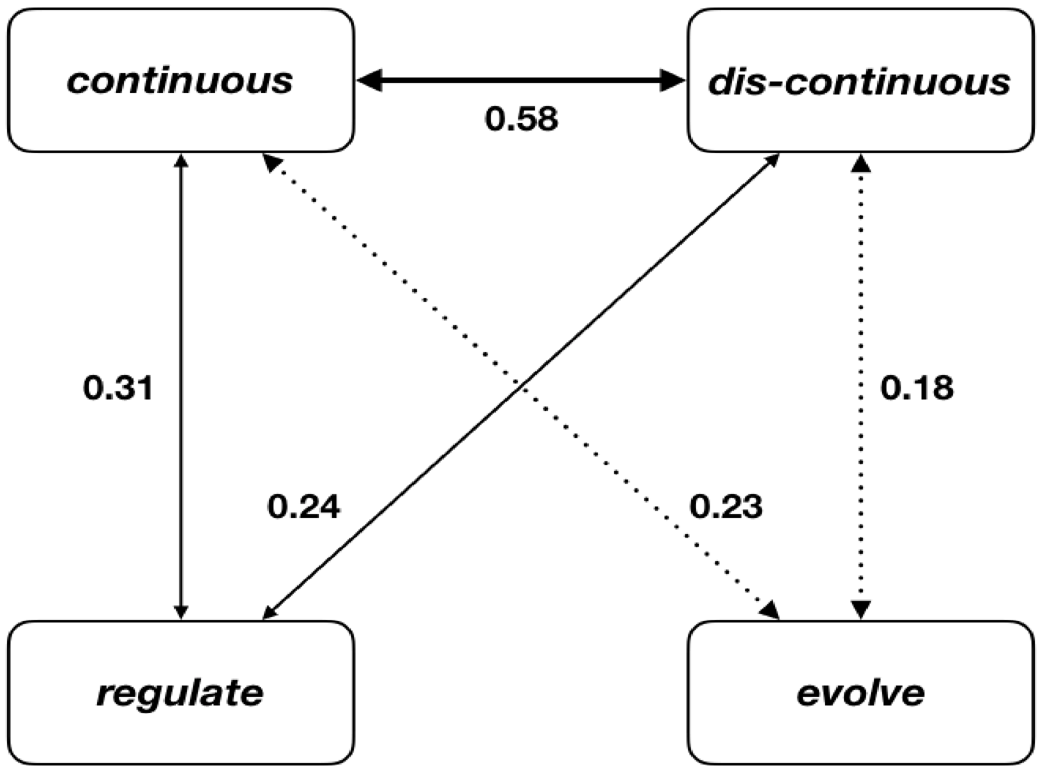

Figure 43. Cosine similarities of some words within and without the cluster formed around *dis-continuous*

Henceforth, I will be using as a measure of word pairs' collocational strength their cosine similarity values obtained from GloVe, a word-embedding model pre-trained on a Wikipedia dump from 2014 and the Gigaword archive of newswire text data (5th edition) (Pennington, Socher, and Manning, 2014). The reasons for this are twofold. First, given the number of word pairs — both bases and derivations — assessing their collocational strengths by directly looking up co-occurrence counts in a corpus is infeasible. Such an endeavour would require 334,615,515 corpus queries overall. Second, cosine similarity is a highly reliable approximation of collocational strength. Intuitively, a cosine similarity equal to 1 indicates that two words' contexts are identical; hence, they must collocate. Conversely, a cosine similarity equal to 0 indicates that two words' contexts do not match; hence, they cannot collocate.

In order to show that the two measures are indeed linearly related, I obtained logDice scores (a well-established statistical association measure for identifying collocations; Rychlý, 2008) and cosine similarity values for 2,469 random word pairs in my data. Their correlation ($r = 0.63$, $p < 0.0001$) is visualised in Figure 44.

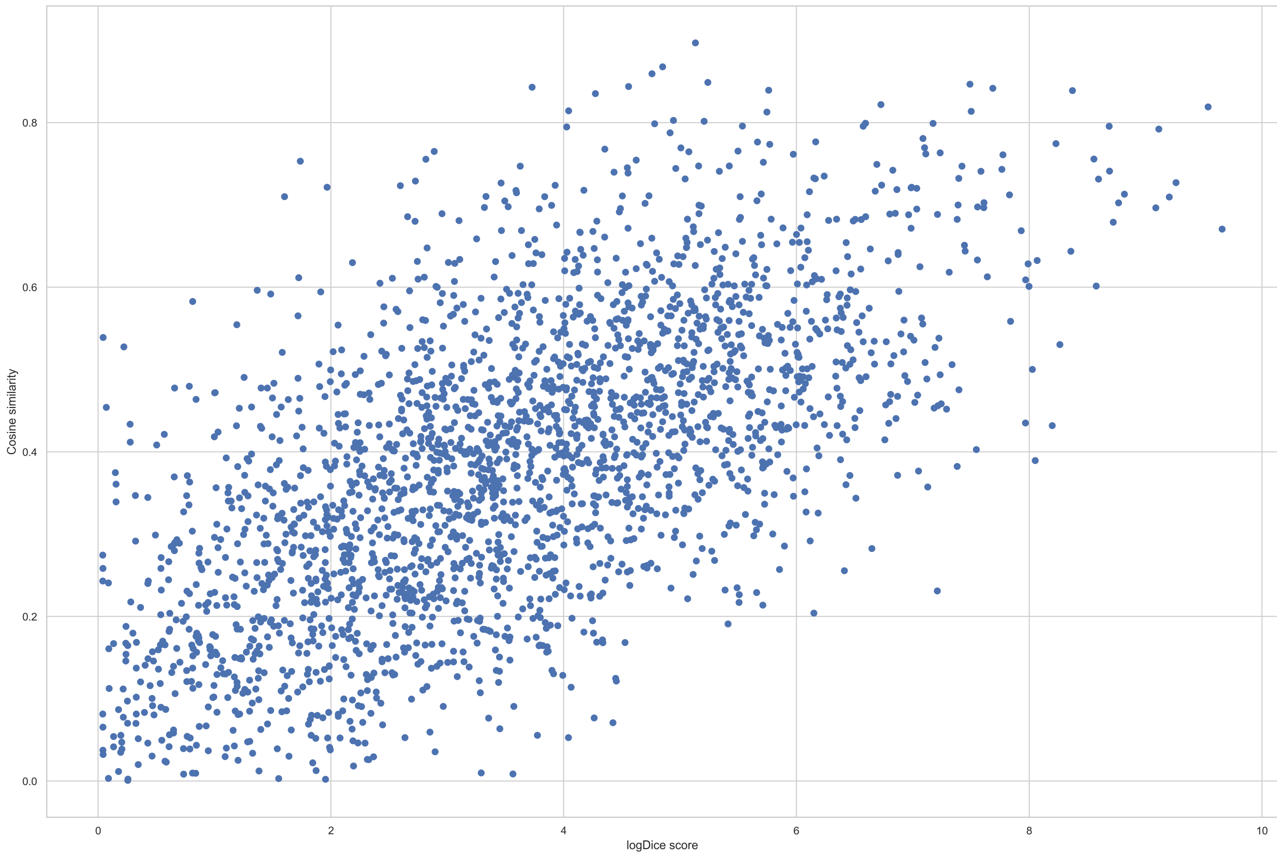

Figure 44. Correlation of logDice scores and cosine similarity values for 2,469 random word pairs

6.3 Methods and models

6.3.1 Inference about clusters

First thing to test is whether the clustering preference that is observable in Figures 40–41 really exists and is not some artefact of the sampling procedure. The problem may be cast into the framework of Bayesian inferencing: given a sample of derivations with a certain prefix and their respective bases, the goal is to infer how likely it is that these derivations form a cluster. To do this, I constructed a Markov network of the architecture specified in Figure 45 using the parametrisation given in Table 30. The model included seven variables: cluster, homogeneity, base frequency, base similarity, derivation frequency, derivation similarity, base/derivation similarity. All the variables were discretised to values in the ranges of [0, 1] or [0, 1, 2]. I will discuss them one by one.

Suppose that seven bases and their same-prefix derivations are randomly sampled from the data. The **cluster** node encodes the variable of interest, namely, whether the sampled derivations belong to the same cluster—C = 1 [‘cluster’]—or not—C = 0 [‘not a cluster’]. The **homogeneity** node communicates whether bases in the sample belong to the same frequency bin. For each prefix, I identified several equally spaced frequency bins for bases, with thresholds set by the following quantiles: [.11, .22, .33, .44, .55, .66, .77, .88], and for derivations, with thresholds set by the following absolute values: [0, 1, 9, 99, 999, 9999, 99999, 999999]. If all bases in the sample belong to the same frequency bin, H = 1 ['homogeneous']; otherwise, H = 0 ['non-homogeneous']. The **base frequency** node stores information on how frequent the sampled bases are. If the median frequency of the bases belongs to one of the bins in {9, 8, 7}, BF = 2 ['of high frequency']; if the median frequency belongs to one of the bins in {6, 5, 4}, BF = 1 ['of medium frequency']; otherwise, BF = 0 ['of low frequency']. The **base similarity** node measures the average collocational strength of the bases. If the mean pairwise cosine similarity of the sampled bases > 0.4, BS = 2 ['strongly connected']; if the mean pairwise cosine similarity < 0.2, BS = 0 ['weakly connected']; otherwise, BS = 1 ['moderately connected']. The **derivation frequency** node stores information on how frequent the sampled derivations are. If at least one derivation belongs to one of the bins in {9, 8, 7, 6}, DF = 2 ['of high frequency']; if the median frequency of the derivations belongs to the bin in {1}, DF = 0 ['not attested']; otherwise, DF = 1 ['of low frequency']. The **derivation similarity** node measures average collocational strength of the sampled derivations. If at least one pair of the derivations has cosine similarity > 0, DS = 1 ['connected']; otherwise, DS = 0 ['not connected']. Finally, the **base/derivation similarity** node indicates whether there is a connection between the bases and derivations in the sample. If at least one base-derivation pair in the sample has cosine similarity > 0, BDS = 1 ['connected']; otherwise, BDS = 0 ['not connected'].

Table 30. Model specification

label	variable	range of values		
C	cluster	0	1	
H	homogeneity	0	1	
BF	base frequency	0	1	2
BS	base similarity	0	1	2
DF	derivation frequency	0	1	2
DS	derivation similarity	0	1	
BDS	base/derivation similarity	0	1	

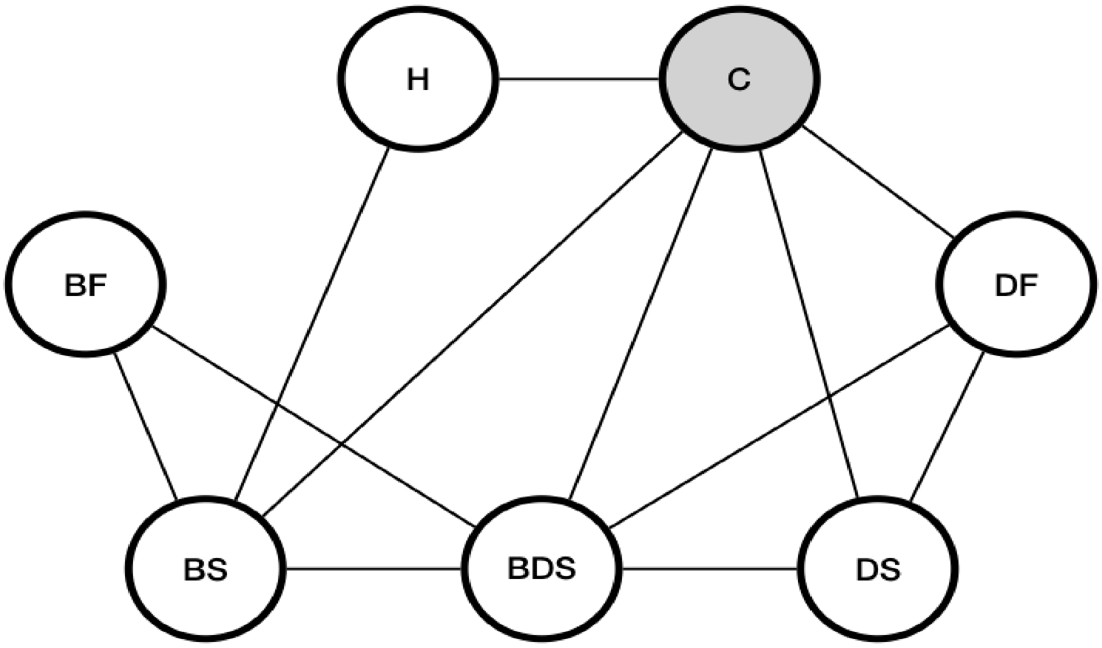

Figure 45. Markov model for identifying clusters of derivations

Each pair of connected nodes in Figure 45 was treated as a factor whose potentials reflect my prior beliefs about how well the respective two variables agree with each other (Koller and Friedman, 2009). For example, in Table 31, it is shown that C and H are believed to agree very well. If bases in the sample do not belong to the same frequency bin (i.e., there is no homogeneity), then the respective derivations are unlikely to cluster, and vice versa. Still, I made allowances for non-zero probabilities of these variables disagreeing with each other.

Table 31. Factorisation of the variables H and C in the Markov model

C	H	phi(C, H)
C = 0	H = 0	1000
C = 0	H = 1	1
C = 1	H = 0	1
C = 1	H = 1	1000

Clusters were also considered to be more likely if (1) the level of base similarity was high, (2) the level of derivation similarity was high, (3) the level of base/derivation similarity was high, and (4) the level of derivation frequency was high. As evidenced by the absence of edge between nodes BF and C in Figure 45, I assumed that the level of base frequency does not influence clustering preference directly but rather does so through the intermediary nodes BS and BDS. This is because of the aforementioned observation that there was no linear relation between base and derivation frequency, and clusters did not only tend to appear in the upper halves of base frequency distributions.

After the model was specified and parametrised, I tested how good it was in predicting formation of clusters. 25,000 random samples of seven base/derivation pairs each were drawn from the data. The probability of selecting each prefix was equal to 1/25. The probability of consecutive (following each other in the frequency rank hierarchy) and non-consecutive sampling of bases was equal to 1/2. For each sample, a prediction was made based on the model by instantiating (setting to evidence) all nodes other than C and inferring the most likely value of C using a variable elimination algorithm. Model predictions were than compared against my cluster labels. 82% of predictions were found to be correct, among them 73% for true clusters and 83% for non-clusters.

These sufficiently accurate results convince me that the clustering preference of English prefixed derivations does exist and that the model detects and replicates the data-generating process fairly well. One can get insight into this process by setting the variables in the model to different values and obtaining marginal distributions P(C) without evidence. Comparisons of these probability distributions show that (1) frequency by itself, either of bases or derivations, plays almost no role in clustering; (2) the presence of at least one base/derivation pair with a high cosine similarity greatly facilitates clustering; and (3) a high level of cosine similarity between either bases or derivations, given the presence of at least one base/derivation pair with high cosine similarity, makes clustering almost certain. These observations are in agreement with my hypothesis.

6.3.2 Inference about derivations' frequency

Another way to test the assumptions underlying my clustering hypothesis is to construct a model capable of predicting the frequency level of a random derivation given information about its base and some other base/derivation pairs, not necessarily neighbouring them in the base frequency rank hierarchy. To accomplish that, I designed a plate model for a pair of base/derivation pairs, so that the model's parameters and structure could be reused as a template for a potentially infinite set of Bayesian networks (Figure 46).

Nodes labelled **Base 1** and **Base 2** contain information about frequency levels of any two randomly selected bases. The respective variables take values in the range [0, 1, 2], according to the same logic as described above for the BF node in the Markov model. That is, if frequency of the first base belongs to one of the bins in {9, 8, 7}, Base 1 = 2; if its frequency belongs to one of the bins in {6, 5, 4}, Base 1 = 1, otherwise Base 1 = 0. Variables **Derivation 1** and **Derivation 2** take values in the range [0, 1]: if frequency of the first derivation belongs to the bin in {1}, Derivation 1 = 0 ['not attested']; otherwise Derivation 1 = 1 ['attested'].

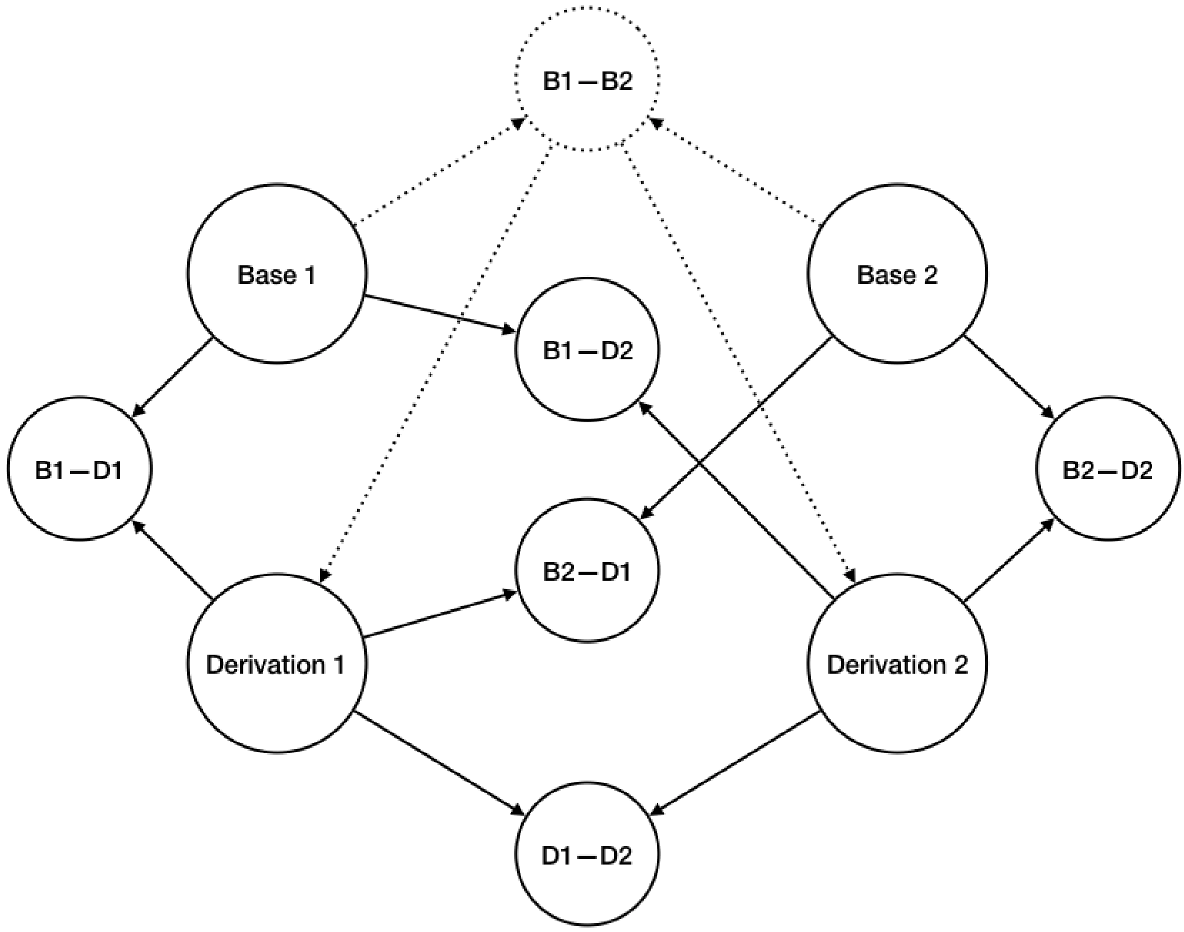

Figure 46. Bayesian model for identifying frequencies of derivations

All other nodes in the model are used to encode collocational strengths of respective pairs of words. Variables **B1–D1**, **B1–D2**, **B2–D1**, **B2–D2**, and **D1–D2** equal 1 ['connected'] if cosine similarity between the words that are their parents in the model is greater than zero and 0 ['not connected'] otherwise. Variable **B1–B2** equals 2 ['strongly connected'] if cosine similarity between the bases is greater than 0.3, 1 ['moderately connected'] if it is less than 0.3 but greater than zero, and 0 ['not connected'] otherwise. Importantly, as the dashed lines in Figure 46 indicate, the node B1–B2 is only added to the template when both sampled bases belong to the same frequency level (i.e., there is homogeneity).

The prior probabilities of the base frequency levels were assumed to be equal: P(Base 1 = 0) = P(Base 1 = 1) = P(Base 1 = 2) = 0.33, due to the sampling procedure. The prior probabilities of the derivation frequency level were obtained from the respective prefixes' probabilistic linguistic productivity measures calculated as suggested in the previous chapter. Thus, given that linguistic productivity of *em-* is estimated as 0.15, for any derivation with this prefix, P(Derivation 1 = 0) = 0.85 and P(Derivation 1 = 1) = 0.15. Conditional probability distributions for all child nodes in the model were specified in such a way so as to reflect one's intuitive belief that greater collocational strength of a pair of words correlates with these words' higher frequency.

In a nutshell, the plate model comprises a number of V-structures of the form *frequency(word_1)* → *cosine_similarity(word_1, word_2)* ← *frequency(word_2)*. Probabilistic reasoning, then, goes along the following lines: 1) if we observe that *word_1* and *word_2* collocate with each other, then we conclude that their frequencies are not independent of each other, 2) if we also observe that one of the words is frequent, then it increases our belief in that the second word is frequent as well.

If my hypothesis about the nature of derivational patterns' self-propagation is correct, then I would expect to find a significant positive correlation between the actual frequencies of derivations with a certain prefix and the likelihoods of these derivations inferred by the model. The process of inference ran as follows. For each base/derivation pair in my data, a random sample containing some other derivations with the same prefix as well as their respective bases was obtained. The base/derivation pair under investigation was then added to the sample, after which each base and each derivation in the sample was paired with each of their counterparts. The template of Figure 46 was applied to all possible combinations of two bases and two derivations. A prediction about the most probable value of the node representing derivation of interest was made by instantiating all other nodes in the model and running a variable elimination algorithm. The size of the sample was limited to 100 pairs because of the algorithm's complexity.

The densities of the obtained values P(Derivation_i = 1) for all 25 English prefixes are presented in Figure 47. The values are split into three groups: high-frequency derivations (blue lines), low-frequency derivations (orange lines), and not attested derivations (green lines). It can be observed that all the subplots in Figure 47 reveal remarkable similarities: 1) high-frequency derivations are characterised, for all prefixes, by higher probabilities of being created, 2) low-frequency derivations are centred near 50% probability threshold, and 3) not attested derivations are shifted furthest to the left along the *x*-axis. Looking somewhat peculiar among the others is the prefix *non-*, but this is, I think, due to the inconsistencies in how GloVe model treats hyphenated and non-hyphenated words with *non-*.

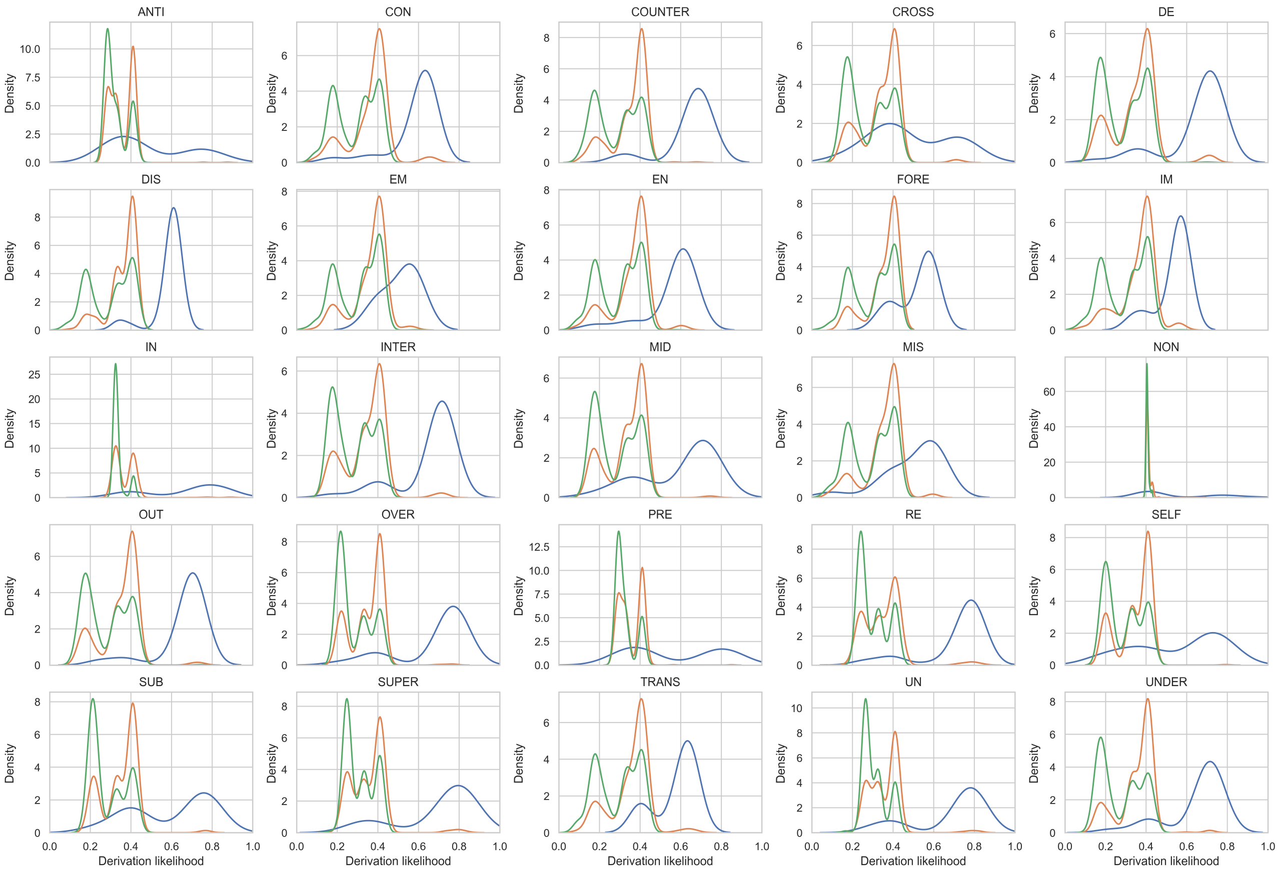

Figure 47. Densities of the derivations' likelihoods for 25 English prefixes

As hypothesised, there is a positive correlation between the actual frequencies of derivations and the likelihoods of these derivations inferred by the model. Besides, three groups of probability values corresponding to the three derivation frequency levels are characterised by significant shifts in locations as confirmed by Kruskal–Wallis test by ranks (Table 32; Conover's test showed $p < 0.001$ for all pairwise comparisons).

By setting P(Derivation = 1) > 0.5 as a cut-off point for separating high-frequency derivations from the rest, I reached a noteworthy accuracy of the model ranging from .92 to .99, depending on the prefix (Table 32). An attempt to automatically identify not attested derivations by setting P(Derivation = 1) < 0.3 as a threshold brought less impressive, though mostly tolerable, especially for more productive patterns, results (Table 32).

Table 32. Spearman correlation coefficients, H-values of Kruskal–Wallis test by ranks, and model's accuracy metrics

prefix	ρ (p-value)	H (p-value)	Accuracy of the model	
			frequent derivations	not attested derivations
anti-	.25 (< 0.001)	64.53 (< 0.001)	.97	.67
con-	.32 (< 0.001)	120.63 (< 0.001)	.98	.53
counter-	.31 (< 0.001)	103.38 (< 0.001)	.99	.59
cross-	.29 (< 0.001)	79.99 (< 0.001)	.98	.62
de-	.30 (< 0.001)	124.42 (< 0.001)	.97	.61
dis-	.31 (< 0.001)	132.01 (< 0.001)	.99	.53
em-	.17 (< 0.001)	35.11 (< 0.001)	.99	.44
en-	.26 (< 0.001)	82.18 (< 0.001)	.99	.50
fore-	.20 (< 0.001)	53.80 (< 0.001)	.99	.46
im-	.22 (< 0.001)	59.37 (< 0.001)	.99	.48
in-	.50 (< 0.001)	199.31 (< 0.001)	.96	.77
inter-	.35 (< 0.001)	143.26 (< 0.001)	.98	.61
mid-	.27 (< 0.001)	78.79 (< 0.001)	.98	.61
mis-	.21 (< 0.001)	54.21 (< 0.001)	.99	.51
non-	.29 (< 0.001)	129.67 (< 0.001)	.92	.88
out-	.32 (< 0.001)	130.96 (< 0.001)	.98	.63
over-	.44 (< 0.001)	217.89 (< 0.001)	.98	.69
pre-	.33 (< 0.001)	103.22 (< 0.001)	.96	.66
re-	.49 (< 0.001)	292.35 (< 0.001)	.96	.68
self-	.32 (< 0.001)	96.41 (< 0.001)	.98	.64
sub-	.42 (< 0.001)	167.94 (< 0.001)	.97	.68
super-	.34 (< 0.001)	108.45 (< 0.001)	.98	.65
trans-	.29 (< 0.001)	97.45 (< 0.001)	.98	.55
un-	.46 (< 0.001)	227.84 (< 0.001)	.97	.70
under-	.39 (< 0.001)	167.61 (< 0.001)	.98	.65

It is clear that there is a greater overlap in the distribution of probabilistic values between the low-frequency and not attested derivations than between either of them and high-frequency derivations. This is not surprising if one takes into account that the majority of low-frequency derivations are words with less than 100 hits in the multi-billion corpus. From this perspective, most of them are possible rather than actual words (Aronoff, 1976; Haspelmath, 2002), so the model correctly estimates their chances of being and not being created as fairly equal. On the other hand, some derivations that are not attested in this particular corpus, theoretically, stand a good chance of being coined at some point, given the network of their discourse connections (provided of course that there are no phonological, morphological, semantic, or pragmatic restrictions).

To show that the model does not rely in its predictions exclusively on the prefixes' linguistic productivity or base frequency levels but rather reads them off the base and derivations' co-occurrence patterns, the probabilities of each prefix being combined with 10 randomly sampled frequent bases are given in Table 33. Bolded are the values of P(Derivation = 1) > 0.5. It can be seen that the highest probabilities are assigned to the derivations that are not only well-attested but also fairly transparent in composition and meaning and thus likely to collocate with their bases.

Table 33. Probabilities of prefix-base combinations inferred by the model

	junction	*border*	*cover*	*ground*	*come*	*action*	*season*	*treat*	*profit*	*view*
anti-	0.287	0.343	0.412	0.418	0.412	0.340	0.414	0.404	0.328	0.430
con-	**0.640**	0.417	0.334	0.407	0.407	0.401	0.404	0.334	0.327	0.404
counter-	0.346	0.408	0.408	0.406	0.326	**0.680**	0.402	0.416	0.322	0.405
cross-	0.190	**0.705**	0.407	0.406	0.342	0.410	0.335	0.311	0.416	0.403
de-	0.153	0.417	0.407	0.412	0.409	0.411	0.400	0.401	0.327	0.400
dis-	**0.645**	0.104	**0.611**	0.401	0.410	0.414	0.410	0.314	0.341	0.406
em-	0.177	0.416	0.403	0.407	0.411	0.403	0.407	0.403	0.414	0.409
en-	0.182	0.407	0.407	0.406	0.409	0.409	0.407	**0.606**	0.325	0.404
fore-	0.186	0.337	0.398	**0.595**	0.404	0.403	0.161	0.333	0.101	0.403
im-	0.415	0.336	0.401	0.401	0.406	0.413	0.412	0.404	0.405	0.407
in-	**0.786**	0.339	0.356	0.409	**0.878**	**0.869**	0.412	0.400	0.337	0.416
inter-	0.239	0.241	0.411	0.411	0.406	**0.705**	0.406	0.410	0.327	**0.714**
mid-	0.167	0.166	0.415	0.411	0.404	0.411	**0.713**	0.409	0.324	0.406
mis-	0.162	0.408	0.342	0.407	0.412	0.404	0.323	**0.596**	0.411	0.404

non-	0.400	0.397	0.426	0.411	0.425	0.435	0.458	0.406	**0.736**	0.433
out-	0.171	0.343	0.406	0.337	**0.705**	0.410	0.401	0.400	0.405	0.405
over-	0.254	0.331	0.409	0.787	**0.770**	0.409	0.325	0.404	0.410	**0.818**
pre-	0.303	0.412	0.418	0.410	0.409	0.417	**0.839**	0.327	0.335	**0.848**
re-	0.243	0.404	**0.797**	0.400	0.419	**0.810**	0.319	**0.805**	0.409	**0.814**
self-	0.204	0.416	0.402	0.413	0.409	0.413	0.335	0.404	0.342	0.419
sub-	0.212	0.406	0.398	0.404	0.416	0.409	0.405	0.216	0.406	0.410
super-	0.319	0.344	0.419	0.423	0.407	0.410	0.404	0.332	0.414	0.421
trans-	0.179	**0.627**	0.405	0.402	0.415	**0.637**	0.354	0.410	0.329	0.408
un-	0.335	0.339	**0.827**	0.405	0.405	0.410	0.413	0.320	0.417	0.406
under-	0.160	0.407	**0.671**	**0.706**	0.343	0.407	0.327	0.407	0.409	0.403

6.4 Why frequency homogeneity matters

Throughout this chapter, I have presented evidence that clusters of derivations are more easily formed with the bases of similar frequency. Now the question is why it might be the case. Since my hypothesis states that derivational patterns propagate themselves via discourse co-occurrences of bases and derivations, one might think that I would expect bases to collocate mostly with their neighbours in the frequency rank hierarchy. Of course, I do not. But it seems that a similar level of entrenchment of bases results in better connectivity of the network they constitute, which, in turn, serves more reliable information spreading.

This can be illustrated as follows. Suppose that there is a pair of words, [BASE] and [PREFIX+BASE], that frequently appear in the same contexts (like the words *continuous* and *dis-continuous* in Figure 43). By virtue of this, the derivational element [PREFIX] becomes easily recognisable, detachable from its base, and ready to be used to form new derivations. For this process to unfold, [PREFIX] needs to travel across the network of bases that tend to collocate with [BASE], from which it has set off on its journey.

One can liken this to the classical problem of the cat-and-mouse Markov chain (Litvak and Robert, 2012). The problem models a cat and a mouse jumping from one box to another or a mouse moving around a maze that consists of some number of connected rooms. The task is to evaluate the expected time of the mouse's survival, or how long the mouse will spend in each room of the maze if it starts in a certain room. In a similar vein, one can think about a derivational element making transitions from one base to another in a Markov chain, with the transitional probabilities given by the respective elements' co-occurrence patterns. Intuitively, for a cluster of derivations to grow, the respective derivational pattern should be continuously refreshed, which means that a Markov chain of bases and derivations should be irreducible and characterised by a low mean recurrence time (the number of transitions necessary to return to each state from where the Markov chain started; Grinstead and Snell, 2012) for all states.

It is instructive in this regard to consider an example visualised in Figure 48. My data on the words with prefix *mid-* suggest the existence of a cluster of derivations growing around the pair *afternoon* — *mid-afternoon* (cosine similarity of 0.56). This cluster includes, among others, such bases as *climb* and *wash*, whose raw frequencies (1,035,927 and 1,031,603 respectively) are very similar to the raw frequency of *afternoon* (1,026,530). Now one might wonder how do I know that, given the presence of the word *wash* in this cluster, the word *water*, which is not only semantically related to *wash* but also much more frequent than all others (10,899,435), does not belong to this cluster?

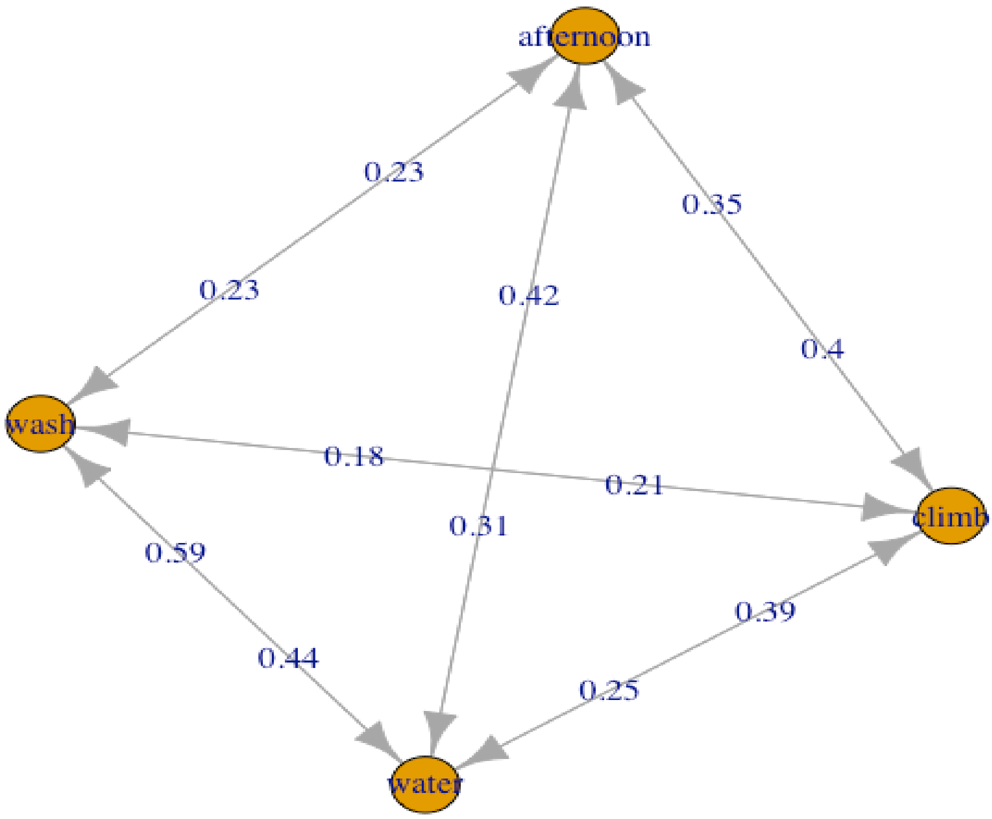

Figure 48. Markov chain including three bases of the same frequency level (*afternoon*, *wash*, *climb*) and one higher-frequency base (*water*)

To answer this question, one should take into account transitional probabilities between different states computed from respective words' cosine similarity values (more on that below) and depicted in Figure 48 along the edges. It is clear that *water*, indeed, is strongly connected to *wash* but much less so to the other members of the cluster. From the Markovian perspective, it means that, if the stationary distribution exists, the system will spend much greater proportion of time in states *water* and *wash*, and the number of transitions necessary to return to each state from where the Markov chain started (mean recurrence time) will be, on average, bigger when this state is included and smaller when it is not. Intuitively, for a cluster to grow, respective derivational pattern should be continuously refreshed, which means that a Markov chain of bases and derivations should be irreducible, with low mean recurrence time for all states.

Now, let us return to the question of why I argue that clusters of derivations are more easily formed with bases of similar frequency. This question might be reformulated as follows. Given any cluster in my data, how do I know that it cannot be extended to include any base of higher or lower frequency than the average frequency of the members of this cluster? My hypothesis is that the observed clusters are balanced in such a way so as to assure irreducibility and minimise the mean recurrence time for all their states.

The criterion of irreducibility implies that the bases of lower-than-cluster-average frequency are less likely to be included in the cluster because such a Markov chain would be reducible, and the signal circulating in the cluster would not reach all states. The criterion of low mean recurrence time implies that the bases of higher-than-cluster-average frequency are less likely to be included in the cluster because such a Markov chain would have a high mean recurrence time (due to the frequent words' wide network of discourse connections), and the signal circulating in the cluster would be forgotten along the way.

To illustrate this idea, some exemplary data are provided in Table 34. Suppose there is a Markov chain 1 where four states are bases that form a cluster of derivations with prefix *mid-* (*afternoon*, *climb*, *wash*, *evolution*) and the fifth state is a derivation with this prefix that most frequently co-occurs with one of the bases (*mid-afternoon*). Suppose further that there are Markov chain 2 and Markov chain 3. In the former, one of the states present in Markov chain 1 (other than the derivation and its base) was randomly selected to be replaced by a random base of a higher frequency rank (*evolution* → *water*). In the latter, the same state was replaced by a random base of a lower frequency rank (*evolution* → *tide*). Now, if one calculates for these chains the value of interest, it will become clear that Markov chain 1, which represents the true cluster, is characterised by the shortest mean recurrence time, averaged across all states.

Table 34. Mean recurrence times for three Markov chains

	fixed states				fixed states' median	added state	all states' median
	mid-afternoon	*afternoon*	*climb*	*wash*			
chain 1 (+ *evolution*)	4.82	2.69	4.38	7.84	4.60	15.14	4.82
chain 2 (+ *water*)	7.48	3.58	6.17	5.66	5.92	4.01	5.92
chain 3 (+ *tide*)	6.98	3.41	5.00	6.11	5.56	4.99	5.00

Table 35. Transition matrix for a random sample

		mid-afternoon	*afternoon*	*climb*	*wash*	*evolution*
mid-afternoon	cos. sim.	—	0.56	0.06	0.01	0.00
	trans. prob.	0.00	0.87	0.10	0.03	0.00
afternoon	cos. sim.	0.56	—	0.33	0.21	0.05
	trans. prob.	0.49	0.00	0.28	0.18	0.05
climb	cos. sim.	0.06	0.33	—	0.16	0.15
	trans. prob.	0.09	0.46	0.00	0.24	0.21
wash	cos. sim.	0.01	0.21	0.16	—	0.00
	trans. prob.	0.04	0.54	0.42	0.00	0.00
evolution	cos. sim.	0.00	0.05	0.15	0.00	—
	trans. prob.	0.00	0.27	0.73	0.00	0.00

In order to see whether this holds true for any cluster in the data, I ran the following simulation. 3,000 random samples labeled as clusters and 3,000 samples labeled as non-clusters were drawn from the data (all prefixes had equal probabilities to be sampled). Each sample included five bases that followed each other in the frequency rank hierarchy. From these bases' derivations, the most frequent one was selected and added to the sample, resulting in six words overall. Each sample was turned into a Markov chain with transitional probabilities approximated by normalising the words' cosine similarity values so that they added up to 1 in each row of the transition matrix (Table 35).

This approach, although naive, nevertheless allowed for modelling the relative likelihoods of making transitions in discourse between any two words that belong to the same sample. Indeed, no matter how many words are added to those in Table 35, the calculated transitional probabilities will only be rescaled and will remain the same with respect to each other.

For each Markov chain, its states' mean recurrence times were computed and averaged. After that, the samples were modified according to the logic described above. The derivation and its base were retained while four other bases were replaced by four randomly selected bases, two of a higher frequency rank and two of a lower frequency rank. Mean recurrence times of the states of these modified Markov chains were computed. As hypothesised, for both types of samples, the latter values were found to be, on average, significantly greater than the former (clusters: $t = -2.5$, $p = 0.01$; non-clusters: $t = -2.63$, $p < 0.01$).

Importantly, the odds of six randomly selected words constituting an irreducible Markov chain are 9.54 times higher for the samples labelled as clusters than for their counterparts. Since in an irreducible Markov chain, the process can go from any state to any state, whatever be the number of steps it requires, and thus all the states in the chain belong to one closed communicating class, it can be confirmed, first, that the groups of the bases I labelled as clusters tend to be discoursally connected with at least one derivation, as well as with each other, and second, that similar level of entrenchment of the bases serves better information exchange within the network they constitute.

6.5 Simulating cluster formation

To show how the process of derivational patterns' spreading might work, I programmed the following computer simulation. First, I created an undirected network $G_0 = (V_0, E_0)$, where V_0 was a set of 995 vertices representing all the bases in my data, and E_0 was a set of unordered pairs of these vertices, such that for any pair of bases v_i and v_j, $i \neq j$, the edge (v_i, v_j) was added to E_0 only if the measure of cosine similarity between v_i and v_j, as calculated by the GloVe model, was found to be greater than 0. This resulted in a very dense (0.80) network with 995 nodes and 396,760 edges.

Next, for each prefix p_i, $i \in \{1, \ldots, 25\}$, I created a network $G_i = (V_0+V_i, E_0+E_i)$, where V_i was a set of the p_i-derivations belonging to the high-frequency group (those with more than 1,000 corpus hits in my data), and E_i was a set of the edges connecting each derivation in V_i to its base in V_0 if the measure of their cosine similarity was greater than 0. All the edges in the network G_i were weighted by the cosine similarity values of their extremities.

After that, the following simulation process was run for each prefix. One node was randomly chosen from the set of nodes in V_i that were considered to be chain initialisers. The probability of each derivation's selection was equal to its relative frequency in the group of high-frequency derivations with the respective prefix, so more frequent words stood a better chance of being drawn.

The possibility of a transition from a selected vertex to one of its nearest neighbours in the network was evaluated by taking all the weights of the edges incident upon the vertex, renormalising them so that they sum up to one, and randomly choosing a candidate from the resulting distribution. It is clear from the structure of the network G_i that if a p_i-derivation had an edge with its base, the transition from the derivation to the base was made with a probability of 1.

The process described above was repeated for each consecutively chosen vertex. Each chain initialised by the randomly selected derivation was limited to 50 transitions; after this, a new derivation was drawn. Overall, I sampled 500 chains for each prefix in my data.

The most important concept for me was that of a 'pattern memory score'. I assumed, in line with my hypothesis, that whenever a derivation and its base co-occur in discourse, the respective derivational pattern is refreshed and remains available for application for some time. However, this is a 'memory-loss' process in the sense that if the invigorated pattern remains unemployed long enough, it is deleted from the operative memory and needs to be retrieved once again.

Hence, whenever a transition from a derivation to its base or vice-versa was recorded during the simulation, I increased the pattern memory score, which was initially set to zero, by five points. This meant that each base reached over the course of five transitions from this moment received +1 to its tally of simulated derivations. The bases that were visited when the pattern memory score equalled zero were passed over with no increase in the number of respective p_i-derivations.

If the above is not a completely inaccurate modelling of the real processes of derivational pattern spreading, then I expect to find that the vertices corresponding to the unattested derivations in my data would be visited significantly less frequently during the simulation than the vertices corresponding to the low-frequency derivations (those with less than 1,000 corpus hits).

Frequency distributions obtained for the derivations with 25 English prefix are plotted against the ranks of the bases in Figure 49. One notable thing about these distributions is their similarity to the distributions of observed frequency counts of derivations with the same prefixes visualised in Figure 40. Most importantly, one can clearly identify the already-familiar multiple-cluster structure, such that the derivation frequency is not proportional to the rank of the respective base for any prefix.

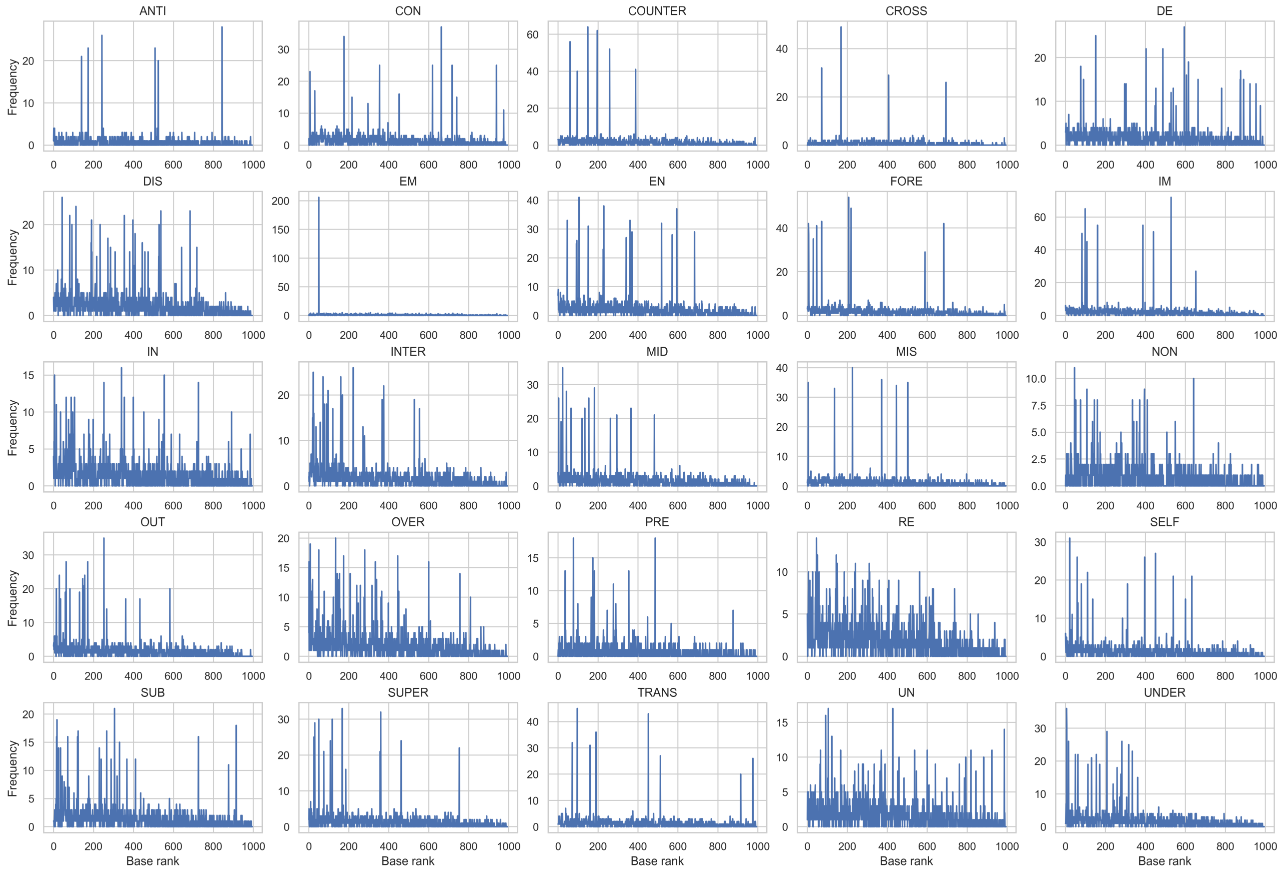

Figure 49. Simulated frequency distributions of derivations with 25 English prefix

One obvious way to assess the accuracy of the simulation process is to correlate the obtained derivation frequency counts with the observed data. For all prefixes, I found a significant positive correlation between the actual and simulated frequencies of derivations, with the coefficients ranging from 0.30 to 0.76. In addition, three frequency groups of derivations are characterised by significant and expected shifts in the locations of their simulated frequency counts, as confirmed by the Kruskal–Wallis test (Table 36; Conover's test showed $p < 0.01$ for all pairwise comparisons).

As previously with probabilistic modelling, one can observe that there is a greater overlap in the distribution of simulated frequency counts between low-frequency and unattested derivations than between either of these and high-frequency derivations. Nevertheless, for each prefix, actual low-frequency derivations are characterised by significantly greater simulated values than their not attested counterparts.

Table 36. Simulated derivations' statistics: coefficients of correlation with observed data, averages of three frequency groups, and H-values of the Kruskal–Wallis test by ranks

prefix	*r* (*p*-value)	M_{HIGH}	M_{LOW}	$M_{UNATTESTED}$	*H* (*p*-value)
anti	.32 (< 0.001)	5.77	0.63	0.40	26.91 (< 0.001)
con	.59 (< 0.001)	11.8	1.35	0.98	50.15 (< 0.001)
counter	.45 (< 0.001)	32.4	1.50	1.05	42.31 (< 0.001)
cross	.30 (< 0.001)	8.52	0.62	0.45	23.59 (< 0.001)
de	.64 (< 0.001)	9.82	1.58	1.12	78.30 (< 0.001)
dis	.75 (< 0.001)	14.11	2.15	1.50	105.8 (< 0.001)
em	.47 (< 0.001)	69.0	1.07	0.79	11.09 (0.003)
en	.76 (< 0.001)	23.36	2.01	1.50	53.27 (< 0.001)
fore	.70 (< 0.001)	27.28	2.04	1.30	55.81 (< 0.001)
im	.76 (< 0.001)	38.90	2.18	1.48	55.49 (< 0.001)
in	.59 (< 0.001)	5.06	1.31	0.64	121.8 (< 0.001)
inter	.60 (< 0.001)	11.77	1.60	1.12	94.65 (< 0.001)
mid	.51 (< 0.001)	14.52	1.47	1.00	72.69 (< 0.001)
mis	.46 (< 0.001)	13.11	1.05	0.76	27.88 (< 0.001)
non	.36 (< 0.001)	2.07	0.64	0.38	37.15 (< 0.001)
out	.66 (< 0.001)	13.10	1.60	0.97	97.18 (< 0.001)
over	.62 (< 0.001)	8.96	1.66	1.04	137.4 (< 0.001)
pre	.42 (< 0.001)	3.74	0.67	0.42	62.80 (< 0.001)
re	.58 (< 0.001)	5.11	1.65	1.10	194.2 (< 0.001)
self	.47 (< 0.001)	9.87	1.17	0.76	80.64 (< 0.001)
sub	.53 (< 0.001)	7.22	1.27	0.90	80.48 (< 0.001)
super	.45 (< 0.001)	13.37	1.31	0.86	53.69 (< 0.001)
trans	.58 (< 0.001)	15.66	1.26	0.89	51.63 (< 0.001)
un	.60 (< 0.001)	6.32	1.49	0.91	167.0 (< 0.001)
under	.62 (< 0.001)	15.18	1.77	1.11	98.83 (< 0.001)

Noteworthy, the accuracy of the simulation was not found to be the same across all prefixes. Specifically, there was some indication of the fact that the strength of association between observed and simulated derivation frequencies tends to be negatively correlated with the prefix's linguistic productivity measured as its probability to combine with a random base (r = -0.36, p = 0.07). This seems to make intuitive sense: the more productive a particular derivational pattern is, the less its probability of application is dependent upon the constant recurrence of invigoration stimuli.

Overall, the results of the simulation are surprisingly good, given the rather bold simplifying assumptions. First, the model was not trained to distinguish between theoretically possible and semantically (or pragmatically) incongruous combinations of prefixes and bases, reproducing exclusively the established discourse co-occurrence network. Second, the pattern refreshment signals were received only from the base-derivation pairings, though frequent co-occurrences of two complex words with the same prefix may give rise to similar effects.

6.6 Diachronic evidence

While programming my model, I assumed that whenever a derivation and its base co-occur in discourse, the respective derivational pattern is refreshed and remains available for application for some time. If the invigorated pattern stays unemployed long enough, it is deleted from operative memory and needs to be retrieved again. That being said, words' collocational preferences are known to be very stable, and the same combinations of words are repeated by language users over and over again. Hence, established networks of derivational element transduction should grow both type- and token-wise over time. If my hypothesis about derivational patterns being more easily applied to those bases that collocate with other members of their morphological family is correct, then, from a diachronic perspective, one would expect to find that the groups of words identified as clusters in my data would gain frequency significantly faster than their non-clustered counterparts over the same period of time.

To check whether the groups of words identified as clusters increased in frequency significantly more than their non-clustered counterparts over the same period of time, I collected frequency counts for all the prefixed derivations in my data from two English web corpora provided by Sketch Engine: *enTenTen15* (2015; 13,190,556,334 words) and *enTenTen20* (2020; 36,561,273,153 words). Given the corpora's different sizes, the numbers of occurrences (hits) per million tokens were obtained for each item. The five years separating two corpora may seem like a short span of time from which to draw reliable conclusions; however, no earlier corpus comparable to the one from which the sample of the bases was obtained was available. In addition, I would argue that, given modern online communication, five years is a sufficient amount of time for productive linguistic patterns to bear fruit.

For each prefix, its derivations' relative frequency counts in 2015 were subtracted from the same derivations' relative frequency counts in 2020. After that, I used Fisher's exact test to determine whether the number of derivations showing any (no less than 0.01 i.p.m.) frequency gain over the span of five years was associated with cluster or non-cluster group. The results are provided in Table 37.

Table 37. Frequency gains of clustered and not clustered derivations

prefixes	type	counts		odds ratio	*p*-value
		clusters	non-clusters		
anti-	gain	29	7	7.93	< 0.001
	no gain	329	630		
con-	gain	3	13	4.78	0.03
	no gain	45	934		
counter-	gain	8	5	13.08	< 0.001
	no gain	107	875		
cross-	gain	18	2	29.82	< 0.001
	no gain	226	749		
de-	gain	27	7	8.49	< 0.001
	no gain	300	661		
dis-	gain	17	15	8.87	< 0.001
	no gain	109	854		
em-	gain	1	1	89.27	0.02
	no gain	11	982		
en-	gain	2	8	2.58	0.22
	no gain	87	898		
fore-	gain	6	5	21.07	< 0.001
	no gain	53	931		
im-	gain	3	9	12.77	< 0.01
	no gain	25	958		
in-	gain	53	9	7.41	< 0.001
	no gain	413	520		

inter-	gain	12	5	6.62	< 0.001
	no gain	260	718		
mid-	gain	27	6	16.01	< 0.001
	no gain	211	751		
mis-	gain	11	3	49.23	< 0.001
	no gain	68	913		
non-	gain	97	2	14.57	< 0.001
	no gain	689	207		
out-	gain	20	3	29.73	< 0.001
	no gain	178	794		
over-	gain	50	8	11.12	< 0.001
	no gain	337	600		
pre-	gain	43	19	3.47	< 0.001
	no gain	368	565		
re-	gain	97	1	63.53	< 0.001
	no gain	542	355		
self-	gain	28	3	12.71	< 0.001
	no gain	408	556		
sub-	gain	40	10	7.77	< 0.001
	no gain	321	624		
super-	gain	34	5	11.20	< 0.001
	no gain	361	595		
trans-	gain	3	6	5.98	0.02
	no gain	76	910		
un-	gain	68	6	12.17	< 0.001
	no gain	444	477		
under-	gain	16	8	7.20	< 0.001
	no gain	211	760		

Remarkably, for all 25 prefixes (including the least productive ones, like *em-* or *con-*), the same pattern is observed: the odds of gaining in frequency are, on average, 11.2 times higher for those derivations that belong to the identified clusters than for those that do not. The difference is significant for all prefixes, with the exception of *en-*. Note that the numbers of clustered lexemes that were used more frequently in 2020 compared to 2015 are almost perfectly correlated with the respective prefixes' productivity values, computed above as their probabilities of combining with a random base (see Table 26): $\rho = 0.94$, $p < 0.001$. However, for non-clustered derivations, no such correlation was observed: $\rho = -0.01$, $p < 0.95$.

For example, this is the case with *de-tune*, the base of which forms a cluster with the verbs *pose* and *regulate*, among others. Here, *regulate* is a strong collocation of *de-regulate* (cosine similarity of 0.5), and both of these words are not uncommonly encountered in the vicinity of *tune* in discourse (cosine similarities of 0.17 and 0.11, respectively), as, for example, in (1) *I got the piano* ***tuned*** *and the action* ***regulated*** or (2) *I am not much a fan of oval track racing, throw some curves into the mix and* ***deregulate*** *a bit and I may start* ***tuning*** *in*. Probably, it is due to this co-occurrence-driven pattern refreshment that the frequency of *de-tune* increased from 0.02 i.p.m. in 2015 to 0.04 i.p.m. in 2020.

6.7 Discussion and conclusion

It is a long-established view that frequently used complex words tend to be structurally and semantically less transparent than infrequently used words (Bybee, 1985, 2007; Hay, 2003). People have argued that since the morphological structure of frequent words is obfuscated, affixes, which are encountered in many frequent items, become less parsable and, thus, lose the ability to combine with new bases (Hay and Baayen, 2002). The current study, however, shows that the reality is more complicated. I argue that high-frequency derivations with an affix, once they are accumulated in a certain number of types, do not block the emergence of new low-frequency coinages but rather facilitate it, paving the way for neologisms. What seems to make the difference in terms of the linguistic productivity of a derivational pattern is not the proportion of infrequent words or parsable words among all words with a specific affix but rather the proportion of high-frequency items that strongly collocate with their bases.

For example, the levels of productivity of the prefixes *de-* and *con-* are apparently different. I assessed their probabilities of combining with a random base as 0.42 and 0.27, respectively (see Table 26). One can find evidence supporting this estimate by looking at the dates of the earliest recorded uses in English of the derivations with these prefixes, which are attested in my samples. It has been noted that dictionary-based measures underestimate actual productivity, at least of productive processes (Baayen and Renouf, 1996; Berg, 2020), so I would not expect to find the majority of the occasionally coined, low-frequency words in my corpus data in even comprehensive dictionaries of modern English. However, it seems natural to expect that, given a random sample of bases, the date of the most recent derivation included in a dictionary will be closer to the current moment in time for a more productive prefix.

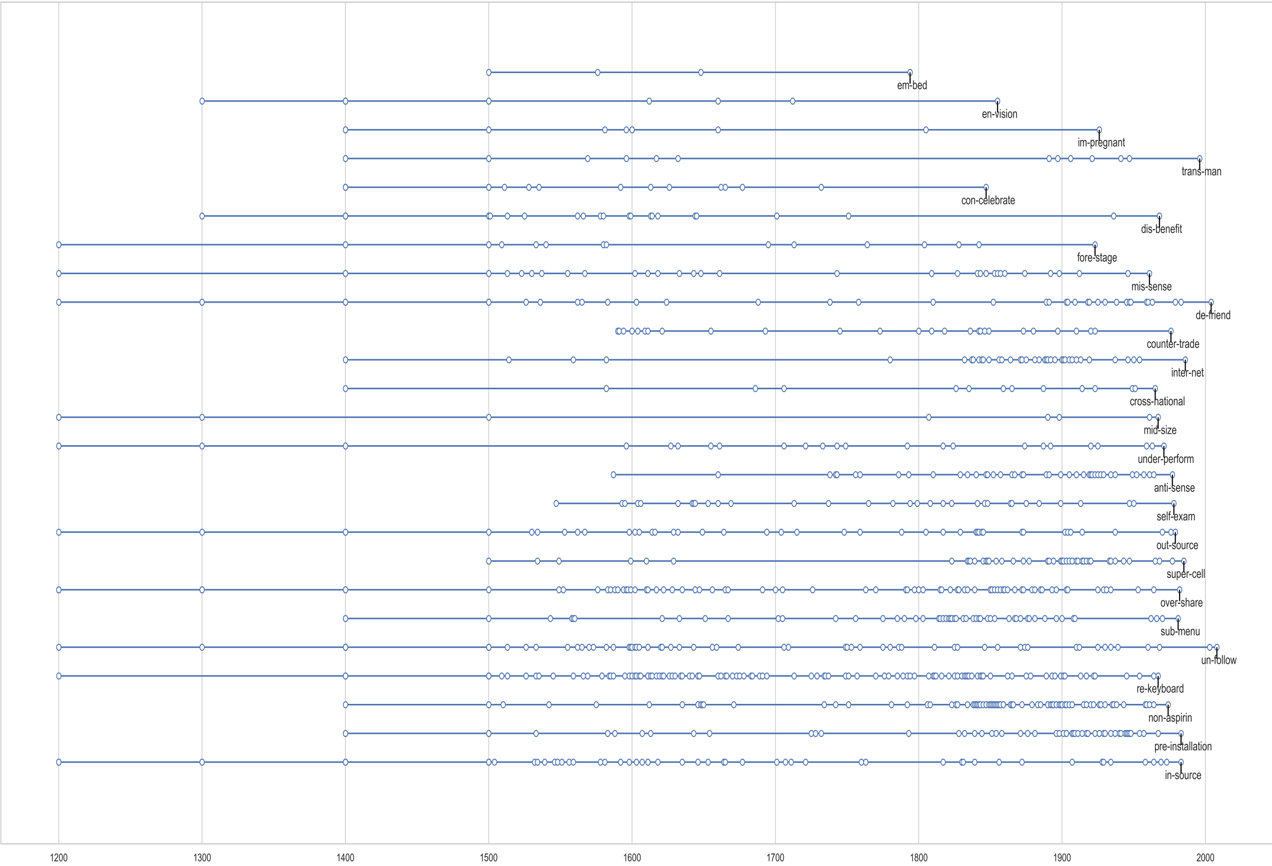

Figure 50. Timeline of the first known uses of the derivations included in Merriam-Webster dictionary

After analysing the first known uses of those derivations with *con-* and *de-* in my data that were included in the Merriam-Webster dictionary (Figure 50), I found that the derivational patterns being compared show significant differences in this regard. The latest derivation with *con-* is *con-celebrate*, the earliest recorded use of which dates back to 1847. The latest derivation with *de-*, *de-friend*, was added to the dictionary much more recently, in 2004. This lends additional support to the belief that the degrees of productivity of the prefixes *de-* and *con-* are not alike.

Importantly, this difference is difficult to explain by taking into account only frequency counts or average levels of parsability of the members of the respective groups. In my random samples, derivations with *de-* include an even greater proportion of frequent (more than 1,000 hits) words than derivations with *con-* (39 vs. 25). On the other hand, the proportion of complex words that are likely to be parsed, as suggested by their derivation to base frequency ratio (Hay, 2001), among all actually attested items with these prefixes is exactly the same: 97%. What really distinguishes my samples of derivations with *de-* and *con-*, apart from the quantity of attested complex words (450 vs. 277), is how many of them are clustered (66% vs. 9%). This, as I have tried to show above, reflects the fact that derivations with *de-*, frequently co-occurring in discourse with their bases, are more numerous, and the average collocational strength in those pairs is greater.

To illustrate this, here are a couple of examples. Confronted with the words *con-tour* and *de-posit*, one who relies only on absolute or relative token frequency would believe that the former contributes much more to the overall productivity of its prefix than the latter: (1) *con-tour* is less frequent (202,361 vs. 1,049,022 hits) and (2) its derivation to base frequency ratio is much smaller (0.07 vs. 12.72). This account, however, falls short of acknowledging the fact that *con-tour* and *de-posit* reveal very different properties with regard to the discourse co-occurrences with their bases (cosine similarities of 0.06 and 0.12, respectively).

Posit and *de-posit* are rather frequently encountered in each other's vicinities, not least because people are interested in understanding the difference in their meanings (these communicative contexts include, among others, dictionary entries; cf. one of the WordNet's synsets: *situate, fix, **posit**, **deposit**: put (something somewhere) firmly*) or want to achieve a rhetorical effect by contrasting them (*Both undo the self-grounding a priori of consciousness and do so in terms of a **de-posited** subject–one that is first **deposited** before **positing** itself and so one that is **de-posited** or displaced in its very position of subject*).

This is evidently not the case with *tour* and *con-tour*. The phonetic similarity (rather than distant etymological relation) of the two may occasionally be used for wordplay, especially with the purpose of branding new products (cf. *The **Contour** 450 is a versatile, mid-sized **touring** kayak and represents a great choice for beginners through to intermediate paddlers*), but the words do not really collocate, and their connection to each other remains obscure for the majority of language users. For this reason, the initial element *con-* in *contour* does not get parsed out and in no way helps the prefix's morphological productivity.

Of course, not all derivations with *con-* in my sample do not co-occur with their bases. For example, the words *con-tend* and *tend* have a very high cosine similarity of 0.53. Judging by how habitually they are used side by side, one would believe that English speakers are fully aware of their historical relationship (*The same people **tend** to **contend** that nothing happens unless it is reported on TV*; *Every male **tends** to **contend** with this, in some way, shape, or form*; *Pratt **contends*** *that researchers* ***tend*** *to view language usage…*; *Wilber* ***contends*** *that Freud* ***tends*** *to reduce the transpersonal to the prepersonal…*; and so on). However, *con-* in *con-tend*, though easily analysable in such contexts, is not applied to other bases of the same level of entrenchment to form new derivations. In my sample, *con-tend* is one of the frequent stand-alone complex words that show no signs of clusters having been formed around them.

In order to clarify what is going on, it is instructive to compare two derivations with *con-*: *con-tend* and *con-version*. The latter is also a strong collocation of its own base, with a cosine similarity of 0.30 (*Different constants are used for the Destination File Format parameter to designate* ***conversion*** *of the source file to different* ***versions*** *of Access*). However, unlike *con-tend*, *con-version* has, in all likelihood, contributed to the emergence of a whole group of derivations with the same prefix, which includes occasionalisms like *con-sense* (*I do not even care about the* ***consense*** *of the words uselessly put together*) or *con-personal* (*Make the relation to God dependent on* ***conpersonal*** *center decisions and experiences*), both *sense* and *personal* being *version*'s close neighbours in the frequency rank hierarchy of my sample.

The difference, again, is that *con-version* attaches to a cluster of base/derivation pairs, members of which usually co-occur in discourse, while *con-tend* does not. This can be easily verified. To do this, I obtained cosine similarity values for (1) all pairwise combinations of bases that I identified as clustered around *con-version* (plus *con-version* itself) and (2) all pairwise combinations of the same number of bases (20) closest to *tend* when arranged in descending frequency order in my sample (plus *tend* and *con-tend*). The former values were found, on average, to be significantly greater than the latter ($M_{con\text{-}version} = 0.37$, $M_{con\text{-}tend} = 0.29$, $t = 7.31$, $p < 0.001$), thus confirming that, for whatever reason (certainly not a semantic one, since no sense or topic relation is observed within either group of words), *version* is more likely to collocate with its frequency neighbours than *tend* with its neighbours.

Equally important as, on average, a higher degree of collocational strength, is the fact that the discourse co-occurrence patterns of the bases clustered around *version* and their derivations with *con-*, when viewed as a Markov chain, serve (1) for a more Gaussian-like distribution of states' visits and (2) for a higher number of visits of the states that correspond to the most frequent derivations and their bases. To provide an example, I chose the same 20 bases grouped around *tend* and the same 20 bases grouped around *version* whose pairwise combinations' cosine similarity values were compared above. Having added the pair *tend*/*contend* to the first set and *version*/*conversion* to the second, I turned both sets into Markov chains in which 22 states corresponded to 22 words and transitional probabilities were approximated by normalising the words' cosine similarity values so that they summed to one in each row of the transition matrix. Then, using the *markovchain* R package (Spedicato, 2017), I sampled two sequences of states coming from the underlying stationary distributions of the Markov chains, with the initial state chosen randomly at the onset of every simulation (Brémaud, 1999).

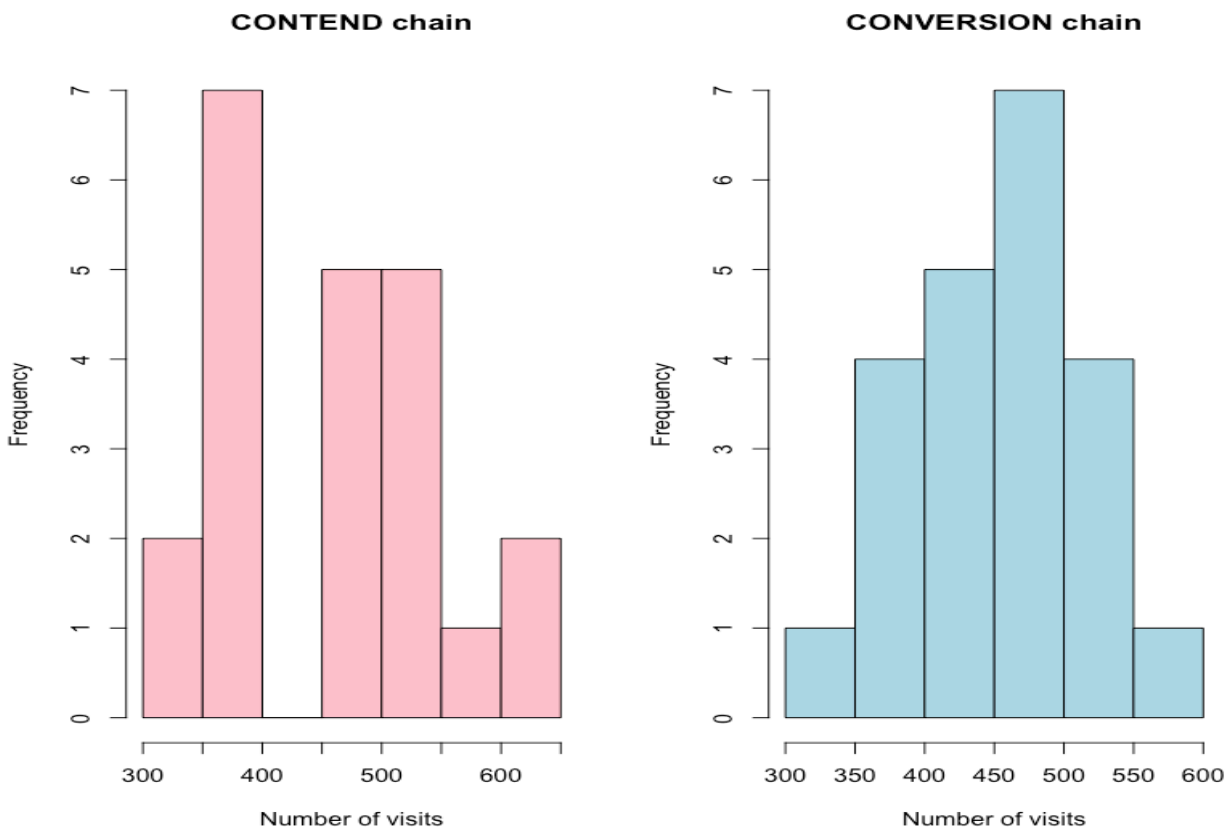

Figure 51. Histograms of the numbers of visits of 22 states in two Markov chains after 10,000 transitions

The two lists of 10,000 items each can be viewed as the result of the modelling of these words' most probable consecutive order of appearance in discourse if one only takes into account their cosine similarities. For example, given that we have encountered the word *asset*, the model predicts that of all the words in the *tend/contend* set, the next to appear will most likely be *spread*. Given that the first two words from this set that we have encountered are *asset* and *spread*, the one to follow will most likely be *limited*, and so on.

One can then calculate how many times each word shows up in its sequence and see whether the two Markov chains are similar with regard to the distribution of the number of visits to different states. As is evident from Figure 51, they are not. The standard deviation of counts in the *con-tend* chain is 1.6 times larger than that of the *con-version* chain (95.12 vs. 58.81). This means that some of the words grouped around *tend* appear very frequently and some very rarely, which, presumably, blocks the process of the derivational pattern's refreshment and reapplication. This is not the case with the *version*/*conversion* chain, where the obtained numbers of visits to different states are clustered more tightly around the mean.

On a final note, the *con-tend* and *con-version* chains differ in one more important respect. The former is constructed from the bases that do not constitute a cluster, that is, the ones that do not have corresponding frequent derivations with *con-*, which means that the *tend*/*contend* pair is the only immediate source of the pattern's invigoration. The latter chain, on the other hand, is comprised of clustered bases, from which it follows that other sources of prefix input are accessible for this particular group of words. As an example, I used the same two sets of words as described above but added one item to each of them. The *version*/*conversion* group was expanded by including the derivation *con-join* with a corpus frequency of 20,514 hits and a cosine similarity of 0.11 with the verb *join*, already in the set. In order to keep the number of states in both chains identical, one base (closest in the frequency rank hierarchy to those already included) was added to the *con-tend* group.

From the extended sets, I constructed two new Markov chains in which 23 states corresponded to 23 words and transitional probabilities were approximated by normalising the words' cosine similarity values so that they summed up to one in each row of the transition matrix. Then, I ran 1,000 simulations on each chain, every time choosing the initial state randomly and recording the first 100 visited states. The differences in the standard deviations of the numbers of visits remained highly significant, with counts from the *con-tend* chain revealing a greater degree of dispersion ($t = 5.73$, $p < 0.001$).

One thing that was conspicuously different, compared to the previous simulation, was the proportion of states visiting of which facilitates pattern recognition. Below are the results of two simulations, one for each chain:

1) CONTEND chain (*tend* — 6 visits, *con-tend* — 1 visit):

camera, limited, hall, majority, struggle, limited, struggle, expand, male, chemical, shoot, ***tend****, limited,* ***tend****, presence, shoot,* ***tend****, download, expand,* ***contend****, vision, distribution, spread, presence, distribution, chemical,* ***tend****, majority, farm, chemical, objective, presence, shoot, expand, limited, download, camera, expand, presence, spread, download, glass, farm,* ***tend****, failure, farm, vision, struggle, farm,* *struggle, quick, majority, quick, limited, distribution, asset, camera, objective, chemical, presence, spread, asset, failure, struggle, failure, vision, camera, chemical, asset, camera, male, hall, quick, secure, distribution, presence, camera, distribution, chemical, male, shoot, glass, secure, objective, presence, camera, male, limited, vision, majority, spread, failure, asset, vision,* ***tend****, expand, distribution, vision, secure, male, distribution*;

2) CONVERSION chain (*con-version* — 3 visits, *version* — 3 visits, *join* — 7 visits, *con-join* — 4 visits):

likely, effort, film, natural, sense, ***join****,* ***conjoin****, difference, simple, relationship, simple, likely, difference, natural, effort,* ***version****, fund, perform, film, relationship, park,* ***conjoin****, road, fund,* ***join****, addition, sale, total, staff, addition, simple, relationship,* ***version****,* ***join****, addition, fund, season, sale, fund,* ***conjoin****, sense,* ***join****, addition, park,* ***join****, relationship,* ***join****, perform,* ***conversion****, particular,* ***conjoin****, film, relationship, sale, park, relationship, difference, season, film, production, staff, difference, park, natural, effort, simple, sense, effort,* ***version****, production, particular,* ***conversion****, sense, likely, particular, fund, production, fund,* ***conversion****, addition, particular, road, addition, total, fund, staff, road, addition, difference,* ***join****, staff, simple, perform, simple, film, season, perform, production, sale, relationship, total*.

Averaged across 1,000 simulations, the difference was as follows: visits of states *tend* and *con-tend* accounted for 9.04 % of all visits in the *con-tend* chain. By contrast, visits of states *version*, *con-version*, *join*, and *con-join* accounted for 16.37 % of all visits in the *con-version* chain. Thus, some sort of self-strengthening effect is obvious here: the more a cluster grows, the more recognisable the pattern becomes, the more frequently existing derivations are used, the more new types are created. It explains why high-frequency complex words with an affix, once they are accumulated in a certain amount of types, do not block the emergence of new low-frequency coinages but rather facilitate it.

7 Parsability revisited and reassessed

7.1 Introduction

In her article ‘Lexical frequency in morphology: Is everything relative?’ from 2001, Hay proposed a simple and elegant way of assessing complex words’ parsability (decomposability). According to Hay, the degree of parsability of a given item depends on the frequency of the derived word relative to its base. With most complex words, the base is more frequent than the derived form, so this relative frequency is less than one. Such words, Hay argues, are more easily decomposed. In the opposite case, when the derived form is more frequent than the base, a whole-word bias in parsing is expected, which has consequences for semantics (such words become less transparent and more polysemous), affix ordering, phonetics (Hay, 2001, 2002, 2003), and morphological productivity (Hay and Baayen, 2002, 2003).

This approach is intuitively appealing and, up until the present day, has been highly accepted in the field (see, for example, Berg, 2013; Pycha, 2013; Diessel, 2019; Saldana, Oseki, and Culbertson, 2021; Zee et al., 2021). However, many researchers who have examined relative frequency effects noted that they exhibit inconsistency and may not hold up across contexts or languages. In fact, over the years, contradicting evidence has been accruing in every domain where relative frequency was believed to play a role.

In phonetics, it is expected that words which are more easily segmentable are less likely to be phonetically reduced (Hay 2001, 2003). However, while some studies indeed found that relative frequency affects, under certain conditions, both affix and base duration (Hay, 2003, 2007; Plag and Ben Hedia, 2018), other studies reported no such effect or even an effect in the opposite direction (Pluymaekers, Ernestus, and Baayen, 2005; Schuppler et al., 2012; Zimmerer, Scharinger, and Reetz, 2014; Ben Hedia and Plag, 2017; Stein and Plag, 2022).

In semantics, relative frequency is viewed as a sign of semantic transparency: if the base is less frequent than the whole form, the output of the derivational process is likely to be less transparent with respect to the semantics of the base (Hay, 2001). However, as a recent distributional semantic study on German event nominalisations discovered, higher relative frequency does not always imply a semantic shift, and conversely, a lower relative frequency is not always associated with semantic transparency (Varvara, Lapesa, and Padó, 2021).

With regard to affix ordering, the so-called parsability or complexity-based ordering hypothesis implies that more parsable affixes do not occur within less parsable affixes because the attachment of a less separable affix to a more separable one is difficult to process (Hay, 2002; Hay and Plag, 2004). However, research on suffix combinations in Bulgarian has shown that Bulgarian suffixes are indeed hierarchically ordered, but the hierarchy they constitute cannot be explained by parsability (Manova, 2010).

Within the domain of morphological productivity, it is a general assumption that there exists a direct positive relationship between the proportion of tokens with a certain affix which are parsed and the productivity of this affix (Hay and Baayen, 2002). Surprisingly, Pustylnikov and Schneider-Wiejowski (2010), after applying the same hapax-based productivity measure that Hay and Baayen (2002) used to several German suffixes, got for the suffix *-nis*, which has almost fallen out of use, a much higher productivity (and hence parsability) value than for the other three noun-forming suffixes under comparison: *-er*, *-ung*, and *-heit/-keit*. This is both counter-intuitive from a language user perspective and contradictory to the traditional view (Lohde, 2006).

Even the very notion of the different perceived complexities of words with high and low derivation to base ratios did not hold universally across languages. For example, in an experiment on Spanish complex words designed in the same manner as proposed by Hay (2001), native Spanish speakers did not rate derived forms with more frequent bases as more complex than derived forms with less frequent bases. However, for L2 Spanish speakers, the base frequency of a derived form did affect decomposition (Deaver, 2013).

In general, relative frequency account remains somewhat controversial. For the purposes of my study, I will additionally outline three issues with the way the derivation to base frequency ratio is usually calculated. First, the resulting value is undesirably dependent upon the absolute frequency of derived forms. As noted by Bybee, ‘at extremely high token frequencies, loss of analysability and transparency will occur independently of relative frequency’ (Bybee, 2010: 46). In other words, when determining a word’s parsability status by calculating its derivation to base frequency ratio, one will always be biased towards judging high-frequency derivations as holistic and low-frequency derivations as decomposable.

Second, Hay treated each word as parsable or non-parsable viewing it in isolation, just by comparing its base and derived frequencies. However, given a whole family of affix-base constructions, it is important to take into account that each affix may combine with multiple bases and each base may combine with multiple affixes. This means that to determine how likely a particular word is to be parsed, more information is necessary besides just its own base and derived frequency. This line of argumentation has been developed in the literature and is usually referred to as the morphological family size effect (Cole, Beauvillain, and Segui, 1989; Schreuder and Baayen, 1997; De Jong, Schreuder, and Baayen, 2000).

The third problem with the derivation to base frequency ratio is that it is undefined for morphological constructions with real affixes and nonce bases. For example, compare the following two words: *sub-measles* and *sub-banksit*. Both of them have frequencies of 0 in the numerator of Hay's equation. Regarding the denominator, the frequency of *measles* is (necessarily) a positive integer, but the frequency of *banksit* is zero. After doing the math, one must conclude that *sub-measles* is very likely to be parsed. The status of *sub-banksit*, however, remains unclear.

This is an unwelcome result taking into account that similar constructions in Russian, with existing prefixes and fictional bases, were rated by participants as semantically transparent and, when given in context, correctly substituted by real words (Monakhov, 2021). Monakhov argued that many Russian prefixed verbs are in reality parsable, despite their derivation to base frequency ratios of greater than one, as in the following examples: *za-kavychitj* 'put in quotes' (148/31), *za-hmeletj* 'get tipsy' (5719/1438), *za-materetj* 'mature' (1681/499), and so on. He claimed that this phenomenon can be explained if we agree that these and many other verbs are not separate lexemes but rather instantiations of one construction with a fixed prefix and an empty slot that can be filled with any relevant lexical material.

These numerical problems seem to be indicative of some conceptual complications. The theory underlying Hay's experiment was that of the dual-route model of perception (Frauenfelder and Schreuder, 1992; Marslen-Wilson and Tyler, 1998; Clahsen, 1999; Pinker and Ullman, 2003; Ullman, 2004; Silva and Clahsen, 2008). This model assumes that speakers might try to decompose a complex word into its parts or access it as a whole. A frequent whole-word representation would speed up the holistic route while a frequent base would facilitate the decomposed route, that is, make the word more likely to be parsed into its constituent parts. Words that are frequently accessed via the decomposed route have their decomposition reinforced. Those that are frequently accessed via the whole word route are felt to be less decomposable.

An alternative interpretation is proposed within the framework of construction morphology (Booij, 2010b), where complex words are seen as constructions on the word level. The view that complex words instantiate morphological constructions can be found in Croft (2001) and Goldberg (2006). Some examples of the constructional analysis of complex words are the analysis of English *be*-verbs in Petre and Cuyckens (2008), the analysis of the phrasal verbs of Germanic languages in Booij (2010a), and the analysis of Russian prefixed verbs in Monakhov (2021).

The main difference between the two approaches, as I see it, is in the allowance for one additional meaning processing mechanism, which construction morphology can make due to its ability to distinguish between fixed elements and slots (variables) (Culicover and Jackendoff, 2005; Jackendoff, 2008; Booij, 2010b; Diessel, 2019). Very simply, for a two-element complex expression — for example, a prefix or particle verb — one can have four possible combinations: (1) both elements are fixed, (2) both elements are variables, (3) the first element is a variable and the second element is fixed, and (4) the first element is fixed and the second element is a variable. Linguistic items of type (1) are non-analysable, non-compositional, and non-productive. They are listed diachronic relics that are not assembled on the fly but are retrieved from the lexicon. Linguistic items of type (2) are, in contrast, analysable, fully compositional, and productive. Up to this point, there is really no divergence between the dual-route model and construction morphology accounts. However, with types (3) and (4), which can be conceptually merged since they differ only in the linear order of elements, the situation is more interesting.

Linguistic items of types (3) and (4) are analysable and (semi?)productive (Jackendoff, 2002) and yet, with regard to their semantics, cannot be called either compositional or non-compositional. They cannot be called compositional in the traditional, Langacker's (1987) sense since their general meaning cannot be inferred from the meaning of their components. Yet it feels somewhat awkward to call them non-compositional because, often, their fixed elements make the same semantic contribution in multiple words (e.g., *around* in the sense of 'not achieving much' in *fiddle around*, *play around*, *fool around*, *mess around* and others listed in Larsen, 2014; cf. McIntyre, 2002).

<table>
<tr><td rowspan="2">non-analysable</td><td colspan="3">analysable</td></tr>
<tr><td>compositional</td><td colspan="2">parsable</td></tr>
<tr><td>type (1)</td><td>type (2)</td><td>type (3)</td><td>type (4)</td></tr>
</table>

Figure 52. Schema of analysability / compositionality / parsability relationship

Moreover, it is well-known that German, Russian, and English non-spatial complex verbs with a certain preverb, prefix, or particle often come in groups of numerous members such that the meanings of derivations are almost identical and yet the meanings of the bases might have nothing in common (Stiebel, 1996; Zeller, 2001). Thus, it makes more sense to call complex linguistic expressions of type (2) compositional and complex linguistic expressions of types (3) and (4) parsable, putting a strong emphasis on the fact that all of them are analysable as opposed to expressions of type (1) (Figure 52).

These considerations allow us to better understand the non-linear relationship between analysability and semantic transparency. The former notion seems to imply the latter (Bauer, 1983; Plag, 2003, 2018; Dressler, 2005; Varvara, Lapesa, and Padó, 2021). Since all linguistic units are form-meaning pairings, our ability to break a complex form into a number of simpler forms crucially depends on our ability to assign meanings to these forms: 'for compositionality to go through, it is necessary that each item in the lexicon is associated with a fixed number (…) of discrete meaning chunks, only one of which is selected in the compositional process' (Taylor, 2012: 42).

Thus, on the surface, there is an interdependency: any analysable pattern is compositional in meaning (Hay, 2001, 2003) and is also linguistically productive (Hay and Baayen, 2002). However, in reality things do not seem to be aligned as conveniently. For example, as noted by Bybee, 'compositionality can be lost while analysability is maintained, indicating that the two measures are independent' (Bybee, 2010: 45). Bybee provided examples of idioms (*pull strings*) and compounds (*air conditioning*, *pipe cleaner*), but with multi-morphemic words, this tendency is no less prominent. In fact, taking into account the idea of the lexicon-syntax continuum in construction grammar (Hoffmann and Trousdale, 2013), it is possible to say more about the relationship between analysability and compositionality than just stating that they are independent of one another.

Thinking about a continuum of possible linguistic item combinations where one pole is occupied by mono-morphemic words and the other by combinations of words, with multi-morphemic words falling in between (see the fragment labelled 1 in Figure 53), it becomes clear that there is a linear increase in compositionality accompanying movement from left to right along the *x*-axis (e.g., *arm* → *forearm* → *arm and leg*). However, the parsability trend is most appropriately modelled with an inverse parabola. It is impossible to talk about the parsability of mono-morphemic words, but it also seems unnatural to call any combination of words that do not constitute a single concept (semantically) parsable (though, of course, this combination remains perfectly morphologically analysable).

The pole of the combination of words is a continuum of its own (see the fragment labelled 3 in Figure 53). This sub-continuum is structured very much like the top-level one. Again, moving from the pole of fixed phrases (e.g., *hapax legomenon*) through the middle point of collocations (e.g., *opera house*) to the pole of free combinations of words (e.g., *two words*), there is linear growth in the compositionality. Again, only collocations here are parsable in the traditional, Langacker's sense.

Finally, this logic can be extended to the sub-continuum of multi-morphemic words (see the fragment labelled 2 in Figure 53). Among them, one can easily distinguish between 1) words of type (1) that behave like idioms, where both elements are fixed (e.g., *con-tact*); 2) words of type (2) that can be thought of as free combinations of morphemes, where both elements are slots (e.g., *non-linear*); and 3) words of types (3) and (4) that resemble collocations, with one element fixed and another one free to vary (e.g., *un-couth* or *em-power*).

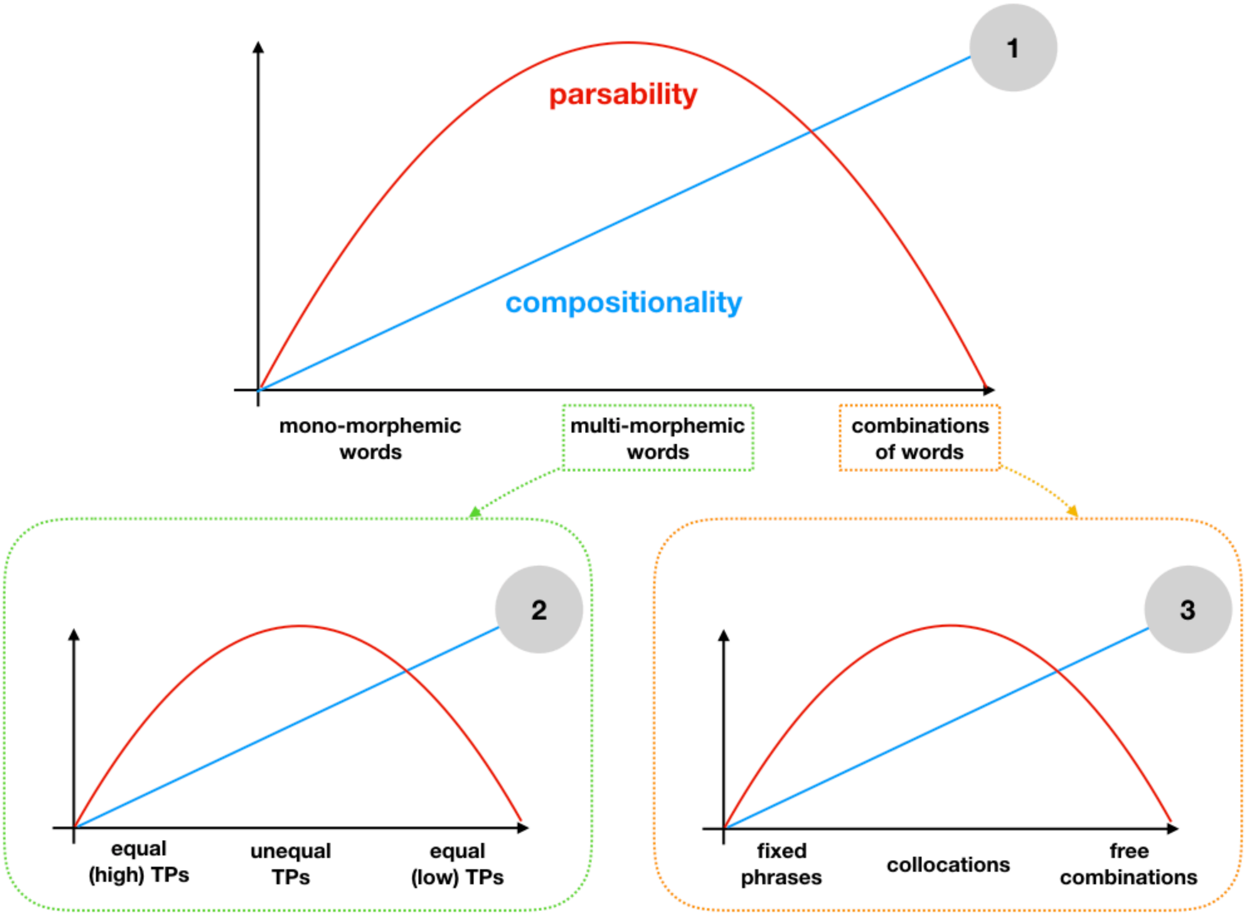

Figure 53. Schema of parsability / compositionality relationship

In general, all of the above implies that parsability and compositionality account for two different models of meaning processing that can be described as follows: parsability as

$$Meaning_{ITEM} = Meaning_{COMPONENT\ 1} + X$$

and compositionality as

$$X = Meaning_{COMPONENT\ 1} + Meaning_{COMPONENT\ 2},$$

where X denotes a semantic element that is not readily available and must be obtained by solving the respective equation.

Returning back to the dual-route model, one might hypothesise that under its account of complex words, parsable expressions of type (3) will most likely be conflated with compositional expressions of type (2), and parsable expressions of type (4) will most likely be conflated with those of type (1). The first is to be expected because type (3) derivations would normally strongly overlap with their bases in semantics and distribution, thus facilitating compositional analysis. For example, in German and Russian complex verbs of this type, the base bears the main burden of lexical meaning, while the preverb shapes and categorises this meaning in terms of primitive semantic concepts (Biskup, 2019).

On the other hand, linguists who have not studied preverbs and verb particles in detail would probably conflate complex verbs of type (4) with those of type (1), (mis)analysing their fixed elements as signalling nothing but telicity (which is clearly wrong even with some so-called 'perfective' particles, cf. *they* {*beat me up* / *hosed the wall down*} *for ten minutes*). Type (4) verbs look more idiomatic than they are because they mostly encode non-spatial meanings, and their preverbs or particles, which are sometimes labelled 'adjunct-like' in the literature, do not fulfil normal arguments of the base verb (Stiebels, 1996; McIntyre, 2007).

As will be shown later, the risk of conflating different populations and glossing over important distinctions becomes prominent when one tries to measure complex words' analysability degrees by calculating their derivation to base frequency ratios. In a sense, the very design of the constructions of types (3) and (4) predetermines the relative frequency relation between the whole form and the base. Since one fixed element normally appears in many words, combined with different elements that fill the respective construction's empty slot (as in Russian *na-pisatj* 'write on', *v-pisatj* 'write in', *nad-pisatj* 'write above', *pod-pisatj* 'write under'), it is expected that in complex words of type (3), where the base is fixed, the derivation to base frequency ratio will tend to be less than one. In contrast, complex words of type (4), where the base serves as a filler (as in German *auf-klären* 'clear up', *auf-bessern* 'polish up', *auf-schaukeln* 'build up', *auf-modeln* 'spruce up'), will most likely reveal derivation to base frequency ratios greater than one.

One way to overcome the conflation problem is to think about complex words' analysability patterns in terms of transitional probabilities, both forward- and backward-going (Pelucchi, Hay, and Saffran, 2009). Thus, for a specific complex word, one would ask, how likely it is that this particular base would be combined with this affix, and how likely it is that this particular affix would be combined with this base? In other words, the goal is to estimate two probabilities: P (affix | base) and P (base | affix). These probabilities can be obtained empirically as relative frequencies, for example, by taking all affixed words in a morphemic dictionary of the respective language and looking up frequencies of interest in the internet corpus of this language. Then, for any word, its P (affix | base) = number of word's tokens / number of tokens of all words with this base and P (base | affix) = number of word's tokens / number of tokens of all words with this affix. It is clear from the first formula, that the derivation to base frequency ratio, when calculated morphological family-wise (see below), is equal to P (affix | base) (Lewis, Solomyak, and Marantz, 2011). This is yet another illustration of the aforementioned conflation problem because complex words of types (1) and (4) reveal equal probabilities of transition from base to affix, as do complex words of types (2) and (3). Two types of constructions within each pair can only be differentiated by taking into account the forward-going transitional probability P (base | affix) (see Figure 54).

Applying the formulae to the two previously given English examples yields the same probability estimations for both of them: P (*sub* | *measles*) = P (*sub* | *banksit*) = 1, P (*measles* | *sub*) ≈ P (*banksit* | *sub*) → 0, which confirms our intuitive belief that complex words with a nonce base and a base that is frequent by itself but extremely unlikely to appear in the empty slot of this particular construction should be equally analysable.

From these considerations, it logically follows that expressions of type (1) will be characterised by comparably high probabilities of transition from affix to base and from base to affix and expressions of type (2) will be characterised by comparably low probabilities of transition in both directions. For expressions of types (3) and (4), these probabilities will diverge. In type (3), where the first element is a variable and the second element is fixed, the probability of transition from base to affix will be low while the probability of transition from affix to base will be high. Conversely, in type (4), where the first element is fixed and the second element is a variable, the probability of transition from base to affix will be high while the probability of transition from affix to base will be low. This discrepancy should come as no surprise since, intuitively, one expects to find that the fixed element communicates less information about the filler than the filler about the fixed element (cf. Gries and Stefanowitsch, 2004).

The expected pattern is schematised in Figure 54 (the two-letter abbreviations proposed therein will be used as a shorthand for respective construction types throughout the rest of the chapter). The continuous nature of probability values makes it clear that there are no distinct classes of complex words with regard to their analysability. Rather, these values indicate how likely each particular word is to be processed according to the respective construction template.

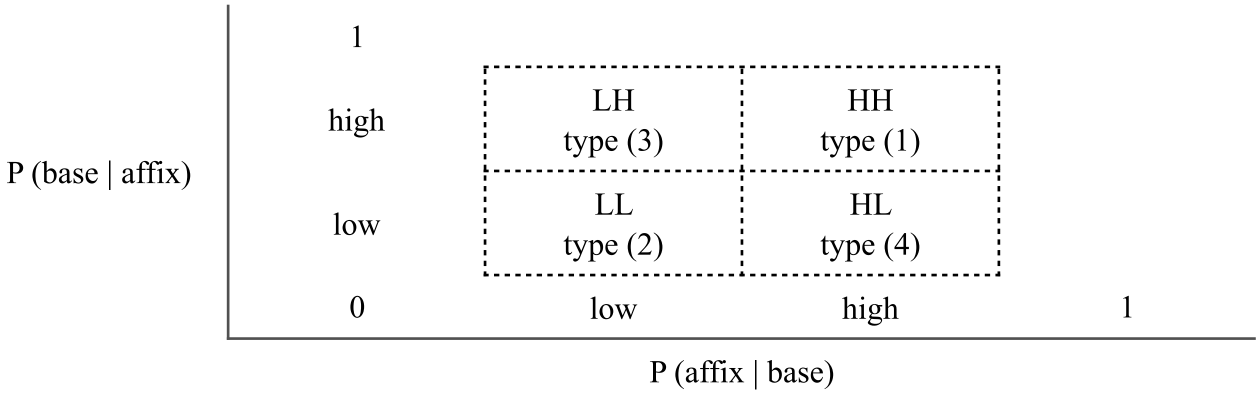

Figure 54. Schema of complex words' transitional probabilities' patterns

In order to combine the two transitional probabilities into one simple numerical measure, one would use the log ratio P (affix | base) / P (base | affix). Given what has already been discussed, the distribution of these measures is expected to be of the following form:

$$LH < -\delta < HH < 0 < LL < \delta < HL,$$

where δ is some positive real number (for the experimental purposes of the current study, it was set to 1).

The rest of the chapter is dedicated to probing into the cognitive reality of the four conjectured construction types. Specifically, I am interested in whether language users perceive complex LH and HL words differently than HH and LL words with regard to their morphological analysability and semantic transparency. Study 1 provides some experimental evidence supporting this claim. The experiments, carried out on English and Russian data, draw heavily on the experimental design proposed by Hay (2001) and on the idea that analysable words are conceived of as more complex — that is, able to broken down into smaller, meaningful units.

In study 2, I address the question of the relationship between two ways of measuring complex words' degrees of analysability: by calculating their derivation to base frequency ratios and by calculating the log ratios of their elements' transitional probabilities. By means of probabilistic modelling and partial replication of Hay's original experiment (2001) I show how the former method might lead to the conflation of different construction types and thus obfuscate the difference between two meaning processing models: one based on the principle of compositionality and another on the principle of parsability.

Finally, in study 3, based on empirical corpus data, I show that the relationship between analysability and productivity is not linear, as it has been frequently described. In fact, two types of analysability might reveal two opposite directions of association with the linguistic productivity of a certain affix. Thus, the preponderance of parsable but not compositional words among the derivations with this affix might serve as a sign that its overall applicability is limited.

7.2 Study 1: Perceived complexity of the different types of complex words

7.2.1 Collecting stimuli

If my hypothesis about the existence of four different construction types governing the processing of two-element complex words were correct, I would expect to find that language users assess respective words' complexity differently. Specifically, if one assumes that subunits of complex words are more easily recognisable when they are variables within particular constructions, then one needs to take into account that HH words have no open slots, both LH and HL words have one, and LL words have two. Therefore, given that LL words are more complex than HH words, it naturally follows that LH and HL words should be placed somewhere in between in this hierarchy.

For each language, I selected 40 stimuli: eight prefixes of different linguistic productivity (this will be described further below) and five construction types (HH, LH, HL, LL plus one pseudo-affixed word) with each prefix. Words were matched for the number of morphemes, and every effort was made to match them for junctural phonotactics, stress patterns, syllable counts, and the frequency of the derived form as well. However, in some cases, not all restrictions could be applied simultaneously. It is possible that the polysemy/homonymy of prefixes could have a non-negligible effect on the results. Nevertheless, since in Russian, LH and HL constructions are semantically differentiated (see below), controlling for their meanings seemed infeasible, and so I did not do this for English data either.

Words were assigned to construction types based on the values of their transitional probabilities' log ratios: (1) LH: log ratio < -1 (with the exception of *out-*, for which -0.73 was the lowest value in my data), (2) HH: -1 < log ratio < 0, (3) LL: 0 < log ratio < 1, and (4) HL: log ratio > 1. The frequencies used to calculate transitional probabilities were obtained from two internet corpora provided by Sketch Engine (Jakubíček et al., 2013): English Web 2018 corpus (*enTenTen18*, more than 21 billion words) and Russian Web 2017 corpus (*ruTenTen17*, more than 9 billion words). English stimuli can be found in Appendix 3, and Russian stimuli in Appendix 4.

One can make some important observations concerning two types of derivation to base frequency ratios by looking at the tables with frequencies and transitional probabilities in the appendices. Not controlling for morphological family size and taking into account only occurrences of the base as a free element leads to a highly unstable and unbounded measure which ranged from 0.003 to 499.4 in my English data and from 0.008 to 55.7 in my Russian data. This is not to mention the dubiousness of the assumption that modern language users are aware of the historical links between some bases and their derivations, for example, that between *tact* and *contact*.

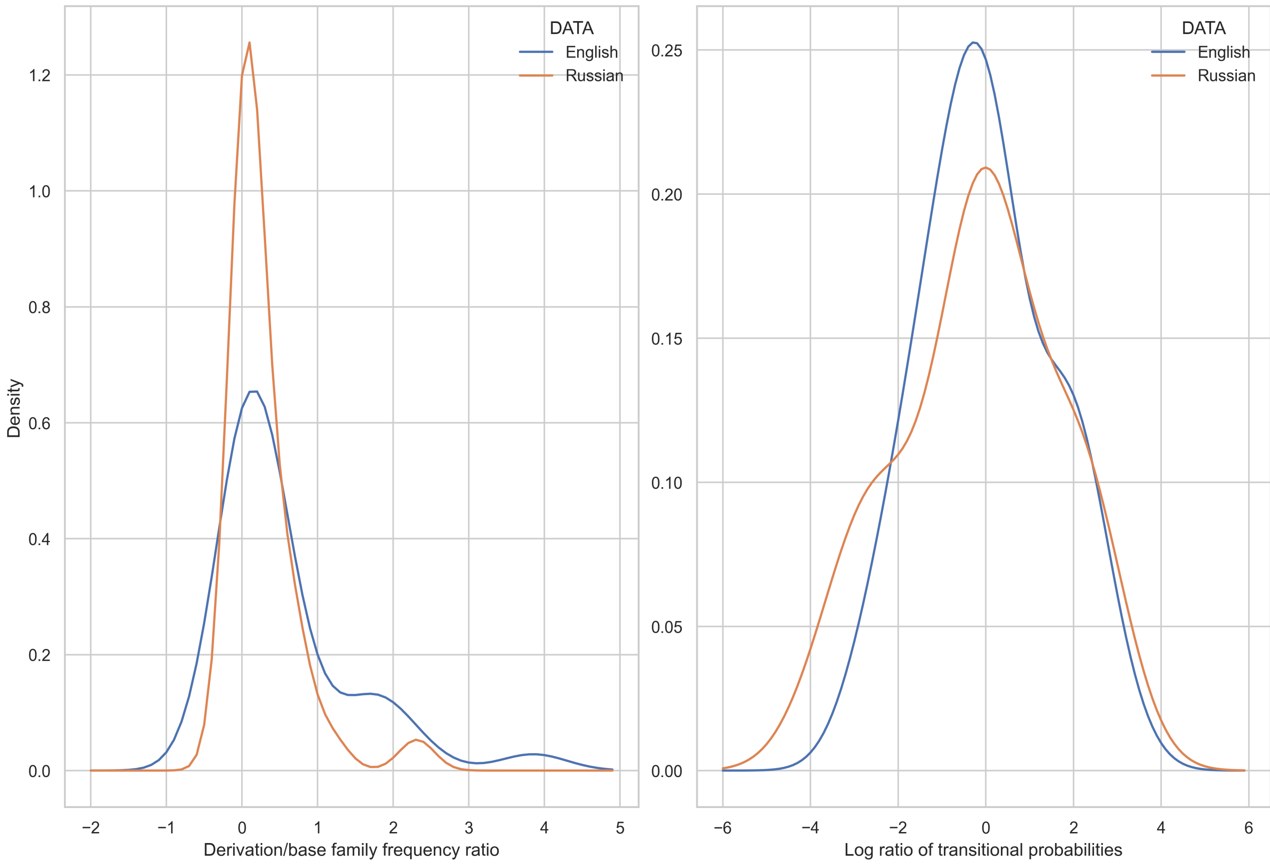

Figure 55. Densities of derivation to base family frequency ratios (left panel) and transitional probabilities' log ratios (right panel), English and Russian

Note: The estimated density curves on the left panel extend to values smaller than zero because of the software's smoothing algorithm (Waskom, 2021).

One can also estimate the base frequency of a particular derivation in a different way, calculating first the cumulative root frequency (Cole, Beauvillain, and Segui, 1989), that is, the overall frequency of all lemmas in which this base occurs, either in a free or bound form, and then subtracting the frequency of the derivation itself (De Jong, Schreuder, and Baayen, 2000). This method has some desirable properties. First, it allows to stabilise the derivation to base frequency ratio. For example, the variance of values in my data was reduced by a factor of 11,075 for English and a factor of 435 for Russian. Moreover, by treating each base as a representative of the whole morphological family, one does not gloss over the fact that speakers, for example, might be able to parse the element -*cede*- out of *precede* not only because it exists as a free form but also (and more importantly) because they encounter it in multiple words with related meanings (e.g., *accede*, *concede*, *recede*, *secede*, etc.).

On the other hand, the actual analysability measures of the complex words in my sample, when calculated as derivation to base family frequency ratios, revealed for both languages the multimodal distribution I expected to find (Figure 55, left panel). The form of this distribution, as well as the results of Kruskal-Wallis and Dunn's tests (not reported here), suggest that parsable expressions of the LH type might indeed be conflated with compositional expressions of the LL type and parsable expressions of the HL type with non-analysable expressions of the HH type. The values of the transitional probabilities' log ratios are, in contrast, normally distributed (Figure 55, right panel) as verified by the Kolmogorov-Smirnov test for goodness of fit in Table 38.

Table 38. Results of the Kolmogorov-Smirnov test for standard normal distribution

	English	Russian
derivation to base family frequency ratios	$D = 0.5$, $p < 0.001$	$D = 0.5$, $p < 0.001$
transitional probabilities' log ratios	$D = 0.18$, $p = 0.19$	$D = 0.20$, $p = 0.10$

7.2.2 Experimental design

The experiments, drawing on the design proposed by Hay (2001), were conducted on English and Russian data. Subjects were presented with pairs of prefixed words and asked to provide intuitions about which member of the pair was more easily decomposable. Experiments participants were gathered via the Amazon Mechanical Turk crowdsourcing platform for English part and the Yandex Toloka crowdsourcing platform for Russian part. The experimental designs for both languages were identical. For English subjects, I repeated the instructions verbatim as they were given in Hay (2001), and for Russian participants, I simply translated them into Russian, having only changed the language examples.

Neither Amazon Mechanical Turk nor Yandex Toloka grant access to their workers' personal data, but they do allow for some coarse-grained social stratification while assembling pools of users. Each word pair in my data was evaluated by 24 native speakers of each respective language, and each set of participants was constructed in such a way so as to conform to the matrix in Table 39 below.

Table 39. Participants' matrix for each word pair

Age	Bachelor's degree (Higher education in Russian part)		High school diploma (Secondary education in Russian part)	
	Male	Female	Male	Female
18–25	2	2	2	2
30–35	2	2	2	2
45–55	2	2	2	2

Both experiments were completed online. Each participant was presented with just one pair of words sharing the same prefix (or pseudo-prefix coinciding with it in form) and asked to type in the word they thought was more complex. Task completion time was not limited (the average duration was 43 seconds for the English part and 53 seconds for the Russian part). Participants were explicitly urged to rely solely on their language intuition and not to consult with any online sources available to them, as there was no such thing as a 'correct answer' in this case. After submitting their task, each participant was rewarded with \$0.02 (which is a usual reward for such simple tasks).

Each word was paired with three of its counterparts of the same prefix and with one pseudo-affixed word with the initial element resembling this prefix. The order of presentation was randomised, so every word had an equal probability of being the first or the second member of the pair to appear. Overall, there were 1,920 English participants and 1,920 Russian participants:

$$\binom{5}{2} \text{ word combinations} \cdot 8 \text{ prefixes} \cdot 24 \text{ subjects} = 1920.$$

7.2.3 Results

Each word pair in my data was evaluated by 24 different people, and so any word in any pair could theoretically win from zero to 24 of these contests. These results lend themselves to different types of analysis. One could treat the overall number of won contests as a score assigned to a word or as the number of people who voted for it. However, this way of reasoning leads to an undesirable loss of information: for example, by only taking into account the fact that nine participants selected a certain word, the valuable information that the other 15 participants, when confronted with this particular pair, gave their preferences to the word's counterpart would be missed.

For this reason, I decided to approach the problem somewhat differently and treat each stimulus in my data as a Bernoulli trial in which each word might win or lose, depending on the probability of success associated with its construction type. Thus, each word, when tested against a word of a different construction type, participated in 24 independent Bernoulli trials with equal probability of success, and the outcome followed the Binomial distribution $X \sim B(n, p)$. Given the obtained experimental results, one could apply the Bayes theorem to calculate the posterior of the probability of success for each word. For example, let us assume that our prior belief is that for all words of a certain construction type, the probability of success θ is equally likely to be 40% or 60%. Given a word that was selected as more complex by nine out of 24 people, it is possible to estimate which value of θ is more likely: 40%,

$$\Pr[\Theta = 0.4 \mid X = 9] = \frac{\Pr[X = 9 \mid \Theta = 0.4]\Pr[\Theta = 0.5]}{\Pr[X = 9]} = \frac{\binom{24}{9}(0.4)^9(1-0.4)^{24-9}(0.5)}{\binom{24}{9}(0.4)^9(1-0.4)^{24-9}(0.5) + \binom{24}{9}(0.6)^9(1-0.6)^{24-9}(0.5)};$$

or 60%,

$$\Pr[\Theta = 0.6 \mid X = 9] = \frac{\Pr[X = 9 \mid \Theta = 0.6]\Pr[\Theta = 0.5]}{\Pr[X = 9]} = \frac{\binom{24}{9}(0.6)^9(1-0.6)^{24-9}(0.5)}{\binom{24}{9}(0.4)^9(1-0.4)^{24-9}(0.5) + \binom{24}{9}(0.6)^9(1-0.6)^{24-9}(0.5)}.$$

This calculation yields the probabilities 0.92 and 0.08, respectively, which confirms our intuitive feeling that if a word was selected just nine times out of 24, then its probability of success should be smaller than 0.5.

However, I am interested not in the θ's point estimates for particular words but rather in the complete probability distributions of θs for the construction types these individual words belong to. That is why I used the Markov chain Monte Carlo sampling approach to construct the following posterior distributions: θ_{LH}, θ_{HH}, θ_{LL}, and θ_{HL}. I expected to find, for both English and Russian words, that the posterior distribution of the probability of success θ_{HH} was centred at some point significantly below 0.5, the posterior distribution of the probability of success θ_{LL} was centred at some point significantly above 0.5, and the probabilities of success θ_{LH} and θ_{HL} were centred somewhere between these two extremities but above 0.5.

For inference, I used the beta-binomial model with the prior on θ specified as coming from the Beta distribution with the following shape parameters: $a = 2$ and $b = 2$. This is analogous to the statement that I expect to see two successes and two failures in a total of four experiments. Thus, I used a non-informative prior that would be easily overwhelmed by the acquired evidence (Neapolitan, 2004). Having performed the inference, I sampled 2,000 θs from the four posterior distributions of interest constructed for each language. The English results are visualised in Figure 56, and the Russian results are in Figure 57. The means and highest density intervals of θs are provided in Table 40. The results for the individual construction pairings can be found in Table 41 (English) and Table 42 (Russian).

Some remarks on notation include the following: (1) Tables 41 and 42 should be read by rows. For example, 0.61 in cell LH—HH of Table 41 means that 61% of participants judged LH words to be more complex when compared to HH words. (2) The numbers in each pair of cells with the reversed positioning of construction labels (e.g., LH—HH and HH—LH) should add up to 1, except for the round-off error. Finally, (3) PA stands for the pseudo-affixed type.

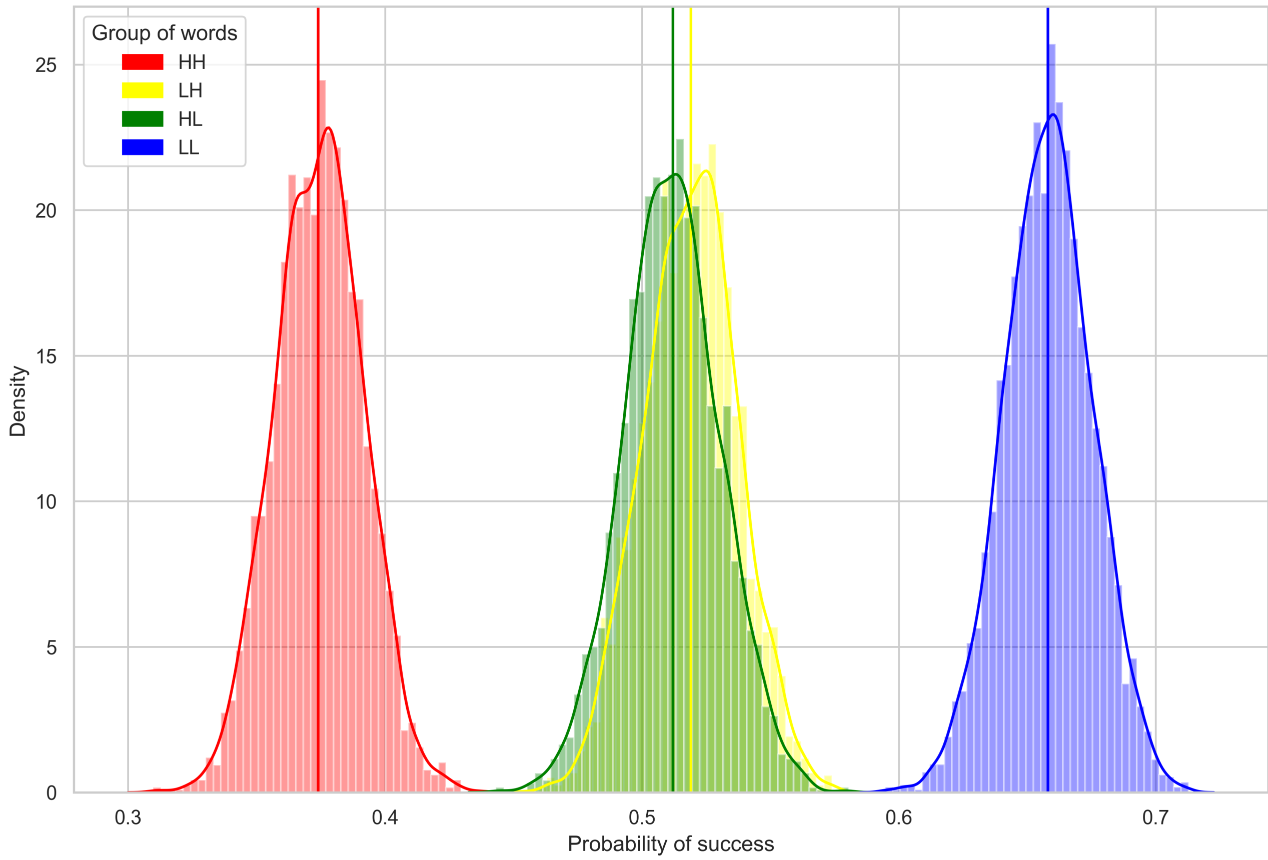

Figure 56. Posterior distributions of the English experiment results

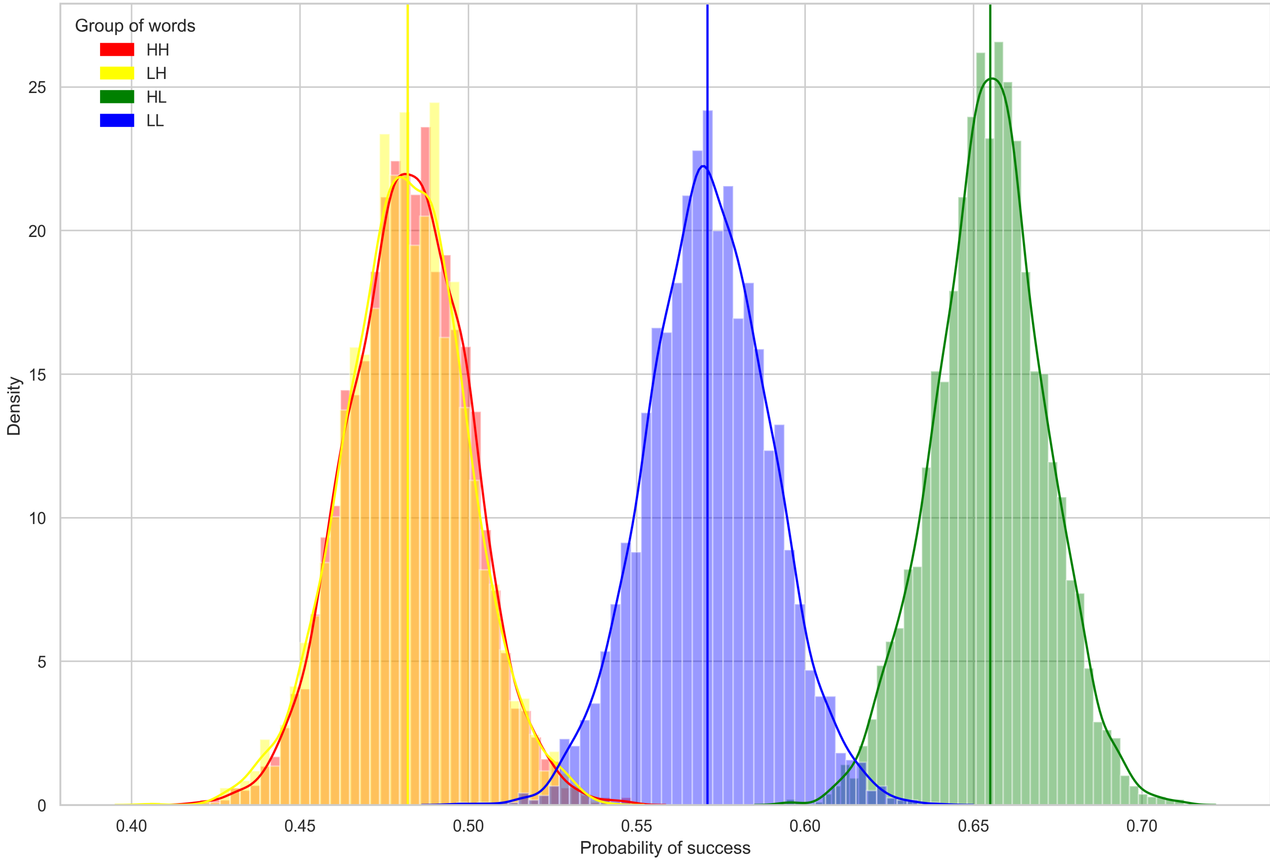

Figure 57. Posterior distributions of the Russian experiment results

Table 40. Means and highest density intervals of θs (English and Russian)

Construction type	Probability of success		
	HDI 3 %	θ	HDI 97 %
English			
HH	0.34	0.37	0.40
LH	0.48	0.51	0.55
HL	0.48	0.51	0.54
LL	0.62	0.65	0.69
Russian			
HH	0.44	0.48	0.51
LH	0.44	0.48	0.51
HL	0.62	0.65	0.68
LL	0.53	0.57	0.60

Table 41. Success ratios in the individual construction pairings (English)

	HH	PA	HL	LH	LL
LH	0.61	0.58	0.50	—	0.37
HH	—	0.43	0.36	0.38	0.31
LL	0.69	0.67	0.65	0.62	—
HL	0.63	0.56	—	0.5	0.35
PA	0.56	—	0.43	0.42	0.33

Table 42. Success ratios in the individual construction pairings (Russian)

	PA	LH	HH	LL	HL
LH	0.65	—	0.47	0.39	0.41
HH	0.65	0.53	—	0.43	0.32
LL	0.70	0.60	0.57	—	0.41
HL	0.77	0.59	0.68	0.59	—
PA	—	0.35	0.35	0.30	0.23

7.2.4 Discussion

Some important things here merit discussion. Let us start with the English part of the experiment. First, the results show that complex words of the LH and HL types are indeed perceived by native speakers as different from words of the HH and LL types with regard to their morphological analysability and semantic transparency. The ranking of the obtained probabilities of success is in agreement with my initial hypothesis that the degree of the construction's perceived complexity would be proportional to the number of empty slots within it. Notably, the distributions of θs for English LH and HL words are lumped together (Figure 56), suggesting that under these experimental conditions, participants exhibited no clear preference in choosing between constructions with the slot for an affix and constructions with the slot for a base.

The results are stable across individual construction pairings. As can be seen in Table 41, (1) no construction type has a greater proportion of successes than LL, (2) the LH and HL contest ended in a tie, and (3) HH lost to every other construction, including, somewhat surprisingly, even pseudo-affixed words. This alignment of the construction types HH < HL / LH < LL is hard to reconcile with the relative frequency account. As I have already noted, if one takes into account only the derivation to base family frequency ratio, then one would expect to find LH and HL words much closer to LL and HH words, respectively, than to each other.

The Russian part of the experiment produced a different hierarchy of construction types that is even more incompatible with the relative frequency view: LH / HH < LL < HL. Here, parsable constructions with an empty slot for a base were consistently rated as more complex than their compositional counterparts. On the other hand, parsable constructions with an empty slot for an affix were merged with non-analysable items.

The striking difference between English and Russian results begs the question of how it can be explained. One possible way to account for this difference is to think back to the proposed model of the parsable constructions' meaning processing:

$Meaning_{ITEM} = Meaning_{COMPONENT\ 1} + X$,

where X denotes a semantic element that is not readily available and has to be obtained by solving the equation

$X = Meaning_{ITEM} - Meaning_{COMPONENT\ 1}$.

In LH words, X is the meaning of an affix, so the participants in the experiment had to assess whether a certain affix brought anything significant to the composite conceptualisation of the derived form once they had accounted for the contribution of the base. By contrast, in HL words, X is the meaning of a base, and the participants had to evaluate its contribution while holding the meaning of the affix fixed.

What distinguishes English LH—HL types opposition from the corresponding Russian one is that these constructions came to be semantically specialised in Russian. Prefixes in the Russian verbs of the LH type mostly encode spatial meanings inherited from prepositions, while the same prefixes in Russian HL verbs tend to have non-spatial, idiosyncratic, construction-specific meanings. From this, it necessarily follows that the fixed elements of the Russian LH constructions (bases) depart from their free counterparts in semantics and distribution to a much lesser extent than the fixed elements of the HL constructions (prefixes) (cf. Kiparsky, 1997; McIntyre, 2015; Monakhov, 2023a). In English, a similar distinction is attested to verb-particle constructions but not to prefixed verbs (Stiebels, 1996; McIntyre, 2007).

This, in fact, can be verified formally using the distributional hypothesis, which states that similarity in meaning results in similarity in linguistic distribution (Firth, 1957). Words that are semantically related tend to be used in similar contexts. Hence, by reverse-engineering the process — that is, coding words' discourse co-occurrence patterns with multi-dimensional vectors and performing certain algebraic operations on them — distributional semantics can induce semantic representations from contexts of use (Boleda, 2020). It is well-established that the similarity of words' vector representations goes beyond simple syntactic regularities (Rehurek and Sojka, 2011; Mikolov et al., 2013; Pennington, Socher, and Manning, 2014) and that vector space models perform well on tasks that involve measuring the similarity of meaning between words, phrases, and documents (Turney and Pantel, 2010).

More importantly for the purposes of this study, vector space models have been used to assess the degrees of compositionality of complex linguistic expressions, notably nominal compounds (Cordeiro et al., 2019) and particle verbs in English (Bannard, 2005) and German (Bott and Schulte im Walde, 2014). The general premise of such analyses is that if the meaning of a multi-word expression is the sum of the meanings of its parts, then a distributional semantic model will reveal significant similarity between the vector for a compositional expression and the combination of the vectors for its parts, computed using some vector operation. Conversely, the lack of such similarity might be interpreted as a manifestation of the complex expression's idiomaticity.

Applying the aforementioned principle to multi-morphemic words seems a straightforward extension. Given the suggested model of the parsable constructions' meaning processing, it is possible to test it by performing simple algebraic operations on semantic vectors representing the experimental stimuli and their subparts. Specifically, the idea is that if one measures the cosine distance between the vectors of the derived form and its fixed element, then in Russian, this distance will be much smaller for LH words than for HL words. In English, on the other hand, there will be no difference between these two types of parsable constructions.

Vector space models have known limitations. Specifically, traditional word2vec (Mikolov et al., 2013) and GloVe (Pennington, Socher, and Manning, 2014) models tend to perform worse when confronted with word formation of morphologically rich languages like German (Köper, Scheible, and Schulte im Walde, 2015) and Russian (Drozd, Gladkova, and Matsuoka, 2016). Of all the existing non-contextualised pre-trained vector models of English and Russian languages, the FastText models seemed best suited for the purposes of this study. While other popular models ignore the morphology of words by learning their vectors, in the FastText model, a vector representation is associated with each character *n*-gram, and words are represented as the sums of these *n*-gram vectors (Bojanowski et al., 2017).

I used identical Continuous-Bag-of-Words (CBOW) FastText models for English and Russian that contained word vectors trained on Common Crawl and Wikipedia, in dimension 300, with position weights, character *n*-grams of length five, a window of size five, and 10 negatives. For each LH stimulus in my data, the cosine distance between its own vector and the vector of its base was recorded. For each HL stimulus, I obtained the cosine distance between its own vector and the vector of the corresponding prefix.

Comparison of the cosine distances' average levels for all Russian words of the LH and HL types indicates that the fixed elements of the former have indeed departed from their free counterparts in semantics and distribution to a much lesser extent than the fixed elements of the latter (M_{LH} = 0.49, M_{HL} = 0.93, t = -17.67, p < 0.001). With English data, as expected, no statistically significant difference between the LH and HL construction types in this regard was observed (M_{LH} = 0.78, M_{HL} = 0.86, t = -1.22, p = 0.23). The densities of the cosine distances for both languages and both construction types can be found in Figure 58.

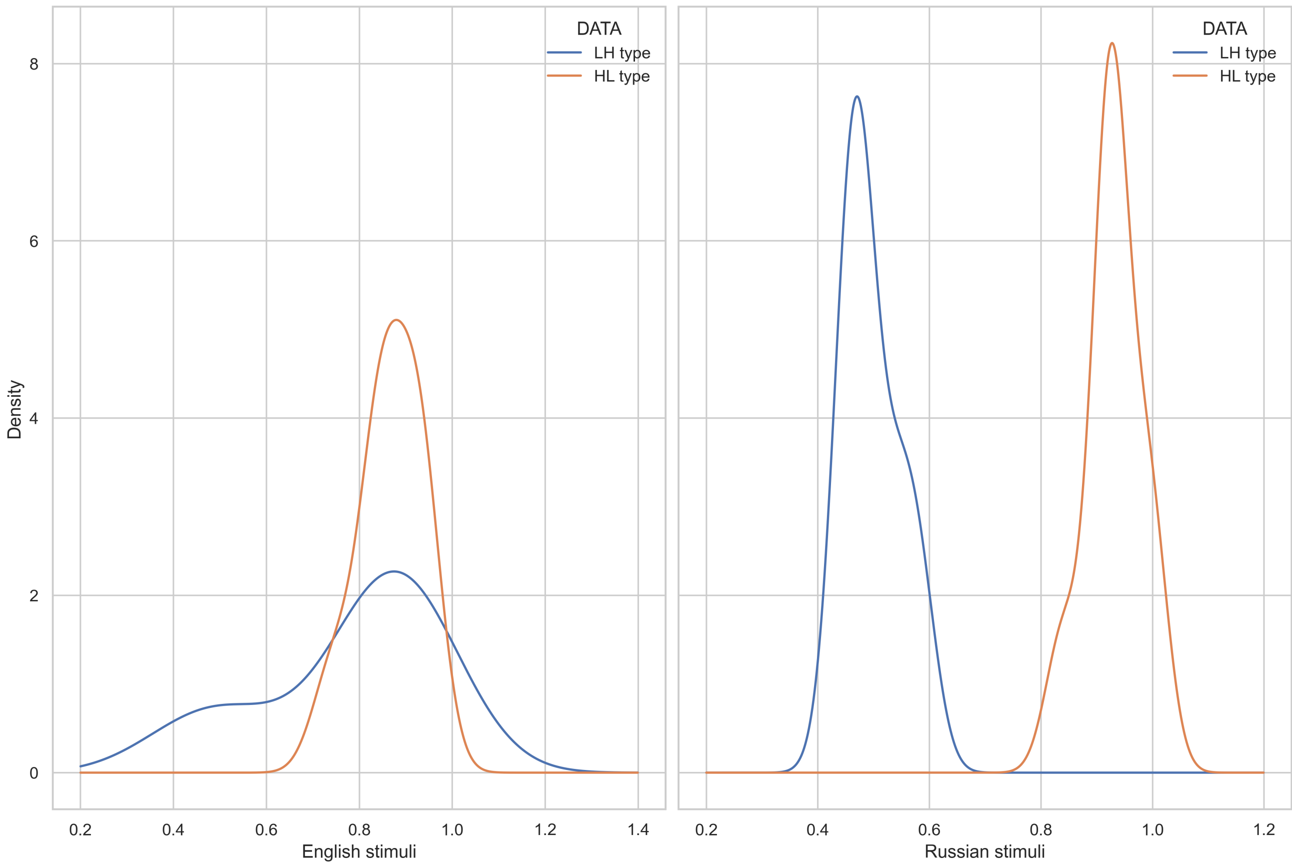

Figure 58. Densities of the cosine distances for English (left panel) and Russian (right panel) stimuli, LH and HL construction types

It makes a lot of intuitive sense that the closer the meaning of a complex linguistic item is to the meaning of one of its components, the harder it will be for the speakers to semanticise the remaining element, which is a prerequisite for judging the item as complex. The results of my Russian and English experiments confirm this view. Russian speakers, for example, should have considered the LH word *na-zhatj* 'press on' as less complex than the HL word *na-vreditj* 'do a lot of harm' because in the former case, the general meaning of the derivation is very much explained away by the meaning of its nested base *zhatj* 'press'. In the latter case, however, the contribution of the fixed element *na-* 'accumulate or produce in great amounts' to the meaning of its host is only of a framework nature.

It is very unlikely that the English participants were confronted with the same complications. For example, *able* in *enable* (LH) does not tell us the whole story of this word, nor does *en-* in *engrave* (HL). Similar difficulties would probably arise for English speakers were they to evaluate the complexity of spatial and non-spatial verb-particle constructions with the same particle. Expressions like *come in* would likely be judged as less complex than *give in*, despite the apparent semantic transparency of the former and the non-transparency of the latter (cf. McCarthy, Keller, and Carroll, 2003).

7.3 Study 2: Disentangling parsability and compositionality

7.3.1 A probabilistic model of complex words' perceived complexity and a partial replication of Hay's experiment (2001)

One useful way to investigate the relationship between two measures of complex words' analysability is by thinking back to Hay's original paper (2001), in which the idea of a derivation to base frequency ratio was initially proposed, and analysing the data therein. It is instructive to look at the table in Appendix 5 with the prefix stimuli from Hay's article (2001). Here, the words in group A are more frequent than the bases they contain, and the words in group B are less frequent than the bases they contain. Thus, the hypothesis that Hay tested was that A words would be rated as less complex than B words. This suggestion was borne out as Hay observed that 'among prefixed pairs, 65% of responses favoured the form for which the base was more frequent than the whole. Only 35% of responses judged forms that were more frequent than their bases to be more complex than their matched counterpart' (Hay, 2001: 1049).

An interesting thing emerges if one looks at the transitional probabilities' log ratios calculated for Hay's experimental stimuli in the same way as I did before and at the construction types assigned to the words depending on these ratios. It turns out that most of the words in group A are of the HL rather than HH type and group B comprises mostly complex words of the LH and LL types rather than just the LL. Given that some non-analysable words are found in group B as well and that, as my experiment has shown, there seems to be no significant difference in the perceived complexity of English HL and LH constructions, one might wonder whether the two groups can indeed be reliably delineated with regard to their members' morphological complexity.

In order to test this, I built a probabilistic model that would, drawing on the evidence obtained during my English experiment, predict a most likely winner in the complexity assessment contest for each of the 17 pairs of words in Hay's data. The model, a fragment of which is visualised in Figure 59, is a Bayesian network with three types of nodes: (1) 11 prefix nodes, that is, nodes that encode how different prefixes encountered either in my or Hay's experiment (*con*-, *de*-, *dis*-, *en*-, *il*-, *im*-, *in*-, *out*-, *pre*-, *re*-, *un*-) affect complexity judgements; (2) five construction type nodes, that is, nodes that encode how four construction types and one pseudo-affixed type (*LH*, *HH*, *HL*, *LL*, *PA*) affect complexity judgements; and (3) 110 contest nodes, that is, nodes that encode the likelihood of a word of a certain construction type being judged as more complex when paired with a word of a different type but the same prefix (*un_HH_PA*, *dis_LH_HH*, etc.).

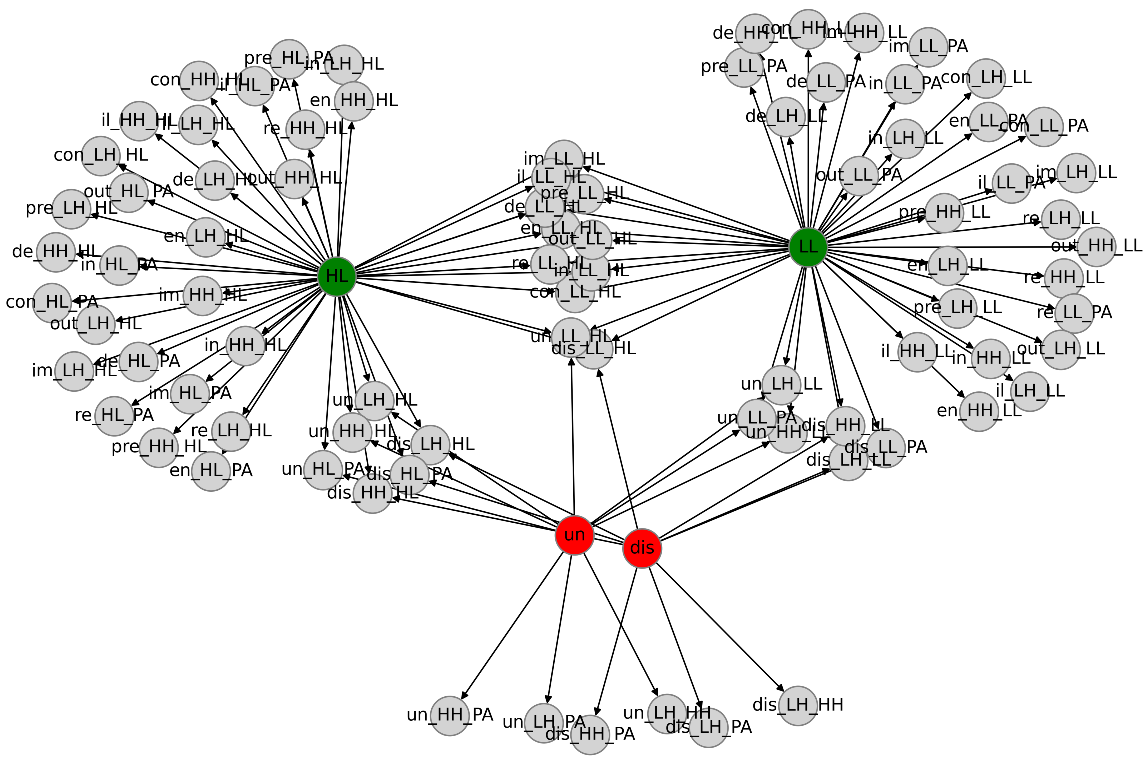

Figure 59. Fragment of the Bayesian network for words' complexity contests

Despite its fairly complicated global structure, on the local level, this Bayesian network is a very simple, state-observational model (Koller and Friedman, 2009) that reproduces the same conditional probability distribution for each contest node given the values of its three parents: one prefix node and two construction type nodes. The prior probabilities in the marginal and joint distributions were specified in an uninformed, commonsensical way: (1) for both prefix and construction type nodes, the probability that they facilitate analysability was set as equal to the probability that they do not, and (2) for contest nodes, the probabilities in the joint distribution simply reflected the (obvious) fact that if a word belongs to a construction type with a greater positive bearing on complexity judgements than the type of its adversary, then the former word is more likely to win the contest. Prefixes, however, might be expected to interact with construction-type pairings in an idiosyncratic manner, making differences between them either more pronounced or more attenuated.

The inference process was two-fold, based on both evidential and causal reasoning (Pearl, Glymour, and Jewell, 2016). First, I used the evidence for contest nodes obtained during my English experiment to infer posterior probability distributions for the prefix and construction type nodes. These distributions, unlike my non-informative priors, were conditioned on observed evidence and hence calibrated to be those under which the results of the experiment were most likely to occur. As a second step, I reverse-engineered the process and, using the updated prefix and construction type nodes' probability distributions, inferred the most likely assignment of values for the contest nodes that corresponded to the 17 prefixed words' pairings in Hay's paper (2001). Finally, after the model had been trained, I used the learnt probabilities to predict the outcome of a hypothetical experiment where 24 participants would be asked to select a more complex word in each of the 17 pairs under investigation.

In order to check the adequacy of the model's predictions, I tested the same 17 pairs of words in an experimental setting identical to the one of my above-described English study. A total of 408 participants (24 x 17), none of which had taken part in the previous experiment, were assembled via the Amazon Mechanical Turk crowdsourcing platform so as to conform to the matrix in Table 39. Each subject was presented with just one pair of words and asked to decide which member of the pair was more complex.

The correlation between the predicted and observed proportions of success was found to be significant ($r = 0.52$, $p = 0.03$). Most importantly, in both hypothetical and real experiments, 55% of responses judged the words from group A to be more complex than their counterparts, and only 45% of responses chose the words from group B. The difference in the number of votes given to each group in the experimental setting was significant ($M_{\text{Group A}} = 13$, $M_{\text{Group B}} = 11$, $t = 2.28$, $p = 0.02$) and accurately predicted by the model ($M_{\text{Group A}} = 13$, $M_{\text{Group B}} = 11$, $t = 2.72$, $p = 0.01$).

Thus, my results are the opposite of what Hay reported: words in group A, though more frequent than the bases they contain, were rated more complex than words in group B, which are less frequent than the bases they contain. It is important to note that the two experiments are not directly comparable. Though the general design and instructions were the same, the number of participants was similar, and the prefixed stimuli were identical, there were two important divergencies. First, my experiment was completed online, and each participant worked with just one pair of words without seeing the whole list of stimuli. Second, I did not test suffixed words and used no filler word pairs.

These divergencies were likely to have one important consequence: my subjects were not primed to perform the same operation of segmenting out two base candidates and comparing them as free elements on each pair of words. I argue that people who have been previously asked to select a more complex word in a filler pair like *family–busily*, when confronted with the pair *uncanny–uncommon*, will be prone to mark *uncanny* (HL, group A) as less complex than *uncommon* (LH, group B). They will do so simply because they have been trained to directly compare *canny* with *common* without taking into account their interaction with the general meaning of the complex word. However, in a non-primed scenario, the reasoning patterns might be more complicated. I will elaborate on this in the following section.

7.3.2 Compositional and parsable models of meaning processing

As discussed above, in analysable parsable words of two types, two different elements occupy slot positions and are likely to be parsed out during the semantic analysis. With LH words, the participants of the experiment had to assess whether a certain affix brings anything significant to the composite conceptualisation of the derived form once they have accounted for the contribution of the base. In contrast, with HL words, the participants had to evaluate the contribution of the base while holding the meaning of the affix fixed.

If we again use the machinery of semantic vector space modelling to reify the alleged difference in the processing of the words *uncanny* and *uncommon* from the previous example, the following set of operations will be needed: (1) subtract the vector of the filler element (*canny* in *uncanny* and *un* in *uncommon*) from the vector of the complex word $\overrightarrow{W}$, (2) check whether the subtrahend vector $\vec{S}$ encodes something meaningful and relatable to the meaning of the derivation, and (3) check whether the difference vector $\overrightarrow{D}$ encodes something meaningful and relatable to the meaning of the derivation. Operation (1) is straightforward. Operations (2) and (3) may be performed by finding the nearest neighbours of the vectors $\vec{S}$ and $\overrightarrow{D}$ and calculating the average cosine similarity of these neighbours' embeddings to the vector representation of the complex word $\overrightarrow{W}$.

This may not be an inaccurate modelling of human reasoning. It seems plausible that the participants of the experiment, in order to come to a decision, first manipulated the filler element of a particular word and tried to assign some meaning to it. Next, they might have wanted to evaluate the general constructional meaning encoded by the fixed element with an empty slot free of any concrete lexical material. In this scenario, the search for nearest neighbours is a reasonable approximation of how people tend to semanticise language units using lexical paraphrases and synonyms (Wiegand, 1992; Mel'čuk and Polguère, 2018).

I tested this approach on the *uncanny–uncommon* pair and found that cosine similarities of the 20 nearest neighbours obtained for *uncanny* in the manner described above (10 for $\vec{S}$ and 10 for $\overrightarrow{D}$) were significantly higher than those obtained for *uncommon*: $M_{uncanny} = 0.76$ and $M_{uncommon} = 0.13$. Among the nearest neighbours of *uncanny*'s $\vec{S}$ and $\overrightarrow{D}$ vectors are *canny*, *shrewd*, *skilful*, *astute*, *deft*, *supernatural*, *eerie*, and so on. By contrast, the nearest neighbours of *uncommon*'s $\vec{S}$ and $\overrightarrow{D}$ vectors are mostly irrelevant and unpredictable items such as *stoicism*, *polyandry*, *expressivity*, *cross-linguistically*, and more.

In order to extrapolate the proposed logic of testing to the whole prefixed dataset in Hay's study (2001), it is necessary to account for two other construction types. HH words that I believe to be non-analysable pose no challenge in this regard since they are more likely to be accessed directly without segmentation. With LL words, however, the meaning processing model is supposed to be different — not parsable as in LH and HL types, but compositional. The rationale behind this model implies arriving at a composite conceptualisation by means of combining the meanings of two distinct elements, so the set of vector operations should be different here: (1) obtain the vectors $\overrightarrow{E}_1$ and $\overrightarrow{E}_2$ of the first and second elements of the complex word $\overrightarrow{W}$, (2) check whether the addend vector $\overrightarrow{E}_1$ encodes something meaningful and relatable to the meaning of the derivation, and (3) check whether the addend vector $\overrightarrow{E}_2$ encodes something meaningful and relatable to the meaning of the derivation. Again, operations (2) and (3) may be performed by finding the nearest neighbours of the vectors $\overrightarrow{E}_1$ and $\overrightarrow{E}_2$ and calculating the average cosine similarity of these neighbours' embeddings to the vector representation of the complex word $\overrightarrow{W}$.

I used the same FastText English model (Bojanowski et al., 2017) as before. For each of the 34 prefixed words in Hay's dataset (2001), I obtained its 20 nearest neighbours, applying in each case those vector modification operations which I outlined above for the respective construction types. Cosine similarity between each of the nearest neighbours and the target complex word was recorded. The average of these 20 similarities constituted the final measure of the word's perceived complexity with a clear interpretation: the larger the value, the more likely the word to be judged complex.

Analysis of the results shows that if this measure were the only driver of choice, then out of 17 contests, words in group A would win 70% of the time — that is, much more often than what actually occurred during my experiment. The difference in the average cosine similarities between groups A and B was found to be large, though slightly above the conventional significance level ($M_{Group\ A} = 0.72$, $M_{Group\ B} = 0.60$, $t = 1.87$, $p = 0.07$). Notably, the differences in average cosine similarities between the paired words of the two groups were strongly positively correlated ($\rho = 0.88$, $p < 0.001$) with the differences in the number of votes cast for respective contestants, as predicted by my Bayesian network.

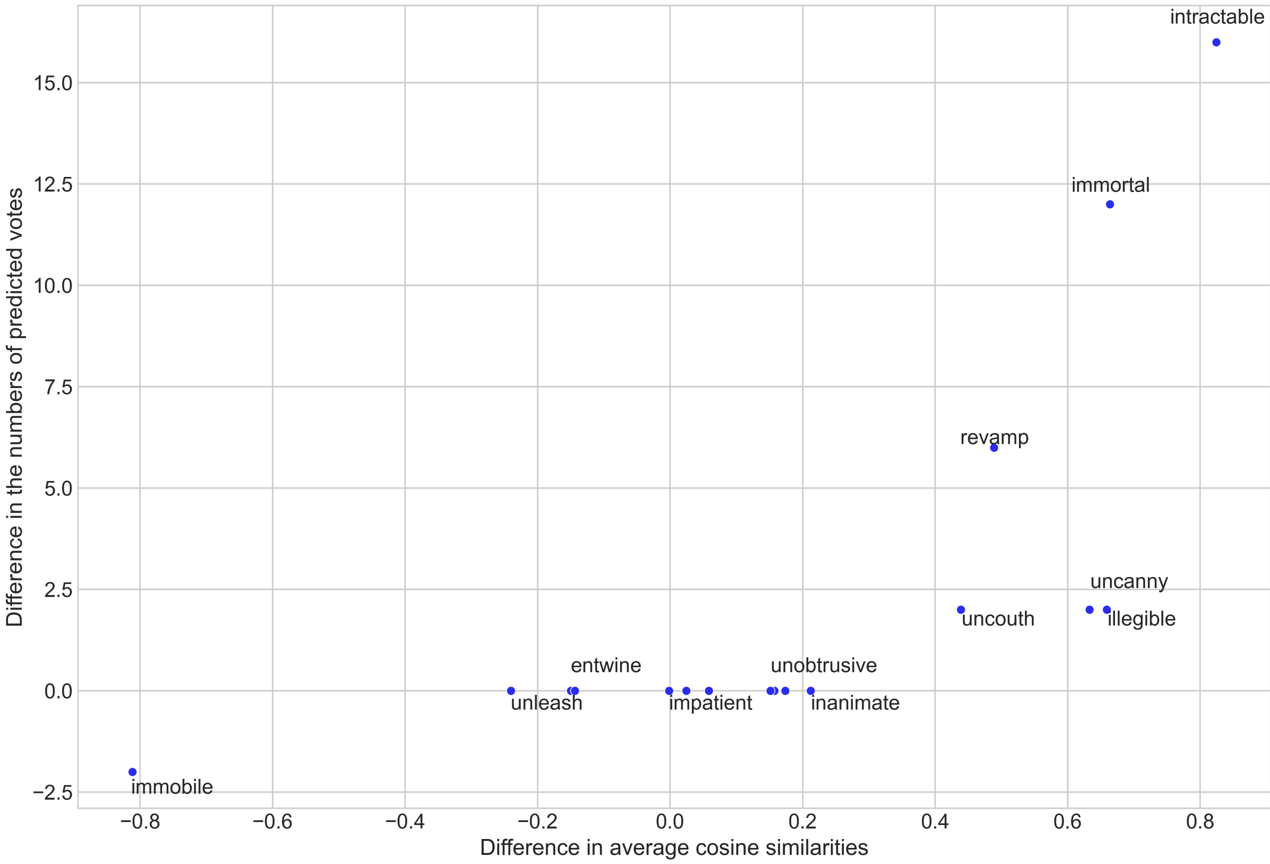

Figure 60. Differences (group A - group B) in cosine similarities and predicted votes

In Figure 60, the dots represent the complex words from group A (they are labelled selectively to avoid overplotting), the *x*-coordinates correspond to the differences between these words' average cosine similarities and those of their counterparts from group B, and the *y*-coordinates correspond to the differences between the number of people (out of 24 hypothetical participants) who, according to my model, would choose these words as more complex and the number of people who would prefer their counterparts from group B.

These results suggest, first, that the cosine similarity-based measure might be a reliable predictor of English complex words' degrees of complexity and, second, that this particular dataset serves as a useful illustration of how relying solely on relative frequency calculations can lead to the conflation of different construction types and thus obfuscate the difference between two meaning processing models, one based on the principle of compositionality and the other on the principle of parsability.

Table 43. Statistics of HH words in group B and their HL counterparts in group A

complex word	construction type	votes (predicted)	votes (observed)	cosine similarity	derivation/ base ratio
group A					
intractable	HL	20	18	0.82	0.641
immortal	HL	18	15	0.66	0.156
revamp	HL	15	12	0.48	0.318
group B					
impractical	HH	4	6	0	0.027
immoral	HH	6	9	0	0.043
retool	HH	9	12	0	0.002

The choice and pairing of 34 prefixed stimuli in this dataset were such that, given the probabilities of success presented in Table 40, words in group B were more likely to win in only seven contests out of 17. As for the other 10, the chances were either approximately equal or in favour of words from group A. Especially telling is the comparison of three HH words that made their way in group B due to low derivation to base frequency ratios with their matched counterparts from group A, which belong to the HL type (Table 43). The Bayesian network predicted that all of these HL words would be considered more complex, which was, to a large extent, borne out in the experiment with human participants. There is an almost perfect positive correlation of values in *votes (predicted)*, *votes (observed)*, and *cosine similarity* columns as well as a less than ideal and unexpectedly positive correlation of the same values with those in the *derivation/base ratio* column.

The presentation of these examples is intended to show that HL words, which constitute the majority of Hay's group A, are in fact not as simple as they might look, though they are quite distinct from LL words. The main point here is that there are two different types of complexity corresponding to the two meaning-processing models — parsable and compositional. The distinction between them, as has been shown, is imprinted in the semantic vector space and can be explained as follows.

LL words like *inadequate* or *reorganise* strongly overlap in semantics and distribution with their bases. One can easily replace *inadequate* with *not adequate* and *reorganise* with *organise again*. However, *uncouth* is not so easily replaceable with *not couth*, or *reiterate* with *iterate again*. Here, as the nearest neighbours of these words' modified vectors suggest, some general sense is encoded by the construction as such while the meaning of the base is only used for concretisation. Consider, for instance, the variability of specific lexical meanings in the set of words aligned with *uncouth*'s vector $\overrightarrow{D}$: *uncivilized*, *insolent*, *amoral*, *unprofessional*, *disrespectful*, *irresponsible*, and *obnoxious*.

To put it simply, *inadequate* is conceptualised as [– ADEQUATE] while *uncouth* is conceptualised as [(DEPRIVED OF DESIRABLE QUALITY) & (THIS QUALITY BEING COUTH)]. The distinction between two models of meaning processing does not necessarily pertain to different relative frequencies of derivations and bases. For example, the derivation to base family frequency ratios of the HL words in Hay's dataset range from 0.01 (*immodest*) to 4.43 (*uncanny*) and the derivation to base family frequency ratios of the LL words range from 0.003 (*immoderate*) to 0.95 (*illegible*). There is a lot of variability in these values, which makes it difficult to separate two construction types using only relative frequency criterion. On the other hand, taking into account the discrepancy in probabilities of transition from base to affix and from affix to base, one can correctly predict, in most cases, the model at work.

To provide an example of a different type of affix from a different language, consider two German nouns of the same stem: *Büchlein* 'small book' and *Bücherei* 'library'. Their respective frequencies in German Web 2020 corpus (*detenten20_rft3*, Sketch Engine) are 47,121 and 67,174 tokens, which, given the frequency of *Buch* 'book' (5,536,382 tokens), forces one to conclude that both words should be equally analysable (and probably indistinguishable under relative frequency account). However, whereas the meaning of *Büchlein* is undoubtedly compositional [BOOK +

SMALL], the meaning of *Bücherei* can hardly be expressed as *[BOOK + (PLACE WHERE IT IS KEPT)]. Rather, it is conceptualised as [(PLACE WHERE PEOPLE DEAL WITH CERTAIN OBJECTS PROFESSIONALLY) & (THIS OBJECT BEING BOOK)] (cf. *Käserei* 'cheese factory', *Mosterei* 'firm producing must, cider, or perry', etc.).

It is possible to predict which model — compositional or parsable — is more likely to be chosen in each case by comparing two words' transitional probabilities. For *Büchlein*,

$P\ (\text{affix} \mid \text{base})_{\textit{Büchlein}} = 47121 / 539800 = 0.08$

$P\ (\text{base} \mid \text{affix})_{\textit{Büchlein}} = 47121 / 721543 = 0.06$

$\varepsilon_{\textit{Büchlein}} = \log(0.08 / 0.06) = 0.28,$

and for *Bücherei*,

$P\ (\text{affix} \mid \text{base})_{\textit{Bücherei}} = 67174 / 539800 = 0.12$

$P\ (\text{base} \mid \text{affix})_{\textit{Bücherei}} = 67174 / 3329967 = 0.02$

$\varepsilon_{\textit{Bücherei}} = \log(0.12 / 0.02) = 1.79.$

Given that $0 < \varepsilon_{\textit{Büchlein}} < 1 < \varepsilon_{\textit{Bücherei}}$, one concludes that *Büchlein* is a compositional expression of the LL type, a free combination of morphemes that do not constitute a conceptual unity. *Bücherei*, on the other hand, belongs to the HL type and has high chances of being treated as a parsable, collocation-like item (Sinclair, 1991; Mason, 2000; Lindquist, 2009) in which an element that tells us less about its counterpart (suffix *-erei* in this example) activates general constructional meaning and an element that has a greater predictive power (stem *Buch*) serves as a filler for the construction's empty slot.

7.4 Study 3: Bringing productivity and parsability together

7.4.1 The contributions of parsable words to their affixes' productivity

Distinguishing between two types of analysable complex words — compositional and parsable — plays an important role in how we understand and describe the mechanics of morphological productivity. One influential theory claiming that the relationship between affixes' productivity and analysability is that of strong positive correlation was first formulated by Hay and Baayen (2002). As a way to evaluate affixes' productivity, they used Baayen's hapax-based measure (Baayen, 1991, 1992, 1994, 2009; Baayen and Lieber, 1991; Baayen and Renouf, 1996; Plag, 2021), and to evaluate affixes' analysability, they proposed the notion of parsing ratio. For each affix, its parsing ratio gives us the probability that a certain word with this affix will be decomposed by a language user during access (Hay and Baayen, 2003). Mathematically, parsing ratios are defined as the proportions of forms (types or tokens) which fall above the so-called parsing line given by the following equation: log(base frequency) = 3.76 + .76 * log(derivation frequency) (Hay and Baayen, 2002).

Using this set of measures, Hay and Baayen found 1) a significant inverse relationship between token frequency and the proportion of tokens which are parsed and 2) a significant positive relationship between the proportion of tokens which are parsed and Baayen's productivity measure. Based on these results, authors claimed that 1) 'the more often you encounter an affix, (…) the less productive that affix is likely to be' and that 2) 'the more often we encounter an affix (…), the less likely we are to parse words containing it' (Hay and Baayen, 2002: 219). Thus, their main result was linking analysability and productivity together.

The notion of parsing ratio builds upon the logic of relative frequency account of analysability. While this approach seems perfectly justified for words of the HH type (which are non-analysable and thus cannot bring anything to the productivity of their affixes) and words of the LL type (which are compositional and hence bear witness to their affixes' wide applicability), the picture is not so clear with parsable words of the LH and HL types. As already discussed, the derivation to base frequency ratio, whether calculated for stand-alone bases or for morphological families, can unpredictably conflate these words either with their non-analysable or compositional counterparts. For example, among my English experimental stimuli, both *decrease* (LH) and *deforest* (LL) made it above the parsing line while both *debunk* (HL) and *describe* (HH) fell below it.

Most importantly, it remains unclear what contribution LH and HL words really make to the overall morphological productivity of their affixes. On the one hand, the derivational elements in HL multi-morphemic words or multi-word expressions in German, Russian, and English, being fixed by construction, are often called 'semiproductive' in the literature (Jackendoff, 2002) in the sense that they have input limitations, that is, do not accommodate every base that is semantically compatible with the preverb, prefix, or particle (McIntyre, 2001; Blom, 2005). On the other hand, as observed in the Russian part of my experiment, derivational elements that fill in empty slots of LH words, though easily analysable, can be completely disregarded by language users if the meaning of the whole construction significantly overlaps with the meaning of its base. Taking all of this into account, I would expect a high proportion of parsable words among all derivations with a certain affix to be indicative of this affix's limited morphological productivity.

7.4.2 Data collection

In order to analyse the relation between the analysability of English and Russian prefixed LH and HL words and the productivity of their prefixes, I used the following two measures. The parsability ratio of a prefix was calculated as the proportion of words for which the absolute difference between P (affix | base) and P (base | affix) was greater than 1% (as a threshold value suggested by my experimental stimuli) among all words with this prefix. The English data, obtained from WordNet (Fellbaum, 1998), comprised a total of 25,816 words with the following 24 prefixes: *anti*-, *con*-, *counter*-, *cross*-, *de*-, *dis*-, *em*-, *en*-, *fore*-, *im*-, *in*-, *inter*-, *mid*-, *mis*-, *non*-, *out*-, *over*-, *pre*-, *re*-, *sub*-, *super*-, *trans*-, *un*-, and *under*-. (Some authors might not view elements like *over*- or *super*- as prefixes but as combining forms; however, I trod here a conventional path, relying on the authoritative opinion of Oxford English Dictionary.) The Russian data, obtained from Tikhonov's morphemic dictionary (Tikhonov, 1985), comprised 9,018 words with the following 27 prefixes: *de*-, *diz*-, *do*-, *iz*-, *na*-, *nad*-, *niz*-, *ob*-, *pere*-, *pre*-, *pro*-, *po*-, *pod*-, *pred*-, *pri*-, *raz*-, *re*-, *s*-, *so*-, *o*-, *ot*-, *u*-, *v*-, *voz*-, *vz*-, *vy*-, and *za*-. The numbers for calculating transitional probabilities were gathered from the same internet corpora of English and Russian that I used while collecting data for the experiments.

As for the linguistic productivity of a prefix, I did not want to use Baayen's hapax-based measure since, as it has been pointed out in the literature, this measure is ill-suited for the comparison of affixes with different token numbers (Gaeta and Ricca, 2006; see also Bauer, 2001; Pustylnikov and Schneider-Wiejowski, 2010). Calculating the ratio of the number of hapax legomena with a given affix to the total number of tokens with that affix is likely to result in overestimating the productivity values of less-frequent constructions, which is undesirable for the purposes of this study. Instead, I assessed the morphological productivity of the prefixes in my data as their probability to combine with a random base (see Chapter 5).

The productivity values for English and Russian prefixes are to be found in Table 44 alongside their parsability ratios.

Table 44. Parsability ratios and productivity values for English and Russian prefixes

	English			Russian	
prefix	parsability	productivity	prefix	parsability	productivity
anti-	0.11	0.54	*de-*	0.53	0.22
con-	0.42	0.26	*diz-*	1.0	0.01

counter-	0.04	0.43	*do-*	0.05	0.68
cross-	0.17	0.45	*iz-*	0.30	0.42
de-	0.40	0.42	*na-*	0.38	0.73
dis-	0.32	0.29	*nad-*	0.18	0.24
em-	0.32	0.15	*niz-*	1.0	0.08
en-	0.26	0.23	*o-*	0.59	0.81
fore-	0.09	0.29	*ob-*	0.30	0.57
im-	0.46	0.25	*ot-*	0.32	0.81
in-	0.40	0.85	*pere-*	0.21	0.63
inter-	0.15	0.44	*pre-*	0.15	0.57
mid-	0.16	0.50	*pro-*	0.42	0.74
mis-	0.09	0.30	*po-*	0.52	0.81
non-	0.09	0.80	*pod-*	0.25	0.60
out-	0.05	0.59	*pred-*	0.40	0.16
over-	0.03	0.63	*pri-*	0.36	0.62
pre-	0.06	0.83	*raz-*	0.26	0.52
re-	0.16	0.68	*re-*	0.26	0.42
sub-	0.11	0.63	*s-*	0.52	0.79
super-	0.04	0.61	*so-*	0.45	0.43
trans-	0.19	0.26	*u-*	0.50	0.63
un-	0.24	0.66	*v-*	0.28	0.67
under-	0.04	0.52	*voz-*	0.25	0.35
—	—	—	*vz-*	0.45	0.17
—	—	—	*vy-*	0.41	0.71
—	—	—	*za-*	0.61	0.84

7.4.3 Analysis of English results

The results presented in Table 44 for the English data are plotted in Figure 61. The observable distribution of dots here has a characteristic U-shape and is reasonably well modelled by a polynomial regression with two terms (Y = 1.42 * x - 1.69 * x + 0.62). This suggests that parsability ratio, calculated as I propose, is related to productivity in a very special way. To better understand what is going on, it is important to remember exactly what this ratio signifies: it describes how many lemmas with a certain prefix comprise elements of which one is more of less fixed and another is more or less free to vary. It logically follows that if a prefix is highly productive, the proportion of such lemmas in its output will be low, as the majority of lemmas will be constructions of the LL type with comparably low transitional probabilities.

This, however, explains only the downward trend in Figure 61. The U-shape pattern suggests that there must be at least one other variable, besides parsability ratio and productivity measure, that influences the distribution of dots in this plot. One cannot but notice that the curve is pulled upwards by the prefix *in*-, specifically by its prepositional variant (e.g., in expressions like *in-place running* or *in-text citation*). Hence, one can hypothesise that the third variable of interest is the frequency of respective prepositions or particles.

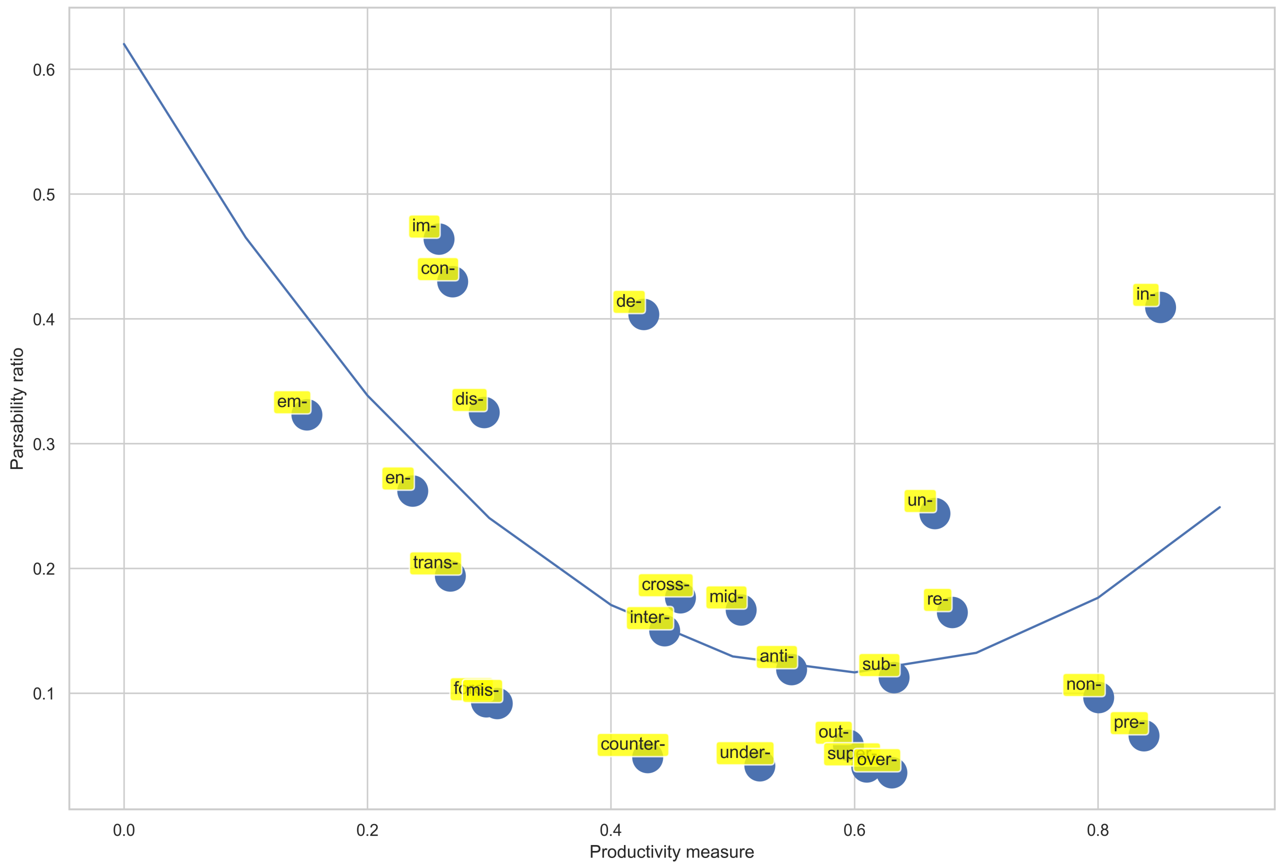

Figure 61. English prefixes' parsability and productivity

Indeed, an ordinary least squares regression model with a prefix's productivity as the response regressed on two interacting independent variables, namely the prefix's parsability ratio and the log-transformed frequency of the respective preposition or particle, accounts for a considerable amount of the total variation (Table 45). The obtained coefficients show that for the prefixes which have no free counterparts or correspond to relatively low-frequency prepositions/particles (*over*, log-transformed frequency of 16.99; *out*, 17.38), the lower the parsability ratio, the greater the linguistic productivity. In contrast, for the prefixes that have high-frequency free counterparts (*in*, 19.87), the higher the parsability ratio, the greater the linguistic productivity.

Table 45. Regression model summary (English prefixes)

term	CI 5%	coefficient	CI 95%	SE	p
intercept	-2.12	-1.60	-1.08	0.25	< 0.01
parsability	-0.68	-0.40	-0.12	0.13	< 0.01
PP frequency	0.01	0.08	0.14	0.02	0.01
interaction	0.005	0.03	0.05	0.01	0.02

Note: $F(3, 20) = 4.74$, $p = 0.01$, $R^2 = 0.41$.

7.4.4 Analysis of Russian results

The Russian language provides many more possibilities for this type of analysis. Of the 27 prefixes in my data, 17 are not only historically related to prepositions, but also have prepositional counterparts in modern Russian: *v-* (*v* 'in, at'), *do-* (*do* 'to, before'), *za-* (*za* 'for, behind'), *iz-* (*iz* 'from, out of'), *na-* (*na* 'on'), *nad-* (*nad* 'over, above'), *o-* (*o* 'about'), *ob-* (*ob* 'about'), *ot-* (*ot* 'from'), *po-* (*po* 'along, by'), *pod-* (*pod* 'under'), *pred-* (*pered* / *pred* 'before, in front of'), *pri-* (*pri* 'by, at'), *pro-* (*pro* 'about, of'), *s-* (*s* 'with'), *so-* (*so* 'with'), and *u-* (*u* 'from, by'). The second group of prefixes, which have no prepositional counterparts in modern Russian, encompasses morphemic borrowings, prefixes of non-prepositional origin and prefixes derived from prepositions that are no longer part of the Russian language.

The methodology of calculating parsability ratios and productivity measures for Russian prefixes was exactly the same as that for English. The results provided in Table 44 are visualised in Figure 62. The picture, overall, bears a remarkable resemblance to the U-shape distribution of English prefixes observed in Figure 61. Again, this distribution can be reasonably well approximated by a polynomial regression with two terms (Y = 2.99 * x - 3.03 * x + 1.001).

More telling is the distribution of prefixes if one takes into account their prepositional or unprepositional natures. For unprepositional prefixes only, a clear negative linear trend is observable: the lower the parsability ratio, the greater the linguistic productivity. This means that the bewildering U-shape pattern is created solely by prepositional prefixes.

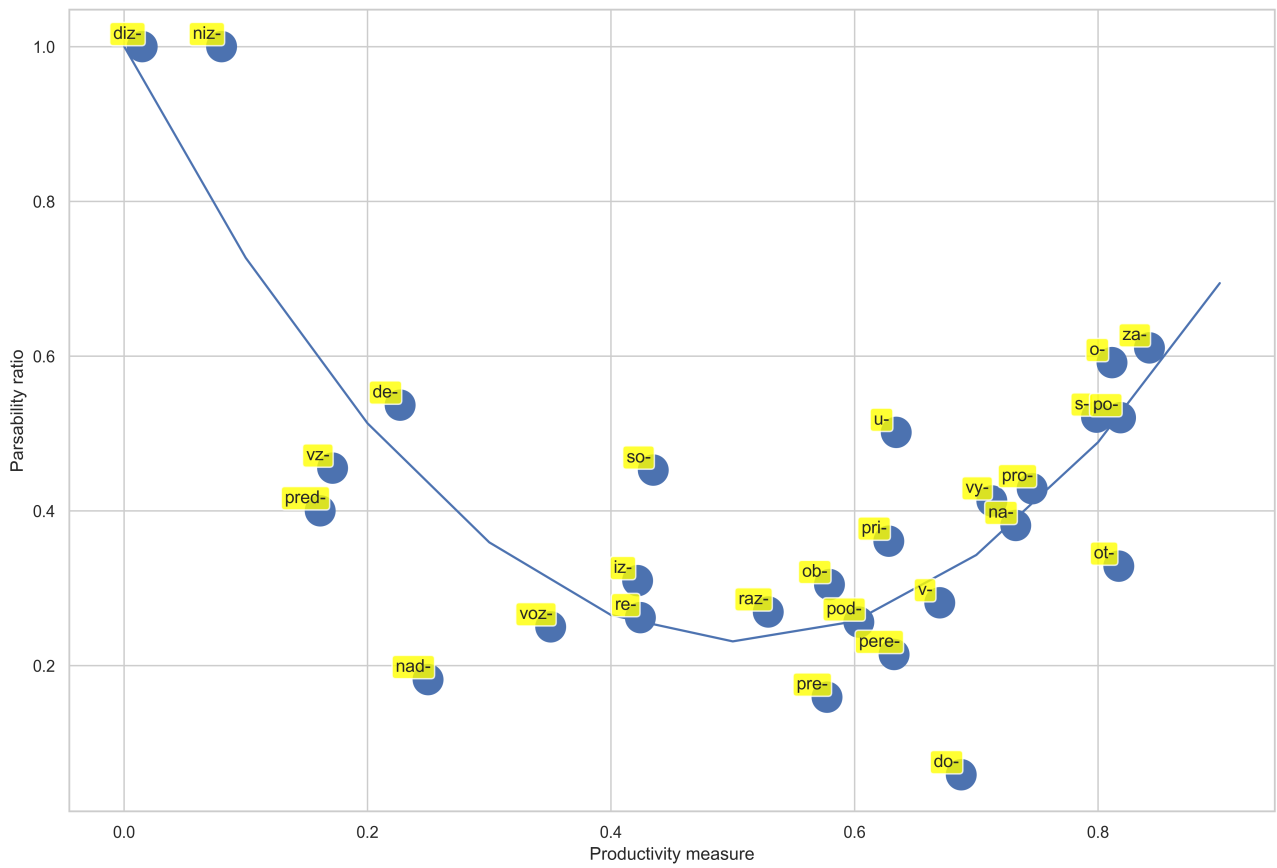

Figure 62. Russian prefixes' parsability and productivity

As before, a regression model with a prefix's productivity as the response regressed on two interacting independent variables — the prefix's parsability ratio and the log-transformed frequency of the respective preposition — explains a significant amount of the total variation (Table 46).

Table 46. Regression model summary (Russian prefixes)

term	CI 5%	coefficient	CI 95%	SE	p
intercept	-3.63	-2.95	-2.27	0.32	< 0.01
parsability	-2.18	-1.59	-0.99	0.28	< 0.01
PP frequency	0.09	0.14	0.19	0.02	< 0.01
interaction	0.05	0.09	0.14	0.02	< 0.01

Note: $F(3, 23) = 15.9$, $p < 0.001$, $R^2 = 0.67$.

Thus, Russian prefixes, corresponding to low-frequency prepositions, behave exactly like unprepositional prefixes: the lower the parsability ratio, the greater the linguistic productivity. On the other hand, those prefixes whose prepositional counterparts are very frequent remain highly productive even though the proportion of lemmas with unilaterally fixed elements in their overall output is significant.

To assess how cognitively plausible this is, let us consider two different prefixes. One is *pred*- 'before', which corresponds to the respective, relatively infrequent (15.7) preposition *pred*. The prefix *pred*- has a parsability ratio of 0.4, indicating that 40% of the words with this prefix are of the LH or HL types. The productivity measure of this prefix, on the other hand, is only 0.16, which means that out of 100 random bases it can be expected to combine only with 16. In comparison, the prefix *o*- 'about, around' has an even higher parsability ratio of 0.59, and yet its linguistic productivity is 0.81.

I would argue that the difference here is due the fact that the corresponding preposition *o* is 6.75 times more frequent than the preposition *pred*. Thus, even though one would expect the prefix *o*- to be unproductive given its high degree of parsability, the presence of a free element coinciding with it in form and partially overlapping in meaning may facilitate the production of new items.

7.5 Conclusion

Hay's work on lexical frequency in morphology was a huge step forward in understanding the mechanisms of morphological processing. The idea that it is relative rather than absolute frequency that affects the decomposability of complex words revealed that high-frequency forms are not necessarily holistic and low-frequency forms are not necessarily decomposable. The former might be accessed via the route of decomposition if the bases they contain are of even higher frequency, and the latter might be accessed as one chunk if they are built of lower-frequency parts.

However, the relative frequency account, while in most cases correctly distinguishing between non-analysable and compositional expressions (HH and LL, in my notation), was not able to register the presence of two other construction types that are comprised of a fixed element and a slot (LH and HL), and instead lumped them together with either LL or HH constructions. However, the findings presented in this study suggest that LH and HL complex word should be treated as schemas in their own rights. The identification of these expressions is important insofar as it allows a distinction to be drawn between two different meaning processing models.

A compositional model of the LL type implies that each of the elements entering into combination is equally free to vary; the combination itself is judged by language users to be semantically complex but transparent. A parsable model of the LH and HL types assigns some very general sense to the construction as such. Multi-morphemic words of these types are similar to collocations in the sense that they also consist of a node (conditionally independent element) and a collocate (conditionally dependent element). Such combinations of linguistic items are also considered semantically complex but less transparent because a collocate's meaning does not generally coincide with the meaning of a respective free element (even if it exists) and must be parsed out from what is available.

The difference between the compositional LL type, on the one hand, and parsable LH and HL types, on the other, has predictable implications for the affixes' morphological productivity. A high proportion of parsable words among all derivations with a certain prefix might be taken as a sign of the prefix's constrained productivity. It is clear that if, among multi-morphemic words with a certain prefix, there are many words whose bases are conditionally dependent upon the prefix — that is, there is a strong sequential link between the elements — the prefix's range of applicability is limited, and the constructional meaning is not general enough to accommodate a wide variety of items in its slot. This relationship may, however, be reversed: if for some prefix there exists in language a corresponding free element that is sufficiently frequent, it can lead to higher productivity even of those prefixes with high parsability ratios.

Clearly, the distinction between the two models of meaning processing is not a clear-cut categorical one but rather a probabilistic continuum. One can predict which model — compositional or parsable — is more likely to be chosen for each word by taking into account the word's two morphological families: one for the affix, another for the base. The words that are characterised by a greater discrepancy between transitional probabilities from affix to base and from base to affix are more likely to be treated as parsable than those with more or less comparable (low) transitional probabilities. Thus, for English prefix-base constructions with *re-*, some points on this cline, arranged in the order of the gap in transitional probabilities narrowing, would be *refurbish* (transitional probabilities' log ratio of 3.38) → *revamp* (3.33) → *rekindle* (2.85) → *reiterate* (2.13) → *reorganise* (0.67), so that *refurbish* is most likely to be parsable and *reorganise* compositional.

One remaining question is whether the current proposal would also be valid for suffixes. Though the scope of the article was limited only to prefixed words in English and Russian, the transitional probabilities' ratio approach does not seem inapplicable to suffixation. Still, I realise that the one German example that I provided is not enough to make any strong statements, and thus this issue requires further investigation.

I will conclude by discussing one terminological subtlety. The terms *analysability*, *decomposability*, and *parsability* are usually used interchangeably, all describing the process whereby the composite conceptualisation is broken down into component parts. However, I think it is more reasonable to differentiate between them in the following manner. *Analysability* is best used as an umbrella term that can be applied simultaneously to both meaning processing models. *Decomposability*, then, could be reserved for referring exclusively to the semantic processing operations induced by the compositional model and *parsability* to those induced by the parsable model.

8 How complex verbs acquire their idiosyncratic meanings

8.1 Introduction

While discussing multi-morphemic words and multi-word expressions, several important concepts should be taken into account: decomposability (analysability, parsability), compositionality of meaning (semantic transparency), and linguistic productivity. *Decomposability* means that such linguistic expressions can be divided by language users into constituent parts and then reassembled from these parts into a whole. Cf. '[analysability is the] recognition of the contribution that each component makes to the composite conceptualisation' (Langacker, 1987: 292).

The notion of decomposability is conditional upon the notion of *compositionality of meaning* because our ability to break a complex form into a number of simpler forms crucially depends on our ability to assign meanings to these forms (Taylor, 2012).

In a similar vein, the notion of decomposability seems to imply the notion of *linguistic productivity*. Since the parts of multi-morphemic words and multi-word expressions are accessible to us as form-meaning pairings, we can readily use them as building blocks to assemble new linguistic items (Laudanna, 1999).

However, if one considers a special case of multi-morphemic words / multi-word expressions — namely complex verbs (Booij and Kemenade, 2003) in German and Russian, as well as verb-particle constructions in English — some problems associated with this line of reasoning become evident. In both German and Russian, there are complex-verb patterns that are (1) decomposable, (2) compositional, and (3) productive. In a prototypical situation, prefixes here encode spatial meanings, inherited from prepositions:

German

(1)	*raus-gehen* out-go 'go out'	(2)	*zu-gehen* to-go 'approach somebody'	(3)	*durch-gehen* through-go 'walk through'

Russian

(4)	*na-pisatj*	(5)	*nad-pisatj*	(6)	*pod-pisatj*
	on-write		above-write		under-write
	'write down'		'write above'		'subscribe'

On the other hand, there are numerous examples of complex-verb patterns that, though totally decomposable, are not absolutely compositional in meaning (cf. Bybee, 2010: 45):

German

(7)	*auf-klären*	(8)	*auf-bessern*	(9)	*auf-schaukeln*
	on-clear		on-improve		on-sway
	'clear up'		'polish up'		'build up'

Russian

(10)	*na-govoritj*	(11)	*na-gotovitj*	(12)	*na-rozhatj*
	on-talk		on-prepare		on-give birth
	'say a lot'		'cook a lot'		'give birth to many'

With regard to morphological productivity, such linguistic expressions are sometimes called 'semiproductive' in the literature (Jackendoff, 2002), in the sense that they (1) have some input limitations — that is, they do not accommodate every base that is semantically compatible with the preverb/prefix/particle (McIntyre, 2001; Blom, 2005) — and (2) are believed to be listed — that is, they have to be memorised.

Semiproductive or not, these complex verbs often constitute very large groups of words (see the detailed discussions of specific preverb, prefix, and particle uses in German, Russian, and English in Kühnhold and Wellmann, 1973; Stiebels, 1996; Krongauz, 1998; Larsen, 2014), which, notably, are open to new members. One can compare the following examples with those listed under (7)–(9) for German and under (10)–(12) for Russian:

(13)	*aus-merkel-n*	(14)	*rum-merkel-n*	(15)	*ver-merkel-n*
	out-Merkel-INF		around-Merkel-INF		ver-Merkel-INF
	‘ignore a problem until it solves itself’		‘do nothing, make no decisions’		‘ruin, waste something completely’

(all examples reflect on the political stance associated with former German chancellor Angela Merkel)

Russian

(16)	*na-priviv-atj*	(17)	*na-zum-itj-sja*	(18)	*na-mitu-sh-nich-atj*
	on-vaccinate-INF		on-zoom-INF-REFL		on-MeToo-Ø-AGT-INF
	‘vaccinate many people’		‘take part in too many zoom meetings’		‘cause a lot of (needless) commotion while being an active part of the #MeToo movement’

Another observation about non-spatial complex verbs is that their bases are sometimes highly idiosyncratic, showing little or no semantic relation to the meaning of the complex verb (cf. English: *make off*, *pack off*, *piss off*, *bugger off*, *skive off*, *slope off*, *spirit off*, *bog off*, *push off*, *shove off*, etc. [examples are from McIntyre, 2002: 111]).

As for the meaning, some scholars argue that preverbs in verbs like German (7)–(9) and (13)–(15) or prefixes in verbs like Russian (10)–(12) and (16)–(18) are meaningless, conveying aspectual or telic interpretations (cf. Spencer and Zaretskaya, 1998). However, many complex verbs that appear to be semantically idiosyncratic when looked at in isolation reveal some interesting regularities when one studies enough verbs with the preverb or prefix in question. McIntyre (2002) argued that many non-spatial uses of particles seem to make the same semantic contribution in multiple particle verbs regardless of whether the contribution of the base is predictable (cf. *fool around* and *muck around*).

Many of these particle senses are construction-specific in that the particle’s semantic contribution is only found in verb-particle constructions and may be further limited to bases with particular semantic properties. (A related idea is Zeller's (2001) suggestion that particles are semantic affixes whose meanings are only licensed by the structural adjacency to a verb.) According to McIntyre, the idiosyncrasy in particle verbs can be induced by construction-specific interpretational rules that he calls 'stipulated composition rules' (McIntyre, 2002: 98).

Such construction-specific meanings are well attested for German, Russian, and English. In many cases, it is even possible to make hypotheses about the routes along which different preverbs/prefixes/particles developed their meanings from spatial ones onward. However, it is not perfectly clear how this construction-specific meaning is born. Even when we can identify the contribution of the base, it often seems somewhat counterintuitive to assign the remaining sememes to the preverb/prefix/particle part. For example, one could argue that, though we cannot derive the meaning of *sing on* by composing the spatial meaning of *on* with *sing*, this is not necessary for compositional analysis. We only need to assume that *on* has a second, non-spatial meaning that makes a consistent contribution to the semantics of the verb (Cappelle, 2005; Larsen, 2016). However, even if one accepts that complex verbs like *sing on* are compositional in this sense, there are other cases where one cannot speak of a 'compositional' combination of a verb and particle since the root is synchronically arbitrary. For instance, *rabbit on* 'talk incessantly' cannot be called compositional, as no other use of *rabbit* has a meaning associated with talking. Similar points can be made regarding combinations with *off*, such as *piss off* and others mentioned above. A Russian example might be *za-sobachitj* 'hit, strike', which is historically related to *sobaka* 'dog', but has no discernible semantic connection with this word.

These language units are parsable but not compositional in the traditional Langacker (1987) sense: the meanings of their elements can be deduced only from the general constructional meaning, and for this to be possible, the latter must be readily available.

The preverb/prefix/particle here is the main driver of the construction, and there is less focus on the base verb or element surfacing as a verb. There is, thus, more freedom to use things that are not normally verbs, and the precise semantic relation between these items and the overall meaning of the construction is less important than in a compositional scenario. This explains why many complex verbs of this type involve bases that are not normally used as verbs without the preverb/prefix/particle. Such linguistic expressions are sometimes called complex denominal verbs in the literature (Stiebels, 1998; Fontanals, 2001; McIntyre, 2015), but the thing is that they do not necessarily incorporate only nouns. Cf. deadjectival *gross someone out* 'arouse disgust in someone' (< *gross* 'disgusting') or *dumb down* 'simplify', or examples where the item surfacing as a verb has no other relevant uses in the language (*divvy up* 'divide something into parts or shares'). It was shown that when a certain Russian prefix-base pattern was primed in discourse, native speakers were able to arrive at the correct interpretation of even those language units in which a real prefix was combined with a nonce base (Monakhov, 2021).

The groups of complex verbs that represent established combinations of productive non-spatial preverbs/prefixes/particles with different bases can often be viewed as the clines of their nested elements' semantic bleaching. By this, I mean that while the semantic contributions of some of these bases are easily interpretable, some other bases act more like placeholders with only expressive but no descriptive meanings (*piss off*, *fuck up*; cf. Brems, 2003), and still other bases no longer exist as independent words (*eke out*, *mete out*). It should also be noted that the arbitrariness of the bases is gradable. For example, *soldier on* ('continue to do something showing bravery, as if one were a soldier') is more clearly motivated than *rabbit on*, and other combinations like *pipe up*, *key in*, or *pan out* are partly motivated for some speakers but totally arbitrary for others.

I have no diachronic evidence to support this claim, but my language intuition tells me that, at least in Russian, these clines tend to run parallel to the time axis so that the most delexicalised instances are the latest to appear. However, not all unmotivated bases are necessarily historically younger than fully compositional formations. There is at least one other potential source of their arbitrariness. In English, some bases originally combined with particles compositionally but then fell out of use, surviving only in particular particle verbs that were memorised. For example, *lap up* comes from the obsolete verb *lape* 'drink' and *eke out* from the obsolete noun *eke* 'supplement', but today, both of them look totally idiosyncratic.

To sum up so far, apart from two clear-cut cases of linguistic expressions being either decomposable, linguistically productive, and compositional in meaning or non-decomposable, non-productive, and non-compositional, there seems to be a special third case of linguistic expressions that are decomposable, (semi?)productive and parsable: their general meanings often cannot be inferred from the meanings of their components, but the meanings of their components can be deduced from their general meanings. Complex verbs of the second, non-compositional and non-productive type are of no particular interest: they are listed diachronic relics that are retrieved from the lexicon. However, the difference between complex verbs of the first and third types merits discussion.

In the previous chapter, I contended that there might exist two different meaning processing models for complex verbs, the distinction between which is not clear-cut and categorical but rather represents an underlying probabilistic continuum. One model implies that each of the elements entering into a combination is equally free to vary; the combination itself is judged by language users to be less semantically complex, more transparent, and tends to be more linguistically productive. Another model assigns some very general sense to the construction as such. Complex verbs of this type are very similar to collocations in the sense that they also consist of a node (conditionally independent element) and a collocate (conditionally dependent element). Such combinations of linguistic items are generally more semantically complex and less transparent because a collocate's meaning does not coincide with the meaning of the respective free element (even if it exists) and has to be parsed out from what is available.

German and Russian provide useful insight into the problem of how the general constructional meaning of complex verbs is acquired. Derivational elements of these verbs can, generally speaking, be subsumed into two categories: spatial and non-spatial. It seems to be a general consensus that non-spatial meanings have developed from spatial ones, not only in German, Russian, and English but also in many other Indo-European languages (Rousseau, 1995; Vincent, 1999; Dehé et al., 2002; Amiot, 2004; Cappelle, 2005; Cuzzolin, Putzu, and Ramat, 2006; Iacobini and Masini 2007; Iacobini, 2009; Köper and Schulte im Walde, 2016; Monakhov, 2021). However, it is far from clear how exactly these processes unfolded.

I hypothesise that at the first stage of development, different preverbs/prefixes/particles with spatial meanings are combined with verbs so that they satisfy these verbs' argument structures (Stiebels, 1996; McIntyre, 2007), thus giving rise to complex verbs whose meaning is the sum of the meanings of their parts. One can compare two very similar sentences from German and Russian (note that the provided English translation also qualifies as an example of this pattern):

German

(19)	*Setzen*	*Sie*	*die*	*Zahnprothese*	*ein.*	[in	den	Mund]
	put	you	the	dental-protheses	in	[in	the	mouth]

'Put your false teeth in. [in the mouth]'

Russian

(20)	*V-stavjte*	*zubnye*	*protezy.*	[v	rot]
	in-put	dental	protheses	[in	mouth]

'Put your false teeth in. [in the mouth]'

Since one verb typically combines with many preverbs/prefixes/particles to encode different spatial meanings (as in Russian *na-pisatj* 'write on', *v-pisatj* 'write in', *nad-pisatj* 'write above', *pod-pisatj* ‘write under’), such instances become generalised as constructions of the form **[_____]**$_{\textbf{PREFIX}}$ **+ BASE** with one empty slot and one fixed element (Diessel, 2019). Next, presumably after the number of unique bases associated with this particular preverb/prefix/particle reaches a certain threshold, a new construction of the form **PREFIX + [_____]**$_{\textbf{BASE:(X>)V}}$ comes into existence by means of abstraction and categorisation.[6] This new construction then licenses certain bases to fill its empty slot, thus serving as a template with an off-the-shelf general (non-spatial) meaning for which the inserted lexical material provides a necessary specification. Importantly, some of these constructions may license the insertion of bases that have already been combined with the same preverb/prefix/particle in its spatial meaning, thus resulting in polysemous complex verbs like the German spatial *auf-nehmen* ‘pick up, lift up’ and the non-spatial *auf-nehmen* ‘open, start’.

In contrast to the spatial type, non-spatial constructions do not involve the satisfaction of verbal arguments. Systematic particle uses that do not fulfil normal arguments of the verb and are sometimes labelled ‘adjunct-like’ in the literature (Stiebels, 1996; McIntyre, 2007) are easy to find in English. Similar cases are also attested in German and Russian:

German

(21)	*Man*	*kocht*	*die*	*Kartoffeln*	*vor.*
	one	cooks	the	potatoes	before / in front of

‘One prepares potatoes ahead of time.’

Russian

(22)	*On*	*za-gotovil*	*mnogo*	*edy*	*vprok.*
	he	before/behind-prepared	a lot of	food	for the future

‘He prepared a lot of food for the future.’

I interpret the process whereby complex verbs of this type are produced as an instantiation on the morphological level of the so-called ‘semantic coherence principle’ of construction grammar (Goldberg, 2006) which implies that constructions attract lexical items compatible with the semantic specifications of certain slots. In particular, this means that each specific complex verb

---

[6] Some remarks on notation are as follows. (1) Henceforth I will be using the term ‘prefix’ in these formalisms because my hypothesis was tested exclusively on Russian data. With German, it would be more conventional to replace ‘prefix’ with ‘preverb’. Applying the model to English verb-particle constructions would require changing not only the name of one element (‘particle’ instead of ‘prefix’) but also the linear order of the elements: BASE + **[_____]**$_{\textbf{PARTICLE}}$ and **[_____]**$_{\textbf{BASE:(X>)V}}$ + PARTICLE. (2) The subscript (X>)V in the second formalism is used to denote that the prefix accepts a base of category (X>)V — a verb or something that is converted into a verb.

with idiosyncratic meaning must be construable as an instance of the more general construction of the form **PREFIX +** $[\_\_\_\_\_]_{\mathrm{BASE:(X>)V}}$. Following this line of reasoning, one can account for the fact that non-spatial complex verbs with a certain preverb/prefix/particle often come in groups of numerous members such that the meanings of derivations are almost identical, although the meanings of their bases might have nothing in common. I believe this analysis is non-contradictory to both Zeller's idea of particles as semantic affixes that have certain selection restrictions stipulating the kinds of verbs they can combine with (Zeller, 2001) and Stiebels's (1996) theory of lexical operations that can change a verb's lexical entry to allow it to accommodate a particle.

The rest of this chapter is dedicated to testing, on Russian data, the hypothesis that complex verbs with spatial and non-spatial meanings represent two different constructions. The chapter is structured as follows. In study 1, I provide experimental evidence that native speakers, when asked to manipulate complex verbs by changing either their prefix or their base, reveal significant preferences for changing the prefixes of spatial verbs and the bases of non-spatial verbs.

In study 2, I draw on the idea that one can identify the construction of a specific complex verb by estimating the ratio of two transitional probabilities: P (prefix | base) and P (base | prefix). I provide empirical, corpus-based evidence that for spatial complex verbs, the probability ratio P (prefix | base) / P (base | prefix) is, on average, less than one, while for non-spatial verbs, it is greater than one. I interpret this difference as confirming the hypothesis of two constructions — $[\_\_\_\_\_]_{\mathrm{PREFIX}}$ **+ BASE** and **PREFIX +** $[\_\_\_\_\_]_{\mathrm{BASE:(X>)V}}$ — since, intuitively, one would expect to find that the fixed element communicates less information about the filler than the filler communicates about the fixed element (cf. Gries and Stefanowitsch, 2004).

In the first part of study 2, I show that the transitional probabilities ratios obtained for 2,566 Russian complex verbs are strongly negatively correlated with these verbs' compositionality scores, which is in line with my initial assumption that complex verbs of the form **PREFIX +** $[\_\_\_\_\_]_{\mathrm{BASE:(X>)V}}$ cannot be called compositional in the traditional sense. In the second part of study 2, two linear regression models are trained to predict the degree of compositionality of a given verb using either the ratio of transitional probabilities or the derivation to base frequency ratio, an alternative measure proposed by Hay (2001, 2003) for assessing the degree of decomposability of multi-morphemic words. The greater predictive power of the former is reported.

8.2 Study 1. Experimental evidence of the existence of two complex verbs' constructions

8.2.1 Hypothesis

One simple way to see whether spatial and non-spatial prefixed verbs are processed differently is to provide native speakers with a respective linguistic item and ask them to write down the first word they can think of that differs from the presented word by either its prefix or its base. If verbs with spatial meanings are indeed constructions of the form **[_____]**$_{\mathbf{PREFIX}}$ **+ BASE** with an empty slot for the prefix, then such a test will reveal participants' preference to manipulate derivational elements of these expressions to produce words with the same base but with different prefixes. Conversely, if verbs with non-spatial meaning are, as I hypothesise, constructions of the form **PREFIX + [_____]**$_{\mathbf{BASE:(X>)V}}$ with an empty slot for the base, then participants will most likely keep the prefix, as a fixed element, unchanged and manipulate the base.

8.2.2 Stimuli

Shvedova (1980) lists 28 verbal prefixes in Russian, of which:

- 17 prefixes are not only historically related to prepositions but also have prepositional counterparts in modern Russian: *v-* (*v* 'in, at'), *do-* (*do* 'to, before'), *za-* (*za* 'for, behind'), *iz-* (*iz* 'from, out of'), *na-* (*na* 'on'), *nad-* (*nad* 'over, above'), *o-* (*o* 'about'), *ob-* (*ob* 'about'), *ot-* (*ot* 'from'), *po-* (*po* 'along, by'), *pod-* (*pod* 'under'), *pred-* (*pered/ pred* 'before, in front of'), *pri-* (*pri* 'by, at'), *pro-* (*pro* 'about, of'), *s-* (*s* 'with'), *so-* (*so* 'with'), and *u-* (*u* 'from, by');
- 11 prefixes have no prepositional counterparts in modern Russian; this group encompasses morphemic borrowings, prefixes that have non-prepositional origin and prefixes derived from prepositions that are no longer part of the Russian language: *de-*, *dis-*, *vz-*, *voz-*, *vy-*, *nedo-*, *niz-*, *pere-*, *pre-*, *raz-*, and *re-*.

Almost all Russian verbal prefixes, both prepositional and non-prepositional, are polysemous, with the number of meanings ranging from 2 (for example, *v-*) to 10 (for example, *pere-*). For the experiment, all meanings of all prefixes listed by Shvedova (1980) were taken into consideration: 91 meanings for prepositional prefixes and 34 meanings for non-prepositional prefixes — 125 in total. For each meaning, one verb was randomly selected from the list of examples provided by Shvedova (1980). The whole set of experimental stimuli can be found in Appendix 6.

8.2.3 Experimental design and participants

The experiment was completed online. I used Yandex Toloka, a Russian crowdsourcing service analogous to Amazon Mechanical Turk, to conduct the experiment. Instructions for the participants read as follows (translated from Russian):

> In each task, you will be given one Russian prefixed verb. Please write, in each case, the first verb you can think of that differs from the presented one by either its prefix or its base. Please note that your input must contain either the same prefix and a different base or the same base and a different prefix. Otherwise, the assignment will not be accepted.

Each verb was presented in 30 tasks to 30 different people. Yandex Toloka does not grant access to their workers' personal data but allows for some coarse-grained social stratification while assembling pools of users. I made sure that the number of males and females in the set of participants was approximately equal, their age ranging from 18 to 55, all of them having obtained at least upper secondary education. The overall number of tasks equalled 3,750 (125 verbs x 30). A total of 186 native speakers of Russian took part in the experiment, and each participant worked with a random selection of 19 to 22 verbs.

8.2.4 Analysis of the results

Ultimately, 166 submissions were excluded as non-conforming to the instruction, resulting in 3,584 accepted answers. For example, for the verb *ot-gremetj* 'stop rumbling', the following entries were submitted:

- 22 instances of verbs with a different prefix and the same base, among them: *po-gremetj* 'rumble for a while' (two instances), *pro-gremetj* 'emit a rumbling sound' (six instances), *za-gremetj* 'start rumbling' (14 instances);
- seven instances of verbs with a different base and the same prefix, among them: *ot-pravljatj* 'send away', *ot-kalyvatj* 'chip away', *ot-gruzhatj* 'load, ship', *ot-zvenetj* 'stop ringing', *ot-vertetj* 'screw off', *ot-davatj* 'give back', *ot-letatj* 'stop flying'.

I am now interested in how much variability, or uncertainty, there is in the prefix and base part of the results. One simple way to determine this is to calculate, separately, the prefix and base entropy by applying the formula

$$E(X) = -\sum_{i=1}^{N} p(x_i)\log_2(p(x_i))\,.$$

The concept of entropy is used to refer to the measure of randomness or disorder within a system. The formula above clearly shows why entropy is also called the measure of 'expected surprise'. Consider two opposite cases. First, there may be a system where many states are equiprobable, which means that their probabilities (relative frequencies) are comparably low. In this case, one's surprise at finding the system in a particular state will always be great because the process of change is random, and no expectations are formed. Alternatively, there might be a system in which one state is much more probable than the others. In this case, one would expect to find the system in its favourite state and would not be surprised if this expectation was confirmed. Obviously, negative logarithms of small probability values are greater than negative logarithms of high probability values, so the measure of expected surprise (entropy) will be greater for highly disordered systems than for stable systems.

In the previous example with the stimulus *ot-gremetj*, there are four unique prefixes in the output that are used the following number of times each (Table 47):

Table 47. Calculating the prefix entropy of experimental results for the verb *ot-gremetj*

prefixes	*za-*	*ot-*	*po-*	*pro-*
counts	14	7	2	6
probabilities	.48	.24	.07	.21

Applying the formula given above, one can calculate the measure of entropy: 1.20. This logic readily extends to the bases. There are eight unique bases in my example that are used the following number of times each (Table 48):

Table 48. Calculating the base entropy of experimental results for the verb *ot-gremetj*

bases	*vertetj*	*gremetj*	*gruzhatj*	*davatj*	*zvenetj*	*kalyvatj*	*letatj*	*pravljatj*
counts	1	22	1	1	1	1	1	1
probs.	.034	.76	.034	.034	.034	.034	.034	.034

As can be seen from the numbers, one base — namely, *gremetj* — totally dominates the distribution, and so the entropy value for the base part of the example is predictably lower than the prefix part's entropy: 1.02.

Now, let us return to the initial hypothesis. What one would expect given the aforementioned idea of the two constructions can be stated as follows: for the verbs with prefixes encoding spatial relations, prefix entropy will be higher than for the verbs with prefixes encoding non-spatial, derived relations, since the former have an empty slot for the prefix. Conversely, base entropy will be higher for the verbs with prefixes encoding non-spatial relations than for the verbs with prefixes encoding spatial relations, since the former have an empty slot for the base. The distributions of the prefix and base entropy values for all the words in my dataset are visualised in Figure 63.

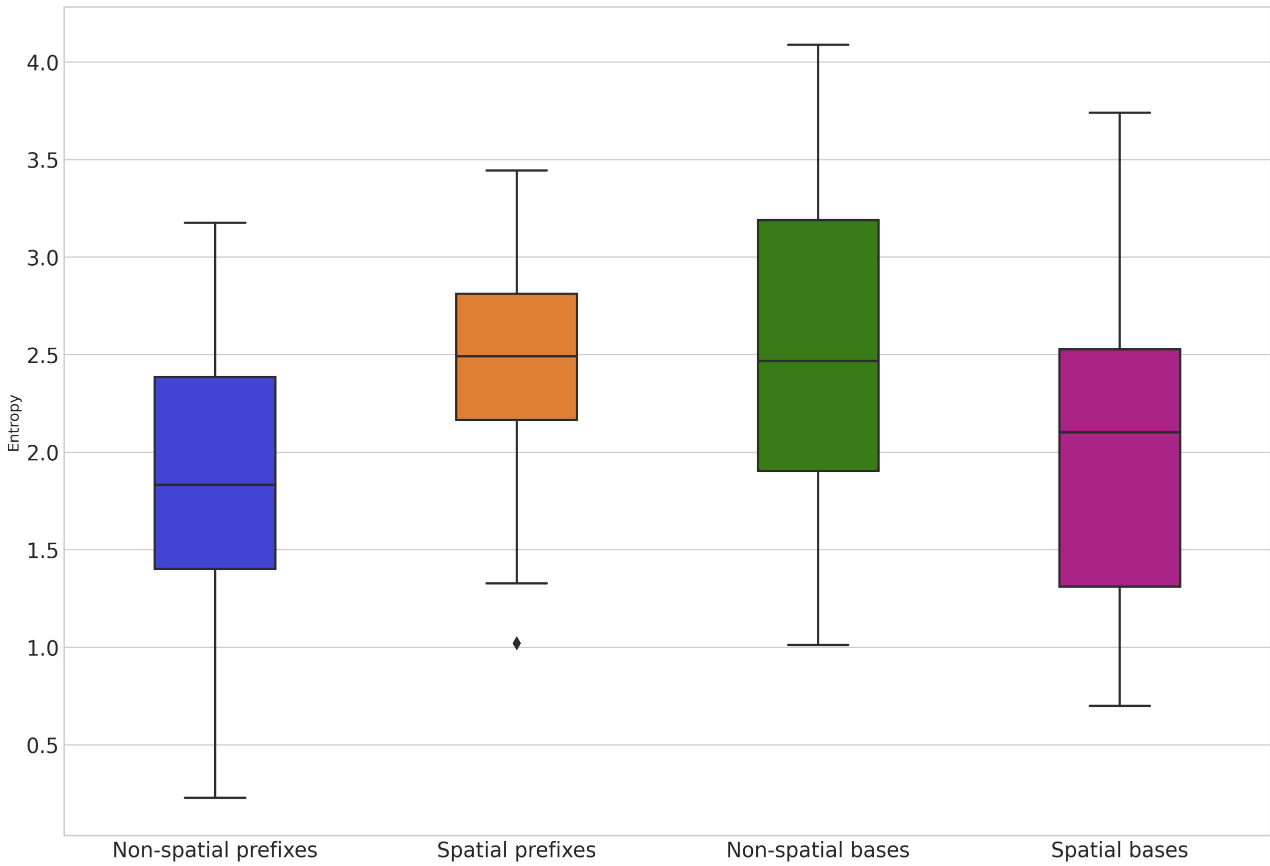

Figure 63. Boxplots of the entropy values

The statistics and *p*-values of the *t*-tests for independent samples are presented in Table 49. They are shown separately for the words with prepositional prefixes, the words with non-prepositional prefixes, and all prefixes without regard to their etymology. In the last row of the table, one can find the median ratios of each word's prefix entropy value to its base entropy value.

Table 49. Inferences on the difference of mean entropy values between verbs with spatial and non-spatial meanings

	all prefixes		prepositional		non-prepositional	
	spatial	non-spatial	spatial	non-spatial	spatial	non-spatial
prefix entropy	$t = 4.51$ ($p < 0.01$)		$t = 3.07$ ($p < 0.01$)		$t = 3.41$ ($p < 0.01$)	
base entropy	$t = -3.60$ ($p < 0.01$)		$t = -2.72$ ($p < 0.01$)		$t = -2.07$ ($p = 0.04$)	
median ratio	1.26	0.72	1.29	0.77	1.11	0.65

It is clear that the null hypothesis that there would be no difference between the prefix and base entropy values of Russian verbs with spatial and non-spatial meanings can be safely rejected. Significant differences were observed not only for the verbs whose prefixes coincide in form with existing prepositions but also for the verbs whose prefixes do not exist as free morphemes. This means that the empty slot in the construction **[_____]$_{\text{PREFIX}}$ + BASE** can be filled both with elements corresponding to real prepositions and with elements that are parsed out from other complex verbs.

I expect that similar results would be observed both 1) in German, where compositional stimuli like *raus-gehen* 'go out' will produce suggestions like *zu-gehen* 'approach somebody' or *durch-gehen* 'walk through' and idiosyncratic stimuli like *an-braten* 'sear, brown' will produce suggestions like *an-brennen* 'light, burn', *an-knabbern* 'nibble at', or *an-kratzen* 'scratch, dent'; and 2) in English, where stimuli like *put down* will result in suggestions like *put on* or *put under* and stimuli like *brush down* will result in suggestions like *clean down*, *scour down*, or *scrub down*.

8.3 Study 2. Corpus evidence in favour of the two-construction account

8.3.1 Degrees of compositionality and transitional probabilities

The aim of this study is to show that the difference between two complex verbs' constructions under investigation manifests itself in the difference between the prefix → base and base → prefix transitional probabilities of the respective complex verbs. I provide evidence that the ratios of these transitional probabilities correlate with the complex verbs' compositionality scores obtained from corpus data and from a word-embedding model in a way supporting my initial assumption that spatial verbs instantiate constructions of the form **[_____]**$_{\mathbf{PREFIX}}$ **+ BASE** and idiosyncratic verbs instantiate constructions of the form **PREFIX + [_____]**$_{\mathbf{BASE:(X>)V}}$.

In computational linguistics, one of the long-standing problems of sense disambiguation is the automatic prediction of literal versus non-literal language usage (Turney et al., 2011; Gutierrez et al., 2016; Hamilton, Leskovec, and Jurafsky, 2016; Schlechtweg et al., 2017; Wang, Durrett, and Erk, 2018). With regard to prefixed verbs, this distinction is epitomised by the division of verbs into groups encoding spatial meanings and verbs encoding non-spatial meanings. I hypothesise that prefixed verbs encoding spatial meanings represent constructions that have an empty slot for prefixes and tend to be compositional, which means that both elements in combination contribute to a general, additive meaning. On the other hand, non-spatial, idiosyncratic verbs represent constructions that have an empty slot for bases. They tend to be non-compositional but parsable: some very general sense is assigned to the construction as such, which results in the meaning of a filler base not coinciding with the meaning of a respective free element (even if it exists) and needing to be parsed out from what is available.

The very nature of these two constructions, with their reversed positioning of the fixed element and the empty slot, suggests that one can identify the construction to which each particular word belongs by estimating two probabilities: P (prefix | base) and P (base | prefix). Suppose that P (prefix | base) / P (base | prefix) = ε. Then, a linguistic item is more likely to be of the form **[_____]**$_{\mathbf{PREFIX}}$ **+ BASE** if $\varepsilon \leq 1$ and more likely to be of the form **PREFIX + [_____]**$_{\mathbf{BASE:(X>)V}}$ if $\varepsilon > 1$. I expect it to be this way and not the other way around because I assume that the fixed element communicates less information about the filler than the filler communicates about the fixed element. Thus, P (prefix | base) must be greater than P (base | prefix) with idiosyncratic verbs and smaller than P (base | prefix) with spatial verbs.

Based on these premises, one might hypothesise that it would be possible to predict the degree of compositionality of a given verb by taking into account its ratio of transitional probabilities, P (prefix | base) / P (base | prefix).

The probabilities P (prefix | base) and P (base | prefix) can be evaluated empirically, for example, by taking all of the prefixed verbs in a morphemic dictionary of the respective language and looking up the frequencies of interest in the internet corpus of this language. Then, for any word, its P (prefix | base) is equal to the number of that word's tokens divided by the number of tokens of all (prefixed) words with this base, and P (base | prefix) is equal to the number of the given word's tokens divided by the number of tokens of all words with this prefix.

To build a model capable of predicting complex verbs' compositionality degrees based on their transitional probabilities P (prefix | base) and P (base | prefix), I obtained all Russian prefixed verbs included in the *Word-Formation Dictionary of the Russian Language* (Tikhonov, 1985). Overall, there were 6,159 verbs. I decided to constrain the task to verbs with prepositional prefixes for the following reasons. In order to train the model, I need some objective measure of how much spatial meaning a certain prefixed verb encodes. Given the amount of data, manual coding seemed infeasible. Therefore, one option was to get an approximation of this measure from a linguistic corpus, relying on the fact that verbs with spatial meanings encoded by prepositional prefixes are often accompanied by prepositions in Russian, unlike their counterparts with idiosyncratic meanings (cf. Bergsma et al., 2010; Biskup, 2015):

(23)	*na-pisatj*	*na*	*bumage*
	on-write	on	paper
	'write on paper'		

(24)	*na-vratj*	*o*	*proizoshedshem*
	on-lie	about	what-happened
	'lie a lot about what has happened'		

Hence, only verbs with prepositional prefixes could be included in the survey. In my data, there were 4,580 such verbs. Using *ruTenTen11*, an internet corpus of Russian from 2011 provided by Sketch Engine and containing more than 14 billion words (Jakubíček et al., 2013), I queried two types of co-occurrence frequencies for each verb: 1) one where a preposition coinciding in form with the verbal prefix is found within the window of four words to the left of the verb, and 2) another where the same preposition is found within the window of four words to the right of the verb. The choice of window size is consistent with the rules of thumb frequently suggested in the literature. These are based on the observation that a smaller window size focuses on how the word is used and learns what other words are functionally similar to it, while a larger window size captures information about the domain or topic of each word (Hvitfeldt and Silge, 2022; Lin et al., 2015; Levy and Goldberg, 2014). All prefixal and prepositional allomorphs were queried in the corpus separately; the numbers were then combined.

To calculate the final measure of compositionality for each verb, one must control for the verb's and the preposition's overall frequency of use because highly frequent verbs and highly frequent prepositions may often co-occur by random chance. To do this, I used the logDice score, a metric from corpus linguistics that is designed to measure collocation strength (Rychlý, 2008). The logDice score has the following useful features: (1) the score does not depend on the total size of a corpus; (2) its theoretical maximum is 14, in cases when all instances of word A co-occur with word B and all instances of word B co-occur with word A; and (3) its negative values mean that there is no statistical significance of the collocation of words A and B.

This metric is easy to calculate and interpret. For example, the total number of co-occurrences of the verb *na-brositj* 'throw on(to)' and the preposition *na* 'on' is 3,206. The preposition *na* has an overall frequency of 2.41e8, and the verb *na-brositj* has an overall frequency of 8,388. The logDice score of this combination of words is 3.465. This means that the verb *na-brositj* and the preposition *na* tend to appear alongside each other. Hence, one can conclude that *na-brositj* is likely to encode compositional spatial meaning. On the other hand, the verb *na-petj* 'sing along' is randomly seen in the vicinity of the preposition *na* (logDice score equal to -0.263), which supports the native speaker's intuition about the general constructional meaning of this linguistic item being non-spatial.

Table 50 provides some further results for illustration. From this table, it becomes clear that, evaluated as I suggest, the compositionality of Russian prefixed verbs can be viewed as a continuum, with some of the verbs retaining much of the prepositional spatial meaning and some drifting far away from it.

Table 50. LogDice scores for some Russian verbs with the prefix *na-*

verb	meaning	logDice score
na-pisatj	'write on'	7.324
na-brositj	'throw on(to)'	3.465
na-valitj	'pile up'	1.844
na-lovitj	'catch a lot'	0.382
na-petj	'sing along'	-0.263
na-soritj	'litter'	-1.349

One problem with the logDice score is that it is undefined for the case of no co-occurrence of a particular combination of words because this requires taking a log of zero. Therefore, I had to exclude all the verbs for which not a single instance of the preposition that coincides in form with the verbal prefix was found within the specified window in the whole corpus. This is justified by the fact that, given the ubiquitousness of prepositions, for sufficiently frequent verbs, at least some co-occurrences should happen by random chance. After pruning, there were 2,566 complex verbs left in the dataset. The distribution of the verbs' compositionality measures (logDice scores) can be found in Figure 64.

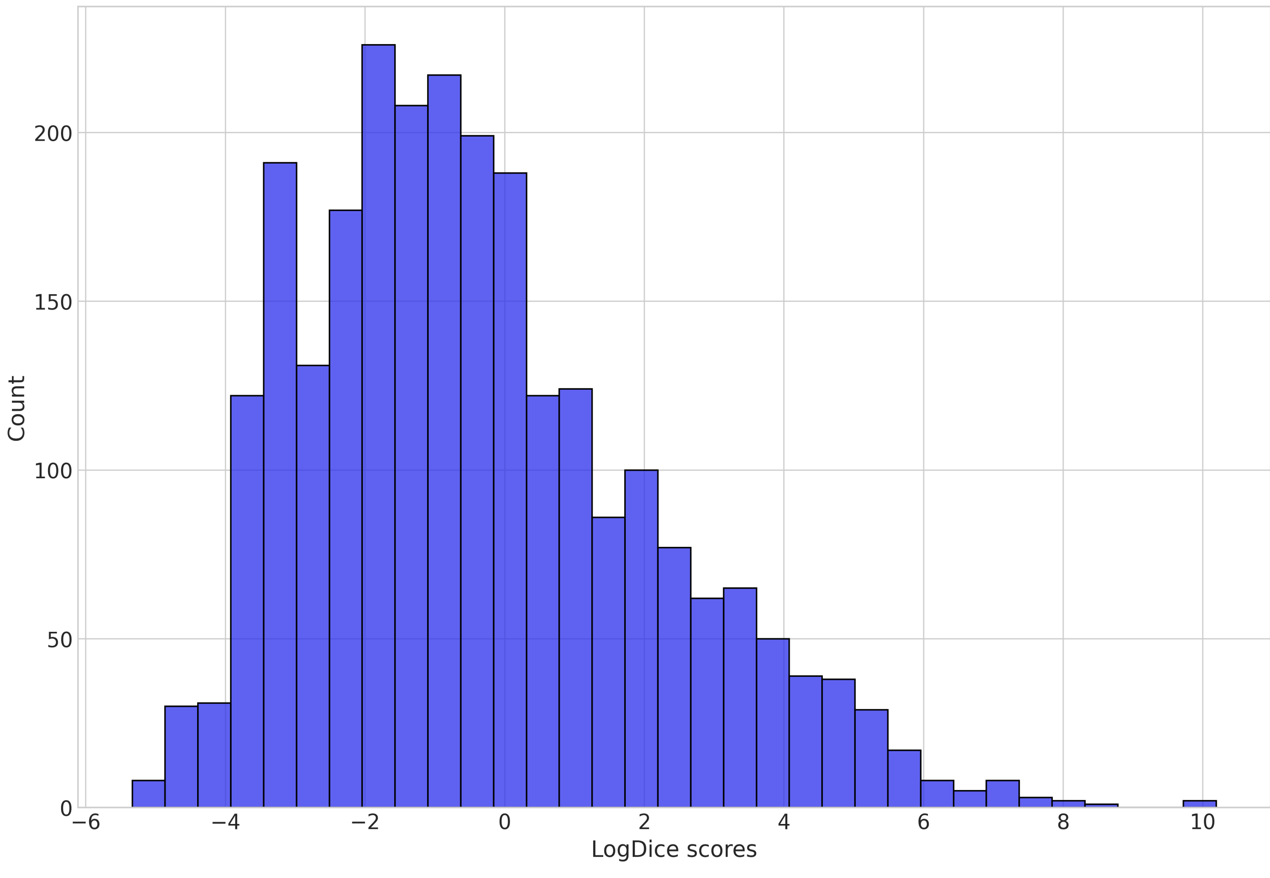

Figure 64. Histogram of the logDice scores

Before creating a predictive model, I wanted to make sure that there indeed exists some relationship between Russian complex verbs' degrees of compositionality and transitional probabilities ratios. Once again, my hypothesis implied that for the prefixes that encode mostly spatial meanings, there would be many verbs with P (prefix | base) ≤ P (base | prefix) so that the ratio of the two, averaged across all items, would be less than one (or zero on the log scale). Alternatively, with the prefixes that encode mostly non-spatial meanings, one would find many lexemes with P (prefix | base) > P (base | prefix) so that the ratio of the two, averaged across all items, would be greater than one.

To illustrate, let us consider two Russian verbs with the same prefix *v-*: *v-chinitj* ‘submit, file (a complaint)’ and *v-bezhatj* ‘run into’. The values of interest for each of them are given in Table 51.

Table 51. Exemplary calculations for the verbs *v-chinitj* and *v-bezhatj*

values	*v-chinitj*	*v-bezhatj*
L — number of word’s tokens	485	17,455
S — number of tokens of all words with this base	512,401	1,768,006
P — number of tokens of all words with this prefix	753,546	753,546
P (prefix | base) = L / S	0.0009	0.009
P (base | prefix) = L / P	0.0006	0.02
ratio P (prefix | base) / P (base | prefix)	1.47	0.42
log ratio P (prefix | base) / P (base | prefix)	0.38	-0.85
logDice score	-3.72	3.48

Based on these calculations, I would expect to find that if the prefix-base construction with *v-* encodes mostly spatial meanings, there will be many verbs like *v-bezhatj*; otherwise, there will be many verbs like *v-chinitj*. The general picture for all prefixes in my data will then be that of the negative correlation between average compositionality measures and average transitional probabilities ratios (log-transformed).

Such a correlation was indeed observed (Figure 65, left-hand panel; ρ = -0.56, p = 0.02). We can easily convince ourselves that prefixes with a high degree of compositionality (*pred-*, *nad-*, *pod-*, *v-*) are characterised by low values of transitional probabilities ratios, which signifies that for the verbs with these prefixes, on average, P (prefix | base) ≤ P (base | prefix). On the other hand, prefixes that represent constructions that have acquired numerous non-spatial meanings over the course of their development (*na-*, *o-*, *po-*, *s-/so-*, *za-*) reveal high values of transitional probabilities ratios, which signifies that for the verbs with these prefixes, on average, P (prefix | base) > P (base | prefix).

Up to this point, I had estimated transitional probability P (prefix | base) as a number of a certain word’s tokens divided by the number of tokens of all (prefixed) words with the respective base and transitional probability P (base | prefix) as a number of a certain word's tokens divided by the number of tokens of all words with the respective prefix. One might wonder whether the relationship between transitional probabilities ratios and compositionality measures is dependent upon the frequency of complex verbs and their bases. This can be easily determined by estimating the same two transitional probabilities from type rather than token frequencies.

P (prefix | base) can be estimated as 1 divided by the number of prefixes that combine with a given base, and P (base | prefix) as 1 divided by the number of bases that combine with a given prefix. By performing these calculations on the dataset of Russian prefixed verbs, I ascertained that type- and token-based transitional probabilities ratios are almost perfectly correlated ($\rho = 0.9$, $p < 0.001$), which means that $P(TP_i \mid C_i, F_i) = P(TP_i \mid C_i)$, where C_i is the compositionality measure of a prefixed verb i, TP_i is this verb's transitional probabilities ratio, and F_i is its token frequency.

8.3.2 Distributional semantic estimates of the degree of compositionality

My assumption that Russian prefixed verbs collocating with prepositions necessarily encode spatial meanings might sound too strong. Indeed, in some cases, unexpectedly high compositionality scores were attested for the verbs with obviously non-spatial meaning. For example, the verb *za-rezatj* 'slaughter, kill with a knife' has a relatively high logDice score of 2.01, and the score of another verb with the same base, *na-rezatj* 'slice', is even higher, at 5.44. However, there is nothing compositional in these verbs' senses. High logDice scores for *za-rezatj* and *na-rezatj* are due to the fact that these words frequently co-occur with prepositions that do not encode spatial meanings themselves, as in the following examples:

(25)	*na-rezatj*	*na*	*kuski*
	on-cut	into	pieces
	'slice into pieces'		

(26)	*za-rezatj*	*za*	*kopejku*
	before/behind-cut	for	kopeck
	'slaughter for nothing'		

The number of such cases is relatively small and is unlikely to undermine the validity of my conclusions in general. However, I wanted to find out whether the results would hold with a different compositionality measure.

An alternative way of obtaining complex verbs' compositionality measures is suggested by the distributional hypothesis, which states that similarity in meaning results in similarity in linguistic distribution (Firth, 1957). Words that are semantically related tend to be used in similar contexts. Hence, by reverse-engineering this process — that is, coding words' discourse co-occurrence patterns with multi-dimensional vectors and performing certain algebraic operations on them — distributional semantics can induce semantic representations from contexts of use (Boleda, 2020). It is well-established that the similarity of words' vector representations goes beyond simple syntactic regularities (Pennington, Socher, and Manning, 2014; Mikolov et al., 2013; Rehurek and Sojka, 2011) and that vector space models perform well on tasks that involve measuring the similarity of meaning between words, phrases, and documents (Turney and Pantel, 2010).

More importantly, for the purposes of my study, vector space models have been used for assessing the degree of compositionality of complex linguistic expressions, notably nominal compounds (Cordeiro et al., 2019) and particle verbs in English (Bannard, 2005) and German (Bott and Schulte im Walde, 2014). These analyses generally assume that multiword expressions are highly variable in compositionality and that if the meanings of some of them can be described as the sum of the meanings of their parts, then a distributional semantic model will reveal significant similarity between vectors for a compositional expression and for the combination of the vectors of its parts, computed using some vector operation. Conversely, the lack of such similarity might be interpreted as a manifestation of complex expressions' idiomaticity.

Applying the aforementioned principle to multi-morphemic words, specifically to Russian complex verbs whose prefixes have corresponding free elements, seems a straightforward extension. My hypothesis is that one can model the difference between spatial and non-spatial prefixed verbs by performing simple algebraic operations on semantic vectors representing the target verbs and their subparts. Specifically, I am interested in estimating the ratio **cosine**$(\overrightarrow{V}, \overrightarrow{C})$ / **cosine**$(\overrightarrow{V}, \overrightarrow{P})$, where $\overrightarrow{V}$ is a vector for the verb, $\overrightarrow{C}$ is a 'compositionality operation vector' obtained by summation of the vectors for the base and preposition coinciding in form with the prefix, and $\overrightarrow{P}$ is a 'parsability operation vector' obtained by subtraction of the vector for the base from the verb vector $\overrightarrow{V}$. I hypothesise that this ratio will be greater for the prefixed verbs that are instantiations of the **[_____]PREFIX + BASE** constructions and smaller for the prefixed verbs that are instantiations of the **PREFIX + [_____]BASE:(X>)V** constructions.

While the assumption about the results of the vector addition operations is intuitively clear, it might not be obvious why I believe that the results of the vector subtraction operations will be of any importance. As discussed in the introduction, my hypothesis implies that constructions with an empty slot for the prefix are instantiated by spatial complex verbs with compositional meaning, and hence, the meaning of the whole expression, in this case, can be most adequately represented as the sum of the meanings of the prefix and the base ('compositionality operation vector'). In contrast, constructions with an empty slot for the base are instantiated by complex verbs with idiosyncratic meaning, such that the prefix is the main driver of the construction, and the base only provides the necessary specification for the general constructional meaning. Hence, even if one removes this semantic specification component ('parsability operation vector'), this procedure should not be completely detrimental to the composite conceptualisation.

To demonstrate, let us consider two groups of 10 nearest neighbours (i.e., words whose vector representations and, by extension, co-occurrence patterns are most similar to the target word) of the vectors obtained as a result of (1) subtraction of the vector of the base *pisatj* 'write' from the vector of the verb *na-pisatj* 'write on' and (2) subtraction of the vector of the base *lovitj* 'catch' from the vector of the verb *na-lovitj* 'catch a lot' (Table 52). The vectors for this example were obtained from the word2vec continuous skip-gram model provided by the *RusVectōrēs* project (https://rusvectores.org; Kutuzov and Kuzmenko, 2017). The model includes functional words and was trained on the Russian National Corpus and the Russian Wikipedia dump of 2018.

Intuitively, the numbers in Table 52 tell us that prefixed verbs with spatial meanings, like *na-pisatj*, strongly overlap in semantics and distribution with their bases. For this reason, there is almost nothing meaningful left in the vectors of these verbs after the key components have been subtracted. The output in the left-hand panel of Table 52 contains mostly irrelevant noise, including proper nouns like *Neil Young*, interjections like *nu-ka* 'come-on.INT', imperative or hortative forms like *davajte-ka* 'let us', and so on.

The situation is very different with non-spatial prefixed verbs like *na-lovitj*. Here, as the items in the right-hand panel of Table 52 suggest, the result of vector subtraction does encode some conceptual entity of its own — some very general sense that is attributed to the construction as such (consider the variability of specific lexical meanings in the set of the words aligned with *na-lovitj*: *na-rubitj*, *na-streljatj*, *na-pech*, *na-kopatj*, *na-gotovitj*).

Table 52. Output of the vector subtraction operations on the verbs *na-pisatj* and *na-lovitj*

na-pisatj		*na-lovitj*	
nearest neighbours of $\overrightarrow{P}$	cosine similarity $(\overrightarrow{V}, \overrightarrow{P})$	nearest neighbours of $\overrightarrow{P}$	cosine similarity $(\overrightarrow{V}, \overrightarrow{P})$
na-pisatj 'write on'	.48	*na-lovitj* 'catch a lot'	.69
kaver-versii 'cover.PL'	.33	*na-rubitj* 'chop a lot'	.40
pozhalujsta 'please'	.32	*na-streljatj* 'shoot a lot'	.38
spoj 'sing.IMP'	.32	*na-pech* 'bake a lot'	.37
davajte-ka 'let us'	.31	*na-kopatj* 'dig a lot'	.33
Neil Young	.31	*na-gotovitj* 'cook a lot'	.33
vos-hoditj 'rise'	.31	*s-varitj* 'boil'	.33
za-pisatj 'record'	.31	*s-vezti* 'bring together'	.33
Cliff Richard	.30	*po-zharitj* 'fry'	.33
nu-ka 'come-on.INT'	.30	*zakusochka* 'little snack'	.32

Note: cosine similarity ranges from 0 to 1 for vectors with positive values.

Note that the vectors obtained by subtraction also show significant differences in terms of their cosine similarities to the initial lemmas' vectors: 0.48 for *na-pisatj* and 0.69 for *na-lovitj*. The cosine similarities of the initial lemmas' vectors and the vectors obtained by summation (preposition + base) are also different, but this difference goes in the opposite direction: 0.57 for *na-pisatj* and 0.48 for *na-lovitj*. Thus, the ratio **cosine**$(\overrightarrow{V}, \overrightarrow{C})$ / **cosine**$(\overrightarrow{V}, \overrightarrow{P})$ is equal to 1.18 for *na-pisatj* and to 0.69 for *na-lovitj*, exactly as I expected.

Some words of caution may be appropriate at this point. Vector space models have well-known limitations. Specifically, traditional word2vec (Mikolov et al., 2013) and GloVe (Pennington, Socher, and Manning, 2014) models tend to perform worse when confronted with word formation in morphologically rich languages like German (Köper, Scheible, and Schulte im Walde, 2015) and Russian (Drozd, Gladkova, and Matsuoka, 2016). Thus, of all the existing non-contextualised pretrained vector models of the Russian language, the FastText model seemed the best suited for the purposes of this study. While other popular models ignore the morphology of words by learning their vectors, in the FastText model, a vector representation is associated with each character n-gram, and words are represented as the sums of these n-gram vectors (Bojanowski et al., 2017)[7].

Using this model, I obtained for each verb in my data its **cosine**($\vec{V}$, $\vec{C}$) / **cosine**($\vec{V}$, $\vec{P}$) ratio and then averaged the values prefix-wise. The results of the correlation analysis of these measures with the log-transformed transitional probabilities ratios are visualised in Figure 65 (right-hand panel). The strength of association is even greater than the one observed for the logDice scores, but the direction of association is the same, in line with my expectations (ρ = -0.69, p = 0.002).

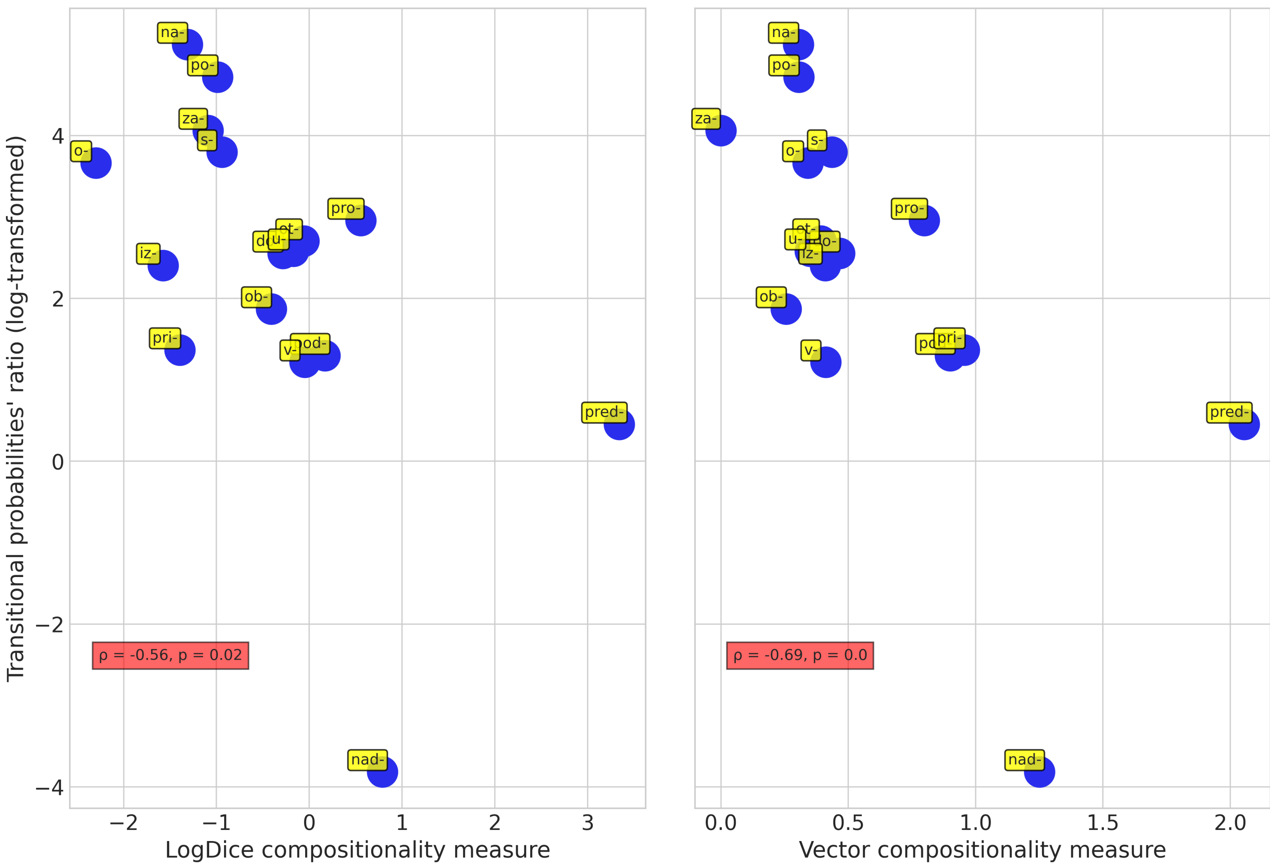

Figure 65. Correlation of transitional probabilities ratios with compositionality measures

One might argue that distributional semantic models represent abstractions over attested data, and further modifying the vectors for assessing multi-morphemic words' degrees of compositionality is yet another step away from the empirical basis. However, while each of my two measures of

[7] The reason I used word2vec model to obtain the examples in Table 52 is a technical one. The Python implementation of the FastText model that I worked with makes it much harder, from a computational perspective, to find nearest neighbours by vector, so I chose to resort to word2vec to illustrate the main idea. However, I believe FastText to be more reliable for the purpose of large-scale data analysis.

compositionality on its own may be considered problematic in some respects, the fact that they, having been obtained independently, agree with each other so well should greatly reinforce one's confidence in either of them. In fact, the similarity between the two plots in Figure 65 is truly remarkable and suggests that both proposed ways of measuring Russian prefixed verbs' degrees of compositionality are reliable.

8.3.3 Automatic prediction of Russian prefixed verbs' meanings

This section will return to the idea of building a predictive model of the prefixed verbs' degrees of compositionality. Taking into account the observed negative correlation between the compositionality measures and transitional probabilities' ratios of Russian prefixed verbs, I hypothesise that individual words with greater compositionality scores (those more likely to encode spatial meanings) would be characterised by a P (prefix | base) approximately equal to or lower than their P (base | prefix), while words that have lower compositionality scores (those more likely to encode non-spatial, construction-specific meanings) would tend to reveal P (prefix | base) values that are greater than their P (base | prefix) values.

It would also be interesting to compare the accuracy of this model with the accuracy of another model built for the same purposes but basing its decisions not on transitional probabilities ratios but on the derivation to base frequency ratio, the measure proposed by Hay (2001, 2003). The logic behind this measure is as follows. According to Hay, the degree of decomposability of a given item depends on the frequency of the derived word relative to its base. With most complex words, the base is more frequent than the derived form, so the relative frequency is less than one. Such words, Hay argues, are more easily decomposed; that is, they are more likely to be accessed via a morpheme-based route. In the opposite case, when the derived form is more frequent than the base, a whole-word bias in parsing is expected. This has consequences for semantics, and such words become less transparent and more polysemous.

Given Hay's method of measuring linguistic relativity, one might hypothesise that it would be possible to automatically classify prefixed verbs encoding spatial and non-spatial meanings by assigning decomposable lexemes with Frequency(base) > Frequency(prefix+base) to the first group and non-transparent lexemes with Frequency(base) < Frequency(prefix+base) to the second group. Comparing the accuracies of these two models will show which way of estimating complex verbs' degree of analysability is more accurate — the one based on transitional probabilities ratio or the one based on derivation to base frequency ratio.

To illustrate, let us consider two Russian verbs with the same base *rezatj* 'cut': *ot-rezatj* 'cut off from' and *za-rezatj* 'slaughter, kill with a knife'. The values of interest for each of them are given in Table 53.

Table 53. Exemplary calculations for the verbs *ot-rezatj* and *za-rezatj*

values	*ot-rezatj*	*za-rezatj*
P (prefix | base)	0.095	0.019
P (base | prefix)	0.117	0.009
ratio	0.095 / 0.117 = 0.806	0.019 / 0.009 = 2.088
logDice score	7.03	2.01
Hay's parsability measure	201,598 / 199,753 = 1.009	38,002 / 199,753 = 0.19

One can say that these two words epitomise the distinction between spatial and non-spatial types. The verb *ot-rezatj* is an instantiation of the construction of the form **[_____]**$_{\textbf{PREFIX}}$ **+ BASE**. Its meaning can be construed as the sum of the meanings of the base and respective preposition *ot* 'away, from'. We can easily convince ourselves that this meaning is compositional by looking at the verb's very high logDice score of 7.03. The verb *za-rezatj*, on the other hand, is different in that it instantiates the **PREFIX + [_____]**$_{\textbf{BASE:(X>)V}}$ construction, which suggests 'to bring someone to an undesirable state (of unfitness, fatigue, exhaustion, death) through an action identified by the base'. This construction's general meaning is obviously non-compositional as it does not inherit anything from the meaning of the corresponding preposition *za* 'behind', which is confirmed by a significantly lower logDice score of 2.01.

The ratios of transitional probabilities seem to capture this semantic difference correctly. The ratio of *ot-rezatj*, 0.806, is much smaller than the ratio of *za-rezatj*, 2.088. This is in line with my hypothesis that for the prefixes that encode mostly spatial meanings, there will be many verbs with P (prefix | base) ≤ P (base | prefix), while with the prefixes that encode mostly non-spatial meanings, one will find many lexemes with P (prefix | base) > P (base | prefix).

Surprisingly, Hay's parsability measures for these complex verbs are at odds with what can be inferred from their transitional probabilities ratios and with compositionality scores as well. Hay's way of assessing decomposability status suggests that *za-rezatj*, with its score of 0.19, is very likely to be parsed (and thus should encode spatial meaning), and *ot-rezatj*, with its score of 1.009, is more likely to be processed holistically (and thus should encode idiosyncratic meaning). A comparison of the two proposed predictive models can help determine whether this inconsistency is a random fluctuation or is indicative of some problems with either measure.

The process of creating the models ran as follows. For each of the 2,566 prefixed verbs in my data, three numerical values were obtained: 1) the logDice compositionality score, 2) the transitional probabilities ratio (log-transformed), and 3) the derivation/base frequency ratio (log-transformed). The data were randomly split into training (80% of observations) and test (20% of observations) sets and two linear regression models were fit to the training set. Model M regressed compositionality scores on transitional probabilities ratios, while model H regressed compositionality scores on derivation to base frequency ratios. The models' coefficients can be found in Table 54 (both models were fit so as to allow for each prefix's specific baseline, but the coefficients for these factor levels are omitted to avoid clutter).

Table 54. Coefficients of the regression models

model M				model H			
coefficient	estimate	SE	*p*	coefficient	estimate	SE	*p*
constant	-.23	.33	.42	constant	-.63	.36	.08
TP ratio	-.46	.02	< 0.001	DB ratio	.18	.01	< 0.001
prefixes	…	…	…	prefixes	…	…	…

Note (Model M): F =38.6, R^2 = 0.24, p < 0.001.
Note (Model H): F = 17.7, R^2 = 0.12, p < 0.001.

The obtained coefficients from both models were used to make separate predictions about the compositionality scores in the test dataset. The predicted and observed compositionality scores were correlated with each other. The resulting plots are presented in Figure 66: the left subplot represents model M, and the right subplot represents model H. The correlation coefficients of the observed and predicted compositionality scores were found to be significant in both cases (both p-values < 0.001). However, the strengths of the relationships are not the same: r =0.52 for model M and r = 0.30 for model H. Hence, model M makes more accurate predictions about the prefixed verbs' compositionality measures compared to model H.

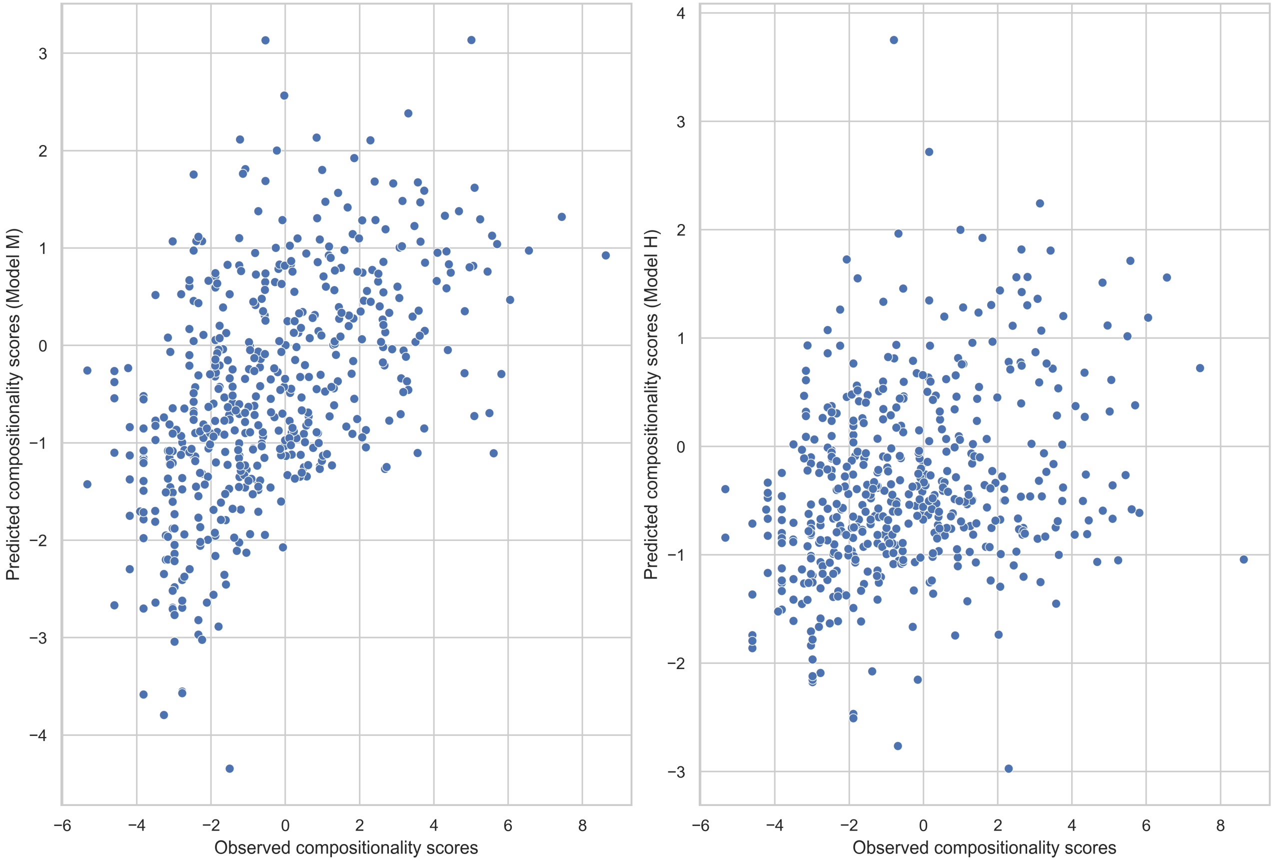

Figure 66. Correlation of the observed and predicted compositionality scores for two models

The most interesting difference between models M and H, however, is in the values of the slope coefficients. Model M predicts that with a one-unit increase in transitional probabilities ratio, the compositionality score will decrease by a factor of -0.46. This is exactly the relationship direction that I expected to find: greater values of the transitional probabilities ratio indicate that for the respective verbs, P (prefix | base) > P (base | prefix), which, in turn, can be taken as a sign that these verbs instantiate constructions of the form **PREFIX + [_____]$_{\mathbf{BASE:(X>)V}}$**, where the lexical meaning of the base only provides a necessary specification for the general constructional meaning.

Model H, on the other hand, predicts that with a one-unit increase in the derivation to base frequency ratio, the compositionality score will increase by a factor of 0.18. Basically, it implies that the less decomposable a word is, the more compositional its meaning will be. This is counterintuitive. The explanation of this anomaly is, however, very simple. Russian prefixed verbs encoding spatial meanings tend to have a higher token frequency than verbs encoding non-spatial meanings: I found a significant positive correlation between the compositionality scores and frequency values of the derived forms in my data (ρ = 0.64, $p < 0.001$). Thus, the phenomenon observed with the verbs *ot-rezatj* and *za-rezatj* (Table 53) was not some random fluctuation but rather a manifestation of the fact that derivation to base frequency ratio may be a biased measure.

One can make sense of the ultimate difference between models M and H by comparing the probabilistic graphical models presented in Figure 67, where *Comp.* stands for the logDice compositionality score, *Freq.* for the token frequency, *TPr.* for the transitional probabilities ratio, and *DBr.* for the derivation to base frequency ratio. Directed edges between the nodes indicate that one node is assumed to exert influence upon another node, while the absence thereof means that no direct relation between two nodes is believed to exist. The numbers are the correlation coefficients.

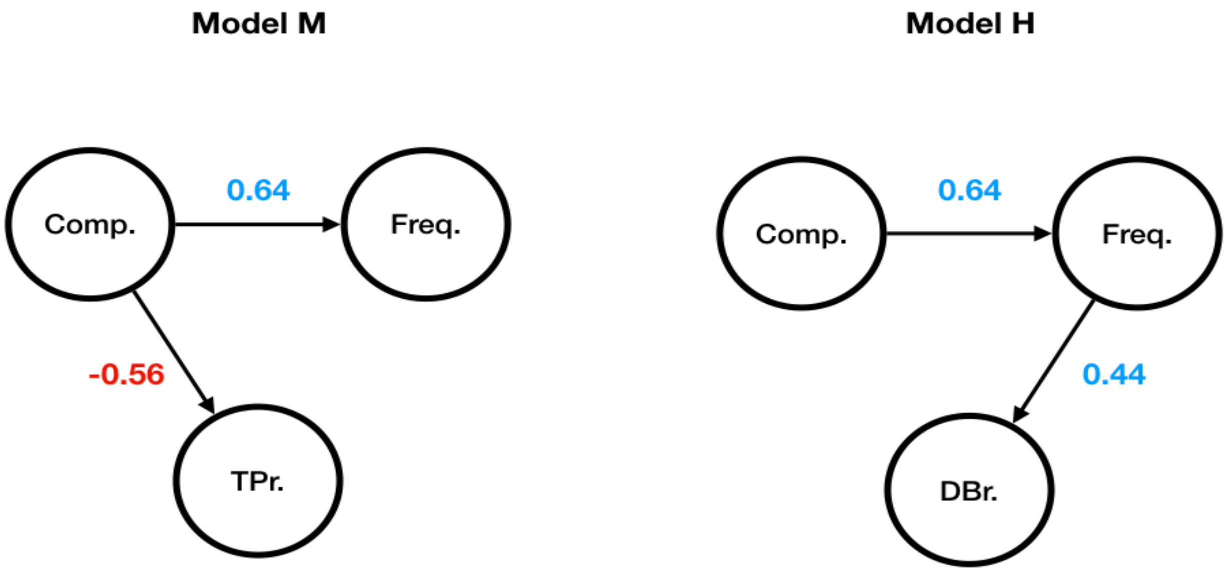

Figure 67. Probabilistic graphical models M and H

Given the structure of the two probabilistic graphical models in Figure 67, it is easy to see that once the token frequency of a particular complex word is observed, its derivation to base frequency ratio does not provide any information about whether the word's meaning is compositional or not (the path between the nodes *Comp*. and *DBr*. is blocked). On the other hand, once the compositionality value of a particular complex word is known, its token frequency no longer influences its transitional probabilities ratio, that is, one can observe P (prefix | base) ≤ P (base | prefix) or P (prefix | base) > P (base | prefix) for both high- and low-frequency words.

8.4 Conclusion

Both experimental and corpus data presented in this chapter suggest that there exist at least two different constructions for Russian complex verbs: $\mathbf{[\underline{\quad\quad}]_{PREFIX} + BASE}$ for verbs with spatial meanings and $\mathbf{PREFIX + [\underline{\quad\quad}]_{BASE:(X>)V}}$ for verbs with idiosyncratic meanings.

In study 1, I provided experimental evidence that native speakers, when asked to manipulate complex verbs by changing either their prefix or their base, reveal a significant preference for changing the prefixes of spatial verbs and the bases of non-spatial verbs. In study 2, I showed that the choice of construction could be predicted by taking into account the discrepancy in probabilities of transition from base to affix and from affix to base. If P (prefix | base) ≤ P (base | prefix), the first construction with an empty slot for the prefix is likely to be chosen. If, on the other hand, P (prefix | base) > P (base | prefix), the second construction with an empty slot for the base becomes more likely. In both cases, the element of the construction that tells us less about its counterpart activates general constructional meaning, while the element that has greater predictive power serves as a filler for the construction's empty slot. I also showed that the distinction between two constructions does not necessarily pertain to the difference in relative frequencies. Token frequencies of bases and lemmas seem to play hardly any role in either of these constructions. Even low-frequency bases may combine with many different prefixes, and low-frequency lemmas adhere to the same constructional patterns as their high-frequency counterparts.

The distinction between two constructions is easily explainable within the framework of construction morphology (Booij, 2010b), where complex words are seen as constructions on the word level. The view that complex words instantiate morphological constructions can be found in Croft (2001) and Goldberg (2006). Some examples of the constructional analysis of complex words include the analysis of English *be*-verbs in Petre and Cuyckens (2008), the analysis of the phrasal verbs of Germanic languages in Booij (2010a), and the analysis of Russian prefixed verbs in Monakhov (2021). Nevertheless, overall, the understanding of the constructional aspects of multi-morphemic word structure is still in its early stages.

One non-trivial contribution to the construction morphology framework made by the current study is the idea that some constructions might arise by way of generalisation over others, which leads to a shift in the positioning of a fixed element and a slot. These results allow us to draw two important conclusions. First, the distinction between two constructions is not a clear-cut categorical one. Rather, there is an underlying probabilistic continuum, and any particular word can be more or less likely to activate either model of meaning processing. Second, there is compelling evidence that constructions with an empty slot for the base are later developments, compared to their compositional counterparts, since the former come into existence by means of analogy and categorisation of the lexical material provided by the latter.

Some limitations of this study must be discussed. First, the potential impact of the examined verbs' polysemy remains unaccounted for. For example, in Russian, there are two instances of the verb *za-brositj*: (1) 'throw something behind some object' and (2) 'stop doing something'. The first one is clearly a compositional expression, an instantiation of the **[_____]**$_{\text{PREFIX}}$ **+ BASE** construction. It is characterised by a high logDice score and, in the experimental setting described in study 1, would most probably produce suggestions like *na-brositj* 'throw on(to) something', *v-brositj* 'throw in(to) something', *pod-brositj* 'throw under something', and so on. In contrast, the second exemplar is an instantiation of the **PREFIX + [_____]**$_{\text{BASE:(X>)V}}$ construction with a simple aspectual meaning. It hardly ever collocates with the respective preposition *za* and, when tested experimentally, would most probably evoke variants like *za-konchitj*, *za-vershitj*, or *za-kljuchitj*, all with the same meaning 'bring something to an end'. Unfortunately, with the methodological toolbox of this study, there is no simple way of disambiguating such instances. However, I do not consider this issue to detract from the main conclusions: all Russian prepositional prefixes have developed clear aspectual, meaning-devoid uses, and so any errors in my corpus scores should be distributed across prefixes in an unbiased manner.

Second, it is unclear how the proposed distinction between two constructions can account for Russian prefixed verbs whose prefixes do not have prepositional counterparts. The etymology of these prefixes is not always self-evident, and the patterns of their co-occurrences with prepositions are not always consistent. Some may encode spatial meanings corresponding to the meanings of prepositions that do not coincide with these prefixes in form (*vy-vesti iz doma* 'to lead out of the house'). Other prefixes cannot encode spatial meanings at all (*raz-bitj na chasti* 'break to pieces').

Third, the abundance of English and German examples provided in the introduction and elsewhere in this chapter should not mislead the reader. Right now, I do not know whether the framework proposed in the study readily extends to German and English data. As already stated, I expect that similar results would be observed in these languages, but this demands further investigation. Thus, the discussion of patterns in languages other than Russian is meant only as a direction for future work.

9 Conclusion

This thesis comprises three big parts, each dedicated to the discussion of one aspect of multi-morphemic words' storing and processing. In the first part, which includes Chapter 2 to 4, I tested whether a construction morphology approach to Russian prefixed verbs is justified, that is, whether they can be analysed as prefix-base constructions. Prefix-base constructions, like any other constructions, must be stored and processed in a network of associations, and access to them must be determined by the activation level of a construction at a particular moment in time. One easy method for activating a construction is through the structural priming of it by means of the same or a similar element preceding it in the discourse. With this in mind, I designed and conducted an experiment, the description of which constitutes the bulk of Chapter 2.

My data included all 28 Russian verbal prefixes, of which 17 have prepositional counterparts in modern Russian and 11 do not. For the experiment, all meanings of all prefixes listed by the Russian Grammar were taken into consideration (91 meanings for prepositional prefixes and 34 meanings for unprepositional prefixes, 125 in total). For each meaning, one sentence containing a respective verb was obtained from the Russian National Corpus, all sentences being approximately of the same length. In each of these sentences, the root of the target prefixed verb was substituted with the nonce root *-banksi-*. Next, two experimental conditions were designed. In the first condition, each of the 125 target sentences was preceded by another sentence obtained from the Russian National Corpus in which the same prefix of the same meaning was used with a different verbal base. In the second condition, the preceding sentences were chosen so that they contained verbs that had different prefixes, or no prefixes at all, but were contextually synonymous to the coded target verb.

The participants of the experiment were asked to evaluate the semantic transparency of a prefixed nonce verb given in minimal context as well as to semanticise it by suggesting an existing Russian verb with the same prefix. Specifically, they were asked to do two things: (1) rate on a scale of 1 to 4, how intuitively well they understood the meaning of the nonce word and (2) substitute the nonce word, as they understood it, with any existing Russian verb, replacing the nonce base *-banksi-* and preserving all other elements (beginning and end) of the verb. From these two types of participants' answers, two scores were calculated: (1) clearness score and (2) correctness score. The correctness score was designed so that it most closely matched the scale of the clearness score. Each submission was ranked on a scale from 1 to 4 where 1 stood for a no answer or an incorrect answer (the prefix was changed in the result of the substitution) and 4 stood for a correctly identified verb.

Having analysed the clearness and correctness scores, I found that the participants of the experiment were very sensitive to the etymological connection between verb prefixes and the prepositions to which they are related. The following hierarchy of Russian prefix-base constructions' decomposability was observed: borrowed prefixes and native prefixes unrelated to prepositions → native prefixes related to prepositions but with non-prepositional meaning → native prefixes related to prepositions and with prepositional meaning. The closer a prefix is to the left extremity of the scale, the higher the chances that the respective prefix-base construction is accessed via lexical link, that is, directly as one chunk. Conversely, the closer a prefix is to the right extremity of the scale, the higher the chances that the respective prefix-base construction is accessed via a sequential link between its morphological parts.

Regardless of experimental condition, prepositional prefixes were distinguished from unprepositional ones. They were rated by the participants as significantly more semantically transparent than their counterparts and also produced a greater number of the correct substitutions of coded words. However, the priming mechanism worked very differently with these two types of constructions. The interpretation of the nonce verbs with prepositional prefixes was significantly facilitated by lexical boost (second experimental condition), while the interpretation of the nonce verbs with unprepositional prefixes was mostly affected by structural priming (first experimental condition).

I think that these priming effects can be explained as follows. With removing prefixed verbs' actual bases and implanting the same nonce base into them, I effectively blocked for these words the direct-access, non-decomposed route. In agreement with my hypothesis, this operation had more dire consequences for verbs with unprepositional prefixes because it turned them into charades that had to be guessed from the context. It is, then, of little surprise that lexical boost in this situation could not provide the participants of the experiment with sufficient information: they must have experienced troubles even with matching priming verb to the target verb. On the other hand, structural priming of the verbs with unprepositional prefixes helped to constructionalise them, opening the route of morphological decomposition and providing participants with a hint at an interpretation. Conversely, the verbs with prepositional prefixes did not really require any structural prop because their prefixes, which coincide in form with very frequent prepositions, are easily detachable from the bases on their own. Lexical boost, on the other hand, helped the participants to strengthen the link between general constructional and specific lexical meaning of respective verbs, thus limiting the space of possible interpretations.

In general, the findings reported in Chapter 2 speak strongly in favour of the idea that morphological structure is gradient and shaped by language use and that morphological decomposition is a matter of degree.

In Chapters 3 and 4, I analysed how morphological decomposition manifests itself in the duration of the inter-morpheme period of silence in Russian prefixed verbs. I started the study reported in Chapter 3 with a hypothesis that in Russian pronunciation, there tends to be a longer silent period between prepositional prefixes and bases (morphologically transparent junctures) than between unprepositional prefixes and bases (morphologically opaque junctures). The reason for that is shown to be grounded in the compositional nature of the former and non-compositional nature of the latter.

My data, again, included all 28 Russian verbal prefixes, 17 prepositional and 11 unprepositional ones. For each of 125 meanings that can be encoded by respective prefix-base constructions, one verb was randomly chosen from the list of paradigm examples with which this meaning is illustrated in Russian Grammar and one sentence with this verb used in this meaning was randomly obtained from the Russian National Corpus. Each sentence was read aloud and recorded by 30 different native speakers. The acoustic waveforms of the target verbs were hand-segmented in Praat. Visually identifiable periods of silence at the boundaries between verbal prefixes and bases were manually coded by two annotators. For each prefixed verb, two values were extracted: A) the total duration of the pronunciation of the given verb in milliseconds and B) the duration of the silent period between the verbal prefix and the base in milliseconds. As a measure of interest, the simple ratio of B to A was calculated in order to control for varying speech rates.

I found that the null hypothesis of no significant difference in the duration of silent periods between verbal prefixes and bases across two types of prefixes can be safely rejected. The prefixes related to prepositions produced, on average, significantly greater ratios than the prefixes unrelated to prepositions. I, however, could not help noticing that the actual variability of the lengths of silent periods resists being reduced to just two homogeneous groups of values. In order to account for this variability, I suggested taking into account that the 'prepositonality' of a prefix-base construction is a matter of degree and that this continuum can be subdivided two-dimensionally: first, along the axis of semantic category, and second, along the axis of compositional type. Thus, in lieu of the original dichotomous division, each of the 125 prefix-base constructions in my data was assigned to a specific slot in a 3-by-3 matrix matching one of the possible semantic categories (literal, metaphorical, and conventional) with one of the possible compositional types (prepositional, non-prepositional, and unprepositional).

I observed that with different compositional types of prefix-base constructions, different meaning categories are characterized by longest silent periods: (1) prepositional constructions have corresponding prepositions for their literal meanings and so tend to employ a greater pause for flagging construction-specific, metaphorical and conventional meanings, (2) unprepositional constructions do not relate to any prepositions at all and thus represent the mirror image of prepositional constructions, flagging most basic literal meanings, (3) nonprepositional constructions fall somewhere in between. On the one hand, they have their corresponding prepositions and so align with prepositional constructions in that prefixes with construction-specific meanings tend to be flagged. On the other hand, their literal meanings do not coincide with the meanings encoded by respective prepositions, and so these constructions align with unprepositional constructions in that the prefix tends to be further detached from the base in pronunciation to accentuate this prefix's preposition-like behaviour. Notably, when confronted with a prefix-base construction that has multiple meanings for a single category/type slot, the participants of my experiment tried to disambiguate those meanings by keeping one silent period as a reference level and shifting another to a lower (if possible) or higher level, depending on the preferences of a particular category/type slot.

In order to find out whether the observed differences pertain to the participants' familiarity with target verbs (their parsability, language frequency, etc.) or with respective prefix-base constructional schemas, in Chapter 4, I conducted a follow-up study designed to verify the results of the previous experiment. I was interested in finding out whether the observed differences would hold if one (1) replaces real bases in target verbs with a nonce base while retaining the prefixes, (2) measures absolute rather than relative length of silent periods, and (3) controls for all phonetic differences in target verbs and considers only the variability that is left unexplained by these factors. To ensure comparability of two experimental settings, the follow-up study retained much of the previous experiment's design. Same 28 Russian verbal prefixes and same 125 construction-meaning pairings came under investigation. Comparable number of participants was employed: each sentence was pronounced, on average, by 29 different native speakers.

In some important aspects, however, experimental design of the follow-up study differed from that of the initial experiment. While previously I illustrated each construction meaning with a sentence containing respective verb from the Russian National Corpus, now I used self-invented formulaic sentences: minimal context, necessary to correctly semanticise the verb (no longer than seven words), SVO order, ditransitive (where possible), same subject (proper noun), all verbs used in past tense. In each verb, an actual base was replaced with a nonce base -*banksi*-. Having obtained the results, I measured the lengths of the the periods of silence at the boundary between prefix and nonce base and fitted a linear regression model to the data to account for the part of variability that was induced by phonetic factors only. From the linear regression output, I obtained the residuals that quantified the variability which was not explained away by the phonetic differences in target verbs. These residuals could be viewed as manifestations of the fact that respective observations came from different populations of prefix-base constructions, and so all analyses to follow were performed on them.

I used Markov chain Monte Carlo methods to build a hierarchical model of the data and create a posterior distribution of the parameters of interest. Specifically, I was interested in estimating quantity $P(\text{Residual} > 0 \mid f_m, f_c, f_b)$, where $\{f_m, f_c, f_b\}$ is the set of all possible values of the variables *meaning category*, *construction type*, and *base type*. This conditional probability is interesting because observing a positive residual means that the length of respective period of silence between prefix and base is greater than what can be explained away by purely phonetic factors.

The main obtained results were as follows. First, I found that three different construction types formed a continuum: it was more likely to observe a positive residual with prepositional prefixes, less likely with non-prepositional prefixes, and even more unlikely with unprepositional prefixes. Second, I found that prepositional and non-prepositional constructions aligned with each other and contrasted with unprepositional constructions with regard to their positive residuals' probabilities for different meaning categories. The former trend may be described as follows: literal < metaphorical < conventional, the latter one as follows: literal > metaphorical > conventional. On the other hand, there was a clear difference between prepositional and non-prepositional constructions themselves. Positive residuals were more likely with literal constructions of non-prepositional type than with literal constructions of prepositional type. Conversely, positive residuals were more likely with metaphorical and conventional constructions of prepositional type than with metaphorical and conventional constructions of non-prepositional type. Thus, all the important results from the study reported in Chapter 3 were confirmed in the follow-up study.

In the second part of the thesis, which encompasses Chapter 5 to 6, I addressed the problem of evaluating the morphological productivity of the prefixes in Russian and English. It is a long-established view that high token frequency represents a sort of stumbling block for affixes' linguistic productivity. It has been argued that affixes encountered in many frequent items become less parsable and, by that, lose their ability to combine with new bases. However, based on my findings, the picture appears to be more complicated: high-frequency derivations with an affix, once they are accumulated in a certain amount of types, do not block the emergence of new low-frequency coinages but rather facilitate them, serving as pathbreakers for neologisms.

In Chapter 5, I tried to show that the unexpected relationship between affixes' frequency and productivity that has been observed in the literature is, most likely, an artefact of the established way of measuring linguistic productivity as the ratio of the number of hapaxes with a certain affix to the number of all tokens with this affix attested in a language corpus. Very simply, if you have an equation of the form $productivity_i = c / T_i$, where c as the number of hapaxes approaches a constant for all affixes and T_i is the total number of derivations with an affix i, it is hard to expect anything other than a negative correlation between $productivity_i$ and T_i. To provide a way out of this circular reasoning, I suggested that linguistic productivity should be viewed as the probability of an affix to combine with a random base. Using the internet corpus of English from 2018 (*ententen18_tt31*), I evaluated the linguistic productivity of 27 Russian prefixes (*dis*-, *niz*-, *pred*-, *vs*-, *de*-, *nad*-, *voz*-, *iz*-, *re*-, *so*-, *ras*-, *pre*-, *ob*-, *pod*-, *pri*-, *pere*-, *u*-, *v*-, *do*-, *vy*-, *na*-, *pro*-, *s*-, *o*-, *ot*-, *po*-, and *za*-; I decided to leave out *nedo*- as the 28th prefix due to the fact that from a computational perspective, it is not always possible to reliably distinguish it from the combination of the negative particle *ne* and the prefix *do*-) and 25 English prefixes (*anti*-, *con*-, *counter*-, *cross*-, *de*-, *dis*-, *em*-, *en*-, *fore*-, *im*-, *in*-, *inter*-, *mid*-, *mis*-, *non*-, *out*-, *over*-, *pre*-, *re*-, *self*-, *sub*-, *super*-, *trans*-, *un*-, and *under*-). For each prefix, three probabilities were obtained: (1) P(X = 0), the probability of no occurrence of the combination of this prefix with a random base in the corpus; (2) P(X = 1), the probability that the combination of this prefix with a random base will be of low frequency; and (3) P(X = 2), the probability that the combination of this prefix with a random base will be frequent.

The true measure of linguistic productivity was estimated in two steps. First, the initial and transition probability distributions for a two-time-slice dynamic Bayesian network were learned on a random sample of 100 random bases obtained from the corpus. Second, the value of P(X = 1) + P(X = 2) was calculated for the 101st random base, given the last base in the sample. I found that, based on the evaluations of these probabilities, all prefixes, when arranged in order of ascending productivity, could be subdivided into three groups. The first group encompasses prefixes with the probabilities hierarchically arranged as P(X = 0) > P(X = 2) > P(X = 1). In the second group, one finds prefixes where the probabilities are aligned in this way: P(X = 2) > P(X = 0) > P(X = 1). Finally, the prefixes that belong to the last group reveal the following pattern: P(X = 2) > P(X = 1) > P(X = 0).

Interestingly, these categorical differences in both languages were found to emerge as manifestations of an inherently gradient structure. Thus, within the first group, the differences between probabilities $P(X = 0)$ and $P(X = 2)$ continuously decrease, while the differences between probabilities $P(X = 2)$ and $P(X = 1)$ continuously increase. Within the second group, a similar mechanism of change can be observed, though with different contrasts. The differences between probabilities $P(X = 2)$ and $P(X = 0)$ become larger, while the differences between probabilities $P(X = 0)$ and $P(X = 1)$ become smaller. Finally, within the third group, the gap between probabilities $P(X = 2)$ and $P(X = 1)$ successively narrows, while the gap between probabilities $P(X = 1)$ and $P(X = 0)$ widens.

All of the above raises an interesting question of how derivational patterns spread. It is well known that morphological rules may be unproductive. No less remarkable, however, is the fact that the productivity of even fully productive Russian or English affixes is not without its limits. Thus, it is not clear why, for example, given the high frequency of the verb *give* and the high productivity of the prefix *re-*, the derivation *re-give* is extremely unpopular, with 0 hits per million tokens in both COCA and *ententen18_tt31*. It is also unclear why, given that the verb *evolve* is more frequent than the verb *regulate*, only *dis-regulate* is actually attested in COCA and *ententen18_tt31*, although there seems to be nothing conceptually improbable or semantically incompatible in the possible combination *dis-evolve*. One might conclude that, even for very productive affixes, there is no simple linear relation between base and derivation frequency. Rather, it is high-frequency items with a certain affix that play a pivotal role in the self-propagating of respective derivational patterns and the structuring of its output, with less-frequent members being grouped around more prominent ones.

Chapter 6 is dedicated to testing this ‘clustering hypothesis’ on English data. The probabilistic estimation of the linguistic productivity of 25 English prefixes presented in the previous chapter seemed to confirm the hypothesis that token frequency as such, contrary to common beliefs, cannot be considered a stumbling block for derivational patterns. The observed dependence of the emergence of low-frequency derivations on the existence of numerous high-frequency derivations with the same affix requires clarification. I believe that this phenomenon can be explained as follows. The greater the number of frequently used words with a certain affix, the higher the chances that some of them will collocate with their own bases. The more persistent these co-occurrences are, the more likely it is that the respective affix will become recognisable, parsable, and applicable, that is, productive. If so, then every instance of such a discourse-conditioned pattern’s invigoration is a short-term memory process, and the range of applicability of the temporarily refreshed pattern should be limited to the nearest context. Hence, one would expect to see many low-frequency coinages clustered around those high-frequency anchors with which they typically collocate. In Chapter 6, I provided simulation-based evidence supporting this claim.

The data for the study were collected as follows. First, 995 random content words (nouns, verbs, and adjectives) without prefixes were sampled from *ententen18_tt31*, their frequency ranging from 48,421,599 to 54 tokens. Each of 25 English prefixes on my list was coupled with each of those 995 bases, so that the bases remained the same for all prefixes. The raw frequencies of all constructed derivations were then queried in the same corpus. To test my hypothesis about how the process of derivational patterns' spreading may work, I programmed the following computer simulation. First, I created an undirected network $G_0 = (V_0, E_0)$, where V_0 was a set of 995 vertices representing all the bases in my data, and E_0 was a set of unordered pairs of these vertices, such that for any pair of bases v_i and v_j, $i \neq j$, the edge (v_i, v_j) was added to E_0 only if the measure of cosine similarity between v_i and v_j, as calculated by the GloVe model, was found to be greater than 0. This resulted in a very dense (0.80) network with 995 nodes and 396,760 edges. Next, for each prefix p_i, $i \in \{1, \ldots, 25\}$, I created a network $G_i = (V_0+V_i, E_0+E_i)$, where V_i was a set of the p_i-derivations belonging to the high-frequency group (those with more than 1,000 corpus hits in my data), and E_i was a set of the edges connecting each derivation in V_i to its base in V_0 if the measure of their cosine similarity was greater than 0. All the edges in the network G_i were weighted by the cosine similarity values of their extremities.

After that, the following simulation process was run for each prefix. One node was randomly chosen from the set of nodes in V_i that were considered to be chain initialisers. The probability of each derivation's selection was equal to its relative frequency in the group of high-frequency derivations with the respective prefix, so more frequent words stood a better chance of being drawn. The possibility of a transition from a selected vertex to one of its nearest neighbours in the network was evaluated by taking all the weights of the edges incident upon the vertex, renormalising them so that they sum up to one, and randomly choosing a candidate from the resulting distribution. It is clear from the structure of the network G_i that if a p_i-derivation had an edge with its base, the transition from the derivation to the base was made with a probability of 1.0. To put it differently, each randomly selected vertex and all its neighbours constituted a Markov chain with transitional probabilities approximated by normalising the words' cosine similarity values so that they added up to 1 in each row of the transition matrix.

The process described above was repeated for each consecutively chosen vertex. Each chain initialised by the randomly selected derivation was limited to 50 transitions; after this, a new derivation was drawn. Overall, I sampled 500 chains for each prefix in my data. The most important concept for me was that of a 'pattern memory score'. I assumed, in line with my hypothesis, that whenever a derivation and its base co-occur in discourse, the respective derivational pattern is refreshed and remains available for application for some time. However, this is a 'memory-loss' process in the sense that if the invigorated pattern remains unemployed long enough, it is deleted from the operative memory and needs to be retrieved once again. Hence, whenever a transition from a derivation to its base or vice-versa was recorded during the simulation, I increased the pattern memory score, which was initially set to zero, by five points. This meant that each base reached over the course of five transitions from this moment received +1 to its tally of simulated derivations. The bases that were visited when the pattern memory score equalled zero were passed over with no increase in the number of respective p_i-derivations.

Having analysed the results, I found for all prefixes a significant positive correlation between the actual and simulated frequencies of derivations, with the coefficients ranging from 0.30 to 0.76. In addition, three frequency groups of derivations (high-frequency, low-frequency, and unattested words) were characterised by significant and expected shifts in the locations of their simulated frequency counts. Specifically, the vertices corresponding to the unattested derivations in my data were visited significantly less frequently during the simulation than the vertices corresponding to the low-frequency derivations (those with less than 1,000 corpus hits).

The findings presented and discussed in Chapters 5 and 6 show that the linguistic productivity of a derivational pattern is dependent not so much on the proportion of infrequent (parsable) words among all words with a specific affix as on the proportion of high-frequency items that strongly collocate with their bases. For example, the levels of productivity of the prefixes *de-* and *con-* are apparently different. I assessed their probabilities of combining with a random base as 0.42 and 0.27, respectively. This difference is difficult to explain by taking into account only frequency counts or average levels of parsability of the members of the respective groups. In my random samples, derivations with *de-* include an even greater proportion of frequent words than derivations with *con-*. On the other hand, the proportion of complex words that are likely to be parsed, as suggested by their derivation to base frequency ratio, among all actually attested items with these prefixes is exactly the same. What really distinguishes my samples of derivations with *de-* and *con-*, apart from the quantity of attested complex words (450 vs. 277), is the fact that derivations with *de-*, frequently co-occurring in discourse with their bases, are more numerous, and the average collocational strength in those pairs is greater.

In the third part of the thesis (Chapters 7 and 8), I addressed the problem of the morphological analysability and semantic transparency of multi-morphemic words. In Chapter 7, I compared the dual-route model of perception of complex words, which was foundational for Hay's work on lexical frequency in morphology, with an alternative interpretation proposed within the framework of construction morphology, where complex words are seen as constructions on the word level. The main difference between the two approaches, as I see it, is in the allowance for one additional meaning processing mechanism, which construction morphology can make due to its ability to distinguish between fixed elements and slots (variables). I argued that for a two-element complex expression — for example, a prefix or particle verb — one can have four possible combinations: (1) both elements are fixed, (2) both elements are variables, (3) the first element is a variable and the second element is fixed, and (4) the first element is fixed and the second element is a variable. Linguistic items of type (1) are non-analysable, non-compositional, and non-productive. They are listed diachronic relics that are not assembled on the fly but are retrieved from the lexicon. Linguistic items of type (2) are, in contrast, analysable, fully compositional, and productive. Up to this point, there is really no divergence between the dual-route model and construction morphology accounts. However, with types (3) and (4), the situation is more interesting. Linguistic items of types (3) and (4) are analysable and (semi?)productive and yet, with regard to their semantics, cannot be called either compositional or non-compositional. They cannot be called compositional in the traditional sense since their general meaning cannot be inferred from the meaning of their components. Yet it feels somewhat awkward to call them non-compositional because, often, their fixed elements make the same semantic contribution in multiple words. Thus, it makes more sense to call complex linguistic expressions of type (2) compositional and complex linguistic expressions of types (3) and (4) parsable, putting a strong emphasis on the fact that all of them are analysable as opposed to expressions of type (1).

I contended that the relative frequency account which evaluates the degree of parsability of a complex word by calculating its derivation to base frequency ratio can only distinguish between non-analysable and compositional expressions. However, it falls short of registering the presence of two other construction types that are comprised of a fixed element and a slot, and instead lumps them together with either compositional or non-analysable constructions. In a sense, the very design of the constructions of types (3) and (4) predetermines the relative frequency relation between the whole form and the base. Since one fixed element normally appears in many words, combined with different elements that fill the respective construction's empty slot, it is expected that in complex words of type (3), where the base is fixed, the derivation to base frequency ratio will tend to be less than one. In contrast, complex words of type (4), where the base serves as a filler, will most likely reveal derivation to base frequency ratios greater than one.

In order to overcome the conflation problem I suggested thinking about complex words' analysability patterns in terms of transitional probabilities, both forward- and backward-going. Thus, for a specific complex word, one would ask, how likely it is that this particular base would be combined with this affix, and how likely it is that this particular affix would be combined with this base? It logically follows that expressions of type (1) will be characterised by comparably high probabilities of transition from affix to base and from base to affix and expressions of type (2) will be characterised by comparably low probabilities of transition in both directions. For expressions of types (3) and (4), these probabilities will diverge. In type (3), where the first element is a variable and the second element is fixed, the probability of transition from base to affix will be low while the probability of transition from affix to base will be high. Conversely, in type (4), where the first element is fixed and the second element is a variable, the probability of transition from base to affix will be high while the probability of transition from affix to base will be low. For the purposes of my studies, I calculated transitional probabilities as relative frequencies, by taking all affixed words in a morphemic dictionary of the respective language and looking up frequencies of interest in the internet corpus of this language. Then, for any word, its P (affix | base) was equal to the number of word's tokens divided be the number of tokens of all words with this base and its P (base | affix) was equal to the number of word's tokens divided by the number of tokens of all words with this affix.

The main bulk of Chapter 7 was dedicated to probing into the cognitive reality of the four conjectured construction types of English and Russian prefixed words. The chapter reports the results of three studies. In the first study, I provided experimental evidence that language users perceived complex words of types (3) and (4) differently than complex words of types (2) and (1) with regard to their morphological analysability and semantic transparency. For both English and Russian, I selected 40 stimuli: eight prefixes of different linguistic productivity and five construction types (those described above plus one pseudo-affixed word) with each prefix. Words were matched for the number of morphemes, and every effort was made to match them for junctural phonotactics, stress patterns, syllable counts, and the frequency of the derived form as well. The experimental designs for both languages were identical. For English subjects, I repeated the instructions verbatim as they were given in Hay (2001), and for Russian participants, I simply translated them into Russian, having only changed the language examples. Both experiments were completed online. Each participant was presented with just one pair of words sharing the same prefix (or pseudo-prefix coinciding with it in form) and asked to type in the word they thought was more complex.

Having analysed the results, I obtained for each construction type its probability of success, that is, the probability of any of its instantiations being judged more complex when paired with an instantiation of a different construction type. With English stimuli, the ranking of the obtained probabilities of success was in agreement with my hypothesis that the degree of the construction's perceived complexity would be proportional to the number of empty slots within it. The observed alignment of the construction types — (1) < (4) / (3) < (2) — is hard to reconcile with the relative frequency account because if one takes into account only the derivation to base frequency ratio, then one would expect to find types (3) and (4) much closer to types (2) and (1), respectively, than to each other. The Russian part of the experiment produced a different hierarchy of construction types that is even more incompatible with the relative frequency view: (3) / (1) < (2) < (4). Here, constructions with an empty slot for a base were consistently rated as more complex than their compositional counterparts. On the other hand, constructions with an empty slot for an affix were merged with non-analysable items.

I explained the puzzling difference between English and Russian results by the following observation. Russian construction types (3) and (4), unlike their English counterparts, came to be semantically specialised. Prefixes in the Russian verbs of type (3) mostly encode spatial meanings inherited from prepositions, while the same prefixes in Russian verbs of type (4) tend to have non-spatial, construction-specific meanings. From this, it necessarily follows that the fixed elements of the Russian constructions of type (3) (bases) depart from their free counterparts in semantics and distribution to a much lesser extent than the fixed elements of the constructions of type (4) (prefixes). It makes a lot of intuitive sense that the closer the meaning of a complex linguistic item is to the meaning of one of its components, the harder it will be for the speakers to semanticise the remaining element, which is a prerequisite for judging the item as complex. The results of my Russian and English experiments confirm this view. Russian speakers, for example, should have considered the type (3) word *na-zhatj* 'press on' as less complex than the type (4) word *na-vreditj* 'do a lot of harm' because in the former case, the general meaning of the derivation is very much explained away by the meaning of its nested base *zhatj* 'press'. In the latter case, however, the contribution of the fixed element *na-* 'accumulate or produce in great amounts' to the meaning of its host is only of a framework nature. It is very unlikely that the English participants were confronted with the same complications.

In the second study, I addressed the question of the relationship between two ways of measuring complex words' degrees of analysability: by calculating their derivation to base frequency ratios and by calculating the log ratios of their elements' transitional probabilities. By means of probabilistic modelling and partial replication of Hay's original experiment (2001) I showed how the former method might lead to the conflation of different construction types. I built a probabilistic model that would, drawing on the evidence obtained during my English experiment, predict a most likely winner in the complexity assessment contest for each of the 17 pairs of words in Hay's data. After the model had been trained, I used the learnt probabilities to predict the outcome of a hypothetical experiment where 24 participants would be asked to select a more complex word in each of the 17 pairs under investigation. In order to check the adequacy of the model's predictions, I tested the same 17 pairs of words in a real experiment with the setting identical to the one of my above-described English study.

The correlation between the predicted and observed proportions of success was found to be significant ($r = 0.52$, $p = 0.03$). Most importantly, in both hypothetical and real experiments, 55% of responses judged the words from Hay's group A to be more complex than their counterparts, and only 45% of responses chose the words from Hay's group B. Thus, my results were the opposite of what Hay reported: words in group A, though more frequent than the bases they contain, were rated more complex than words in group B, which are less frequent than the bases they contain. Though Hay's and my experiments are not directly comparable, I take the discrepancy in our findings as a useful illustration of how relying solely on relative frequency calculations can lead to the conflation of different construction types. If one calculates transitional probabilities' log ratios for Hay's experimental stimuli in the same way as I did before and, depending on these ratios, assigns to the words their most likely construction types, one will discover that most of the words in group A are of type (4) rather than type (1). On the other hand, group B comprises complex words of types (3) and (2) rather than just (2). Given that some non-analysable words are found in group B as well and that, as my experiment has shown, there seems to be no significant difference in the perceived complexity of English HL and LH constructions of types (3) and (4), it is no wonder that the two groups cannot be unequivocally delineated with regard to their members' morphological complexity.

Finally, in the third study, I showed that the relationship between analysability and productivity is not linear, as it has been frequently described. In fact, the preponderance of words of construction types (3) and (4) among the derivations with a certain affix might serve as a sign of this affix's constrained productivity. In order to analyse the relation between the analysability of

English and Russian prefixed words of types (3) and (4) and the productivity of their prefixes, I used two measures. The parsability ratio of a prefix was calculated as the proportion of words for which the absolute difference between P (affix | base) and P (base | affix) was greater than 1% (as a threshold value suggested by my experimental stimuli) among all words with this prefix. The English data comprised a total of 25,816 words with the following 24 prefixes: *anti*-, *con*-, *counter*-, *cross*-, *de*-, *dis*-, *em*-, *en*-, *fore*-, *im*-, *in*-, *inter*-, *mid*-, *mis*-, *non*-, *out*-, *over*-, *pre*-, *re*-, *sub*-, *super*-, *trans*-, *un*-, and *under*-. The Russian data comprised 9,018 words with the following 27 prefixes: *de*-, *diz*-, *do*-, *iz*-, *na*-, *nad*-, *niz*-, *ob*-, *pere*-, *pre*-, *pro*-, *po*-, *pod*-, *pred*-, *pri*-, *raz*-, *re*-, *s*-, *so*-, *o*-, *ot*-, *u*-, *v*-, *voz*-, *vz*-, *vy*-, and *za*-. As for the linguistic productivity of a prefix, I did not want to use Baayen's hapax-based measure since, as it has been pointed out in the literature, this measure is ill-suited for the comparison of affixes with different token numbers. Instead, I assessed the morphological productivity of the prefixes in my data as their probability to combine with a random base (see Chapter 5). Besides these two values, for the purposes of regression modelling, I factored in the frequencies of prepositions and particles that coincide in form with respective prefixes.

My findings for both languages revealed a very special relationship between multi-morphemic words' parsability ratio, calculated as I proposed, and the linguistic productivity of their prefixes. Specifically, for the prefixes which have no free counterparts or correspond to relatively low-frequency prepositions/particles, the lower the parsability ratio, the greater the linguistic productivity. In contrast, for the prefixes that have high-frequency free counterparts, the higher the parsability ratio, the greater the linguistic productivity. I concluded that if, among multi-morphemic words with a certain prefix, there are many words whose bases are conditionally dependent upon the prefix — that is, there is a strong sequential link between the elements — the prefix's range of applicability is limited, and the constructional meaning is not general enough to accommodate a wide variety of items in its slot. This relationship may, however, be reversed: if for some prefix there exists in language a corresponding free element that is sufficiently frequent, it can lead to higher productivity even of those prefixes with high parsability ratios.

The results reported in Chapter 7 are important insofar as they allow a distinction to be drawn between two different models of the processing of meaning of analysable complex words. Construction type (2), on the one hand, implies that each of the elements entering into combination is equally free to vary; the combination itself is judged by language users to be semantically complex but transparent. Hence, this model can be called compositional. Construction types (3) and (4), on the other hand, assign some very general sense to the construction as such. Multi-morphemic words of these types are similar to collocations in the sense that they also consist of a node (conditionally independent element) and a collocate (conditionally dependent element). Such combinations of linguistic items are also considered semantically complex but less transparent because a collocate's meaning does not generally coincide with the meaning of a respective free element (even if it exists) and must be parsed out from what is available. Hence, I suggested to name this model parsable. Clearly, the distinction between the two models of meaning processing is not a clear-cut categorical one but rather a probabilistic continuum. One can predict which model — compositional or parsable — is more likely to be chosen for each word by taking into account the word's two morphological families: one for the affix, another for the base. The words that are characterised by a greater discrepancy between transitional probabilities from affix to base and from base to affix are more likely to be treated as parsable than those with more or less comparable (low) transitional probabilities.

Chapter 8 is dedicated to showing how two parsable construction types can become semantically specialised. Both experimental and corpus data presented in this chapter suggest that there exist at least two different constructions for Russian complex verbs: **[_____]**$_{\textbf{PREFIX}}$ **+ BASE** for verbs with spatial meanings and **PREFIX + [_____]**$_{\textbf{BASE:(X>)V}}$ for verbs with idiosyncratic meanings. The chapter reports the results of two studies. In the first study, I provided experimental evidence that native speakers, when asked to manipulate complex verbs by changing either their prefix or their base, reveal a significant preference for changing the prefixes of spatial verbs and the bases of non-spatial verbs. In the second study, I showed that the choice of construction could be predicted by taking into account the discrepancy in probabilities of transition from base to affix and from affix to base. If P (prefix | base) ≤ P (base | prefix), the first construction with an empty slot for the prefix is likely to be chosen. If, on the other hand, P (prefix | base) > P (base | prefix), the second construction with an empty slot for the base becomes more likely. In both cases, the element of the construction that tells us less about its counterpart activates general constructional meaning, while the element that has greater predictive power serves as a filler for the construction's empty slot. I also showed that the distinction between two constructions does not necessarily pertain to the difference in relative frequencies. Token frequencies of bases and lemmas seem to play hardly any role in either of these constructions. Even low-frequency bases may combine with many different prefixes, and low-frequency lemmas adhere to the same constructional patterns as their high-frequency counterparts.

One non-trivial contribution to the construction morphology framework made by the studies reported in Chapter 8 is the idea that some constructions might arise by way of generalisation over others, which leads to a shift in the positioning of a fixed element and a slot. I hypothesised that at the first stage of development, different preverbs/prefixes/particles with spatial meanings are combined with verbs so that they satisfy these verbs' argument structures, thus giving rise to complex verbs whose meaning is the sum of the meanings of their parts. Since one verb typically combines with many preverbs/prefixes/particles to encode different spatial meanings, such instances become generalised as constructions of the form **[_____]**$_{\mathbf{PREFIX}}$ **+ BASE** with one empty slot and one fixed element. Next, presumably after the number of unique bases associated with this particular preverb/prefix/particle reaches a certain threshold, a new construction of the form **PREFIX + [_____]**$_{\mathbf{BASE:(X>)V}}$ comes into existence by means of abstraction and categorisation. This new construction then licenses certain bases to fill its empty slot, thus serving as a template with an off-the-shelf general (non-spatial) meaning for which the inserted lexical material provides a necessary specification. Importantly, some of these constructions may license the insertion of bases that have already been combined with the same preverb/prefix/particle in its spatial meaning, thus resulting in polysemous complex verbs.

References

Aminova, A. A. (1988). *Glagol'nye sinonimy russkogo jazyka. Razvitie glagol'noj sinonimii v slovo-obrazovatel'nom aspekte* [Verb Synonyms of Russian: Development of Synonyms in Derivational Aspect]. Kazan.

Amiot, D. (2004). Between compounding and derivation: Elements of word-formation corresponding to prepositions. *Morphology and its demarcations, Feb 2004*. Vienna, pp. 183–213.

Arkadiev, P. M., & Shluinsky, A. B. (2015). Verb-Classifying Aspectual Sybases: Towards a Typology. *Vestnik SPbGU 3*: 4–24.

Arcodia, G. F. & Basciano, B. (2018). The Construction Morphology Analysis of Chinese Word Formation. In G. Booij (Ed), *The Construction of Words. Studies in Morphology*. Vol. 4. Springer, Cham. https://doi.org/10.1007/978-3-319-74394-3_9.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge: MIT Press.

Audring, J. (2022). Advances in Morphological Theory: Construction Morphology and Relational Morphology. *Annual Review of Linguistics 8(1)*: 39-58.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47: 31–56.

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2015). Comprehension without segmentation: A proof of concept with naive discrimination learning. *Language, Cognition, and Neuroscience 31(1)*: 106–128.

Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review 118*: 438–481.

Baayen, H. (2009). Corpus linguistics in morphology: Morphological productivity. In A. Lüdeling, & M. Kyto (Eds.). *Corpus linguistics. An international handbook*. Berlin: Mouton De Gruyter, pp. 900–919.

Baayen, H., & Schreuder, R. (2000). Towards a psycholinguistic computational model for morphological parsing. In K. I. B. Spärck Jones, G. J. M. Gazdar, & R. M. Needham (Eds.), Computers, language and speech: formal theories and statistical data [Special issue]. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 358(1769)*: 1281–1293.

Baayen, H., & Renouf, A. (1996). Chronicling *The Times*: Productive lexical innovations in an English newspaper. *Language 72*: 69–96.

Baayen, H. (1994). Productivity in language production. *Language and Cognitive Processes 9*: 447–469.

Baayen, H. (1993). On frequency, transparency and productivity. In G. Booij, & J. van Marle (Eds). *Yearbook of morphology 1992*. Dordrecht: Kluwer, pp. 181–208.

Baayen, H. (1992). Quantitative aspects of morphological productivity. In G. Booij, & J. van Marle (Eds.). *Yearbook of Morphology 1991*. Dordrecht: Kluwer Academic Publishers, pp. 109–150.

Baayen, H., & Lieber, R. (1991). Productivity and English word-formation: a corpus-based study. *Linguistics 29*: 801–843.

Babko-Malaya, O. (1999). *Zero morphology: A study of aspect, argument structure, and case* (Doctoral dissertation, Rutgers, The State University of New Jersey). New Brunswick.

Bagasheva, A. (2015). On verbocentric nominal compounds denoting humans in Bulgarian. *SKASE Journal of Theoretical Linguistics 12(3)*: 74–104.

Bannard, C. (2005). Learning about the meaning of verb–particle constructions from corpora. *Computer Speech and Language 19(4)*: 467–478.

Barðdal, J. (2006). Predicting the productivity of argument structure constructions. *Berkeley Linguistics Society 32*: 467–478.

Basnight-Brown, D., Chen, H., Shu, H., Kostić, A., & Feldman, L. B. (2007). Monolingual and Bilingual Recognition of Regular and Irregular English Verbs: Does Sensitivity to Word Form Vary with Language Experience? *Journal of Memory and Language 57*: 65-80.

Bauer, L. (2001). *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, L. (1983). *English word-formation*. Cambridge: Cambridge University Press.

Beliakov, V. (1999). Krongauz M. A. Pristavki i Glagoly v Russkom Jazyke: Semanticheskaja Grammatika [Krongauz M. A. Prefixes and Verbs in Russian Language: Semantic Grammar], *Revue des études slaves 71(1)*: 212–217.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustic Society of America 113*: 1001–1024.

Ben Hedia, S., & Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics 62: 34–49*. https://doi.org/10.1016/j.wocn.2017.02.002.

Berg, K. (2020). Changes in the productivity of word-formation patterns: Some methodological remarks. *Linguistics 58(4)*:1117–1150. https://doi.org/10.1515/ling-2020-0148.

Berg, K. (2013). Graphemic alternations in English as a reflex of morphological structure. *Morphology 23*: 387–408. https://doi.org/10.1007/s11525-013-9229-1.

Bergsma, S., Bhargava, A., He, H., & Kondrak, G. (2010). Predicting the Semantic Compositionality of Prefix Verbs. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA: MIT, Association for Computational Linguistics, pp. 293–303.

Bird, H., Lambon Ralph, M. A., Seidenberg, M. S., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and past tense morphology: What's the connection? *Journal of Memory and Language 48(3)*: 502–526.

Biskup, P. (2019). *Prepositions, Case and Verbal Prefixes The case of Slavic*. Amsterdam: John Benjamins.

Biskup, P. (2015). On (Non-)Compositionality of Prefixed Verbs. In M. Ziková, P. Caha, & M. Dočekal (Eds.), *Slavic Languages in the Perspective of Formal Grammar: Proceedings of FDSL 10.5, Brno 2014*. Frankfurt am Main: Peter Lang, pp. 59–78.

Biskup, P. (2012). Slavic Prefixes and Adjectival Participles. In M. Ziková, & M. Dočekal (Eds.), *Slavic Languages in Formal Grammar: Proceedings of FDSL 8.5, Brno 2010*. Frankfurt am Main: Peter Lang, pp. 271–289.

Blom, C. (2005). *Complex predicates in Dutch*. Utrecht: LOT.

Bloomfield, L. (1933). *Language*. New York: Holt.

Bochner, H. (1993). *Simplicity in generative morphology*. Berlin: Mouton de Gruyter.

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology 18*: 355–387.

Boersma, P. & Weenink, D. (2020). *Praat: Doing phonetics by computer* [Computer program]. Version 6.1.21. Retrieved from http://www.praat.org/ (April 2020).

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). *Enriching Word Vectors with Subword Information*. arXiv:1607.04606v2. https://doi.org/10.48550/arXiv.1607.04606.

Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics 6*: 213–234.

Bolinger, D. (1948). On defining the morpheme. *Word 4*, 18–23.

Booij, G. (2010a). Construction morphology. *Language and Linguistics Compass 4/7*: 543–555.

Booij, G. (2010b). *Construction Morphology*. Oxford: Oxford University Press.

Booij, G., & Van Kemenade, A. (2003). Preverbs: an introduction. In: Booij, G., Van Marle, J. (eds) *Yearbook of Morphology 2003*. Springer, Dordrecht.

Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension 50(2)*. Retrieved from https://www.joe.org/joe/2012april/tt2.php (26 November 2020).

Bott, S., & Schulte im Walde, S. (2014). Optimizing a distributional semantic model for the prediction of German particle verb compositionality. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 509–516.

Brémaud, P. (1999). *Discrete-Time Markov Models*. Berlin: Springer.

Brems, L. (2003). Measure noun constructions: An instance of semantically-driven grammaticalization. *International Journal of Corpus Linguistics 8(2)*: 283–312**.**

Bundgaard-Nielsen, R. L., & Baker, B. J. (2020). Pause acceptability indicates word-internal structure in Wubuy. *Cognition 198*: 1–10. https://doi.org/10.1016/j.cognition.2019.104167.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed), *Language Production*. London: Academic Press, pp. 257–294.

Bybee, J. (2010). *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.

Bybee, J. (2007). *Frequency of Use and the Organization of Language*. Cambridge: Cambridge University Press.

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

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes 10(5)*: 425–455.

Bybee, J. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds), *Theoretical morphology: approaches in modern linguistics*. Orlando: Academic Press, pp. 119–141.

Bybee, J. (1985). *Morphology. A study of the relation between meaning and form*. Amsterdam, Philadelphia: John Benjamins.

Cappelle, B. (2005) Particle Patterns in English. Ph.D. Dissertation: University of Leuven. https://kulak.kuleuven.be/nl/onderzoek/Onderzoeksdomeinen/language-literature-translation/

letteren/onderzoektaalkunde/fest/descriptiveenglishgrammar/bert-cappelle/bert-cappelle-proefschrift.pdf].

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences 3(3)*: 106–116.

Cetnarowska, B. (2020). Competition between synthetic NN compounds and NN.gen phrasal nouns in Polish: semantic niches, hapax legomena and low-level construction schemas. In L. Körtvélyessy & P. Štekauer (Eds), *Complex Words: Advances in Morphology*. Cambridge: Cambridge University Press, pp. 241–259.

Chukovskij, K. (2001[1928]). *Ot Dvuh do Pjati* [From Two to Five]. Moscow: Terra—Knizhnyj klub.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Studies 22*: 991–1060.

Clopper, C. & Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M Zellers. (Eds), *Rethinking Reduction*. Berlin, Boston: Mouton de Gruyter, pp. 25–72. https://doi.org/10.1515/9783110524178-002.

Cole, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefi⍰xed and suffi⍰xed derived words: A differential frequency effect. *Journal of Memory and Language 28*: 1–13.

Cordeiro, S., Villavicencio, A., Idiart, M., & Ramisch, C. (2019). Unsupervised compositionality prediction of nominal compounds. *Computational Linguistics 45(1)*: 1–57.

Croft, W. (2001). *Radical construction grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press.

Cuzzolin, P. L., Putzu, I., & Ramat, P. (2006). The Indo-European adverb in diachronic and typological perspective. *Indogermanische Forschungen 111*: 1–38.

Dalton-Puffer, C. & Plag, I. (2000). Categorywise, some compound-type morphemes seem to be rather suffix-like: On the status of *-ful*, *-type*, and *-wise* in present day English. *Folia Linguistica 34(3-4)*: 225-244. https://doi.org/10.1515/flin.2000.34.3-4.225.

De Jong, N. H., Schreuder, R., & Baayen, H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes 15*: 329–365.

De Smet, H. (2020). What predicts productivity? Theory meets individuals. *Cognitive Linguistics 31(2)*: 251-278. https://doi.org/10.1515/cog-2019-0026.

Deaver, Guinevere J. (2013). The effects of frequency on dual-route versus single-route processing of morphologically complex terms: A usage-based experiment. *Theses and Dissertations: 4282*.

Dehé, N. et al. (Eds.) (2002.) *Verb-Particle Explorations*. Berlin/New York: Mouton de Gruyter.

Di Sciullo, A.-M., & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.

Diessel, H. (2019). *The grammar network. How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.

Dietz, K. (2015). Historical word-formation in English. In P. O. Müller et al. (Eds), *Word-Formation: An International Handbook of the Languages of Europe*. Volume 3. Berlin: Mouton de Gruyter, pp. 1914–1930.

Divjak, D. (2022). *Frequency in Language. Memory, Attention and Learning*. Cambridge: Cambridge University Press.

Dressler, W. U. (2005). Word-formation in natural morphology. In P. Štekauer, & R. Lieber (Eds.), *Handbook of word-formation*. Berlin: Springer, pp. 267–284.

Drozd, A., Gladkova, A., & Matsuoka, S. (2016). Word embeddings, analogies, and machine learning: Beyond king - man + woman = queen. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. Osaka: The COLING 2016 Organizing Committee, pp. 3519–3530.

Fellbaum, C. (Ed.) (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fernández-Domínguez, J., Díaz-Negrillo, A., & Štekauer, P. (2007). How is Low Morphological Productivity Measured? *Atlantis 29(1)*: 29–54.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: Sage.

Fillmore, Ch. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: the case of 'let alone'. *Language 64(3)*: 501–538.

Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis*. Blackwell, Oxford, pp. 1–32.

Fontanals, J. M. (2001). Preverbs in Complex Denominal Verbs: Lexical Adjuncts or Core Predicates? *Catalan Working Papers in Linguistics 9*: 37–51.

Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood density effects in masked form-priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition 20*: 844–863.

Frauenfelder, U., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. Booij, & J. van Marle (Eds.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer, pp. 165–183.

Fruchterman, T., & Reingold, E. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience 21(11)*: 1129–1164.

Gaeta, L., & Angster, M. (2019). Stripping paradigmatic relations out of the syntax. *Morphology 28*: 249–270.

Gaeta, L., & Ricca, D. (2006). Productivity in Italian word formation: A variable-corpus approach. *Linguistics 44*, 57–89.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research 42(3)*: 237–288.

Goldberg, A. E. (2006). *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, A. E. (1995). *A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gries, S., & Stefanowitsch, A. (2004). Extending collostructional analysis. A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics 9(1)*: 97–129.

Grinstead, C. M., & Snell, J. L. (2012). *Introduction to Probability*. American Mathematical Soc., 2012.

Guillaume, P. (1973). The development of formal elements in the child's speech. In C. A. Ferguson & D. I. Slobin (Eds), *Studies of child language development*. New York: Holt, Rinehart and Winston, pp. 240–251.

Gutierrez, E. D., Shutova, E., Marghetis, T., & Bergen, B. (2016). Literal and metaphorical senses in compositional distributional semantic models. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin, pp. 183–193.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review 111(3)*: 662–720.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.* Berlin, pp. 1489–1501.

Hartmann, S. (2019). Compound worlds and metaphor landscapes: affixoids, allostructions, and higher-order generalizations. *Word Structure 12(3)*: 297–333.

Haspelmath, M. (2023). On what a construction is. *Constructions 15(1)*. https://doi.org/10.24338/cons-539.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics 19*: 1–33.

Haspelmath, M. (2002). *Understanding Morphology*. New York: Oxford University Press.

Hay, J. (2007). The phonetics of *un*. In J. Munat (Ed.), *Studies in Functional and Structural Linguistics: Vol. 58. Lexical creativity, texts and contexts*. John Benjamins, pp. 39–57. https://doi.org/10.1075/sfsl.58.09hay.

Hay, J. (2003). *Causes and consequences of word structure*. London/ New York: Routledge.

Hay, J. (2002). From speech perception to morphology: Affix-ordering revisited. *Language 78(3)*: 527–555.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics 39(6)*: 1041–1070.

Hay, J., & Baayen, H. (2003). Phonotactics, parsing and productivity. *Rivista di Linguistica 15.1*: 99-130.

Hay, J., & Baayen, H. (2002). Parsing and productivity. In G. Booij, & J. van Marle (Eds), *Yearbook of morphology 2001*. Dordrecht: Kluwer, pp. 203–236.

Hay, J., & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory 22*: 565–596.

Hayya, J., Armstrong, D., & Gressis, N. (1975). A Note on the ratio of two normally distributed variables. *Management Science 21(11)*: 1338–1341.

Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.

Hoffmann, T., & Trousdale, G. (2013). *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Hüning, M., & Booij, G. (2014). From compounding to derivation: the emergence of derivational affixes through 'constructionalization'. *Folia Linguistica 48*: 579–604.

Hvitfeldt, E., & Silge, J. (2022). *Supervised Machine Learning for Text Analysis in R*. CRC Press.

Iacobini, C. (2009). Phrasal verbs between syntax and lexicon. *Rivista di Linguistica 21(1)*: 97–117.

Iacobini, C., & Masini, F. (2007). Verb-particle Constructions and Prefixed Verbs in Italian: Typology, Diachrony and Semantics. In G. Booij et al. (Eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5), Fréjus 15–18 September 2005*. Bologna: University of Bologna, 2007.

Jackendoff, R. (2008). Construction after construction and its theoretical challenge. *Language 84*: 8–28.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-Particle Explorations*. Berlin: Mouton de Gruyter, pp. 67–94.

Jackendoff, R. S. & Audring, J. (2020). *The Texture of the Lexicon: Relational Morphology in the Parallel Architecture*. Oxford: Oxford University Press.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. In A. Hardie, & R. Love (Eds.), *7th International Corpus Linguistics Conference CL. Abstract book*. Lancaster: UCREL, pp. 125–127.

Janda, L. A. (1986). *A semantic analysis of the Russian verbal prefixes (za-, pere-, do- and ot-)*. Munich: Otto Sagner.

Jansen, L. (2007). Aristotle's Categories. *Topoi 26*: 153–158. Available online: https://doi.org/10.1007/s11245-006-9009-1.

Jurafsky, D., Bell, A., Gregory, M. L. & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, pp. 229–254.

Kenesei, I. (2007). Semiwords and affixoids: the territory between word and affix. *Acta Linguistica Hungarica 54*: 263–93.

Kiparsky, P. (1997). Remarks on denominal verbs. In: A. Alsina, J. Bresnan, & P. Sells (Eds.), *Complex Predicates*. Stanford, CA: CSLI Publications, pp. 473–499.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models. principles and techniques*. Cambridge, Massachusetts: The MIT Press.

Konietschke, F., Placzek, M., Schaarschmidt, F., & Hothorn, L. (2015). nparcomp: An R Software Package for Non-parametric Multiple Comparisons and Simultaneous Confidence Intervals. *Journal of Statistical Software*, *64(9)*: 1–17. http://dx.doi.org/10.18637/jss.v064.i09.

Koo, Terry, & Mae, L. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine 15(2)*: 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Köper, M., Scheible, C., & Schulte im Walde, S. (2015). Multilingual reliability and “semantic” structure of continuous word spaces. In *Proceedings of the 11th International Conference on Computational Semantics*. London: Queen Mary University of London, pp. 40–45.

Köper, M., & Schulte, S. im Walde. (2016). Distinguishing Literal and Non-Literal Usage of German Particle Verbs. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Lan- guage Technologies*. San Diego, California, USA, pp. 353–362.

Krongauz, M. A. (1998). *Pristavki i Glagoly v Russkom Jazyke: Semanticheskaja Grammatika* [Prefixes and Verbs in Russian Language: Semantic Grammar]. Moscow: Jazyki russkoj kul’tury.

Kühnhold, I., & Wellmann, H. (1973). *Deutsche Wortbildungstypen und Tendenzen in der Gegenwartssprache, 1. Teil: das Verb*. Düsseldorf: Pädagogischer Verlag Schwamm.

Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language* 66: 588–611.

Kutuzov A., & Kuzmenko E. (2017). WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In: Ignatov D. et al. (Eds.), *Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science*, vol 661. Springer, pp. 155–161.

Kühnhold, I., & Wellmann, H. (1973). *Deutsche Wortbildungstypen und Tendenzen in der Gegenwartssprache, 1. Teil: das Verb*. Düsseldorf: Pädagogischer Verlag Schwamm.

Langacker, R. (1987). *Foundations of cognitive grammar: theoretical prerequisites*. Vol. I. Stanford: Stanford University Press.

Langacker, R. (2009). Cognitive (Construction) grammar. *Cognitive Linguistics 20(1)*: 167–176. http://dx.doi.org/10.1515/COGL.2009.010.

Larsen, D. (2014). *Particles and particle-verb constructions in English and other Germanic languages*. Dissertation, University of Delaware. https://udspace.udel.edu/items/f2323297-1eb7-46e1-8e3c-f540aa6c66d9.

Laudanna, A. (1999). Regular versus irregular inflection: A question of levels. *Behavioral and brain sciences 22(6)*: 1028–1029.

Lehmann, V. (2009). Aspekt und Tempus im Slavischen. In S. Kempgen, P. Kosta, T. Berger, & K. Gutschmidt (Eds.), *Die slavischen Sprachen—The Slavic languages. Ein internationales Handbuch zu Ihrer Struktur, ihrer Geschichte und ihrer Erforschung—An international handbook of their structure, their history and their investigation (HSK 32.1)*. Berlin, New York: Mouton de Gruyter, pp. 526–556.

Levy, O., & Goldberg, Y. (2014). Dependency-Based Word Embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Baltimore, Maryland: Association for Computational Linguistics, pp. 302–308. https://www.aclweb.org/anthology/P14-2050.

Lewis, G., Solomyak, O., & Marantz, A. (2011). The neural basis of obligatory decomposition of suffixed words. *Brain and Language 118(3)*: 118–127. https://doi.org/10.1016/j.bandl.2011.04.004.

Lieber, R. (1992). *Deconstructing morphology*. Chicago: University of Chicago Press.

Lin, C.-C., Ammar, W., Dyer, C., & Levin, L. (2015). Unsupervised POS Induction with Word Embeddings. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Denver, Colorado: Association for Computational Linguistics, pp. 1311–1316.

Lindquist, H. (2009). *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.

Litvak, N., & Robert, Ph. (2012). A scaling analysis of a cat and mouse Markov chain. *The Annals of Applied Probability 22(2)*: 792–826.

Lohde, M. (2006). *Wortbildung des modernen Deutschen*. Tübingen: Narr.

Lorenz, D., & Tizón-Couto, D. (2017). Coalescence and contraction of V-to-Vinf sequences in American English—Evidence from spoken language. *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/cllt-2015-0067.

Manova, S. (2010). Suffix combinations in Bulgarian: parsability and hierarchy-based ordering. *Morphology 20*: 267–296.

Markova, A. (2011). On the nature of Bulgarian prefixes: Ordering and modification in multiple prefixation. *Word Structure 4(2)*: 244–271.

Marslen-Wilson, W., & Tyler, L. K. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences 2(11)*: 428–435.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review 101*: 3–33.

Marzi, C., Ferro, M., & Pirrelli V. (2014). Morphological structure through lexical parsability. *Lingue e Linguaggio XIII.2*: 263–290.

Masini, F., & Micheli, S. M. (2020). The morphological expression of approximation: the emerging simil- construction in Italian. *Word Structure 13(3)*: 371–402.

Mason, O. (2000). Parameters of collocation: the word in the centre of gravity. In J. M. Kirk (Ed.), *Corpora Galore. Analyses and Techniques in Describing English*. Amsterdam/Atlanta, GA:Rodopi, pp. 267–280.

Matushansky, O. (2002). On formal identity of Russian prefixes and prepositions. *MIT Working Papers in Linguistics 42*: 217–253.

Matzinger, T., Ritt, N., & Fitch, W.T. (2020). Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates. *PLoS ONE 15(4)*: e0230710. https://doi.org/ 10.1371/journal.pone.0230710.

McCarthy D., Keller, B., & Carroll, J. (2003). Detecting a Continuum of Compositionality in Phrasal Verbs. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*. Sapporo, Japan. Association for Computational Linguistics, pp. 73–80.

McIntyre, A. (2015). Denominal verbs: an overview. In S. Olsen, P. Müller, I. Ohnheiser, & F. Rainer (Eds.), *Word-formation: An International Handbook of the Languages of Europe*. Berlin: Mouton de Gruyter, pp. 434–450.

McIntyre, A. (2007). Particle verbs and argument structure. *Language and Linguistics Compass* 1/4: 350–367.

McIntyre, A. (2002). Idiosyncrasy in particle verbs. In N. Dehé, R. Jackendoff, A. McIntyre, S. Urban (Eds.), *Verb-Particle Explorations*. Berlin: Mouton de Gruyter, pp. 95–118.

McIntyre, A. (2001). Argument blockages induced by verb particles in English and German: event modification and secondary predication. *Structural aspects of semantically complex verbs*, pp. 131–164. Germanistik Online Datenbank. Retrieved 2023-01-15, from https://www.degruyter.com/database/GERMANISTIK/entry/ogerm.ga02144014_3/html.

Mel'čuk, I., & Polguère, A. (2018). Theory and practice of lexicographic definition. *Journal of Cognitive Science 19(4)*: 417–470.

Mendaza, R. (2015). Matching productivity indexes and diachronic evolution: The Old English affixes ful-, -isc, -cund, and -ful. *Canadian Journal of Linguistics / Revue Canadienne De Linguistique 60(1)*, 1–24.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR*. arXiv:1301.3781.

Milin, P., Smolka, E. & Feldman, L. B. (2017). Models of Lexical Access and Morphological Processing. In E. M. Fernández and H. S. Cairns (Eds), *The Handbook of Psycholinguistics*. Wiley-Blackwell, pp. 240–268.

Miloslavskij, I. G. (1980). *Voprosy slovoobrazovatel'nogo sinteza* [Problems of Word-Formation Synthesis]. Moscow: Izdatelstvo Moskovskogo universiteta.

Moder, C.L. (1992). *Productivity and Categorization in Morphological Classes*. PhD thesis, State University of New York at Buffalo, NY.

Monakhov, S. (2023a). How complex verbs acquire their idiosyncratic meanings. *Language and Speech*. https://doi.org/10.1177/00238309231199994. (in print)

Monakhov, S. (2023b). Probabilistic Method of Measuring Linguistic Productivity. *arXiv:2308.12643*. https://doi.org/10.48550/arXiv.2308.12643.

Monakhov, S. (2021). Russian prefixed verbs as constructional schemas. *Russian Linguistics 45*: 45–73.

Neapolitan, R. (2004). *Learning Bayesian Networks*. New York: Pearson Prentice Hall.

Neapolitan, R., & Jiang, X. (2007). *Probabilistic methods for financial and marketing informatics*. San Francisco: Elsevier. https://doi.org/10.1016/B978-0-12-370477-1.X5016-6.

Núñez, R. E. & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences 17*: 220–229.

Patterson, K., Lambon Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia 39(7)*: 709–724.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Wiley.

Pearl, J. (2009). *Causality. Models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: eight-month-old infants track backward transitional probabilities. *Cognition 113(2)*: 244–247. https://doi.org/10.1016/j.cognition.2009.07.011.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543.

Petre, P., & Cuyckens, H. (2008). Bedusted, yet not beheaded: the role of be-'s constructional properties in its conservation. In A. Bergs, & G. Diewald (Eds.), *Constructions and Language Change*. Berlin: Mouton de Gruyter, pp. 133–169.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming. A critical review. *Psychological Bulletin 134(3)*: 427–459.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language 39*: 633–651.

Pierrehumbert, J. B., & Granell, R. (2018). On hapax legomena and morphological productivity. In *Proceedings of the 15th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, Brussels, Belgium, October 31, 2018*, pp. 125–130.

Pinker, S., & Ullman, M. (2003). Beyond one model per phenomenon. *Trends in Cognitive Sciences 7(3)*: 108–109.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6: 456–463.

Pinker, S. (1999). *Words and Rules*. New York: Harper Perennial.

Plag, I. (2021). Productivity. In B. Aarts, A. McMahon, & L. Hinrichs (Eds). *The Handbook of English Linguistics* (2nd ed.). Wiley, pp. 483–499.

Plag, I. (2018). *Word-formation in English* (2nd ed.). Cambridge: Cambridge University Press.

Plag, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press.

Plag, I., & Baayen, H. (2009). Suffix Ordering and Morphological Processing. *Language 85(1)*: 109–152. doi:10.1353/lan.0.0087.

Plag, I., & Ben Hedia, S. (2018). The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In S. Arndt-Lappe, A. Braun, C. Moulin, & E. Winter-Froemel (Eds.), *Expanding the lexicon: Linguistic innovation, morphological productivity, and ludicity*. Berlin: Mouton de Gruyter, pp. 93–116. https://doi.org/10.1515/9783110501933-095.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review 103(1)*: 56–115.

Pluymaekers, M., Ernestus, M., & Baayen, H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustic Society of America 118*: 2561–2569.

Pustylnikov, O., & Schneider-Wiejowski, K. (2010). Measuring Morphological Productivity. *Studies in Quantitative Linguistics 5*: 1–9.

Pycha, A. (2013). Mechanisms for remembering roots versus affixes in complex words. *Journal of the Acoustical Society of America 133*: 3568. https://doi.org/10.1121/1.4806524.

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Radimský, J. (2020). A paradigmatic account of lexical innovation: the role of repeated components in French N+N compounds. In J. Audring, N. Kousoukos, & C. Manouilidou (Eds), *Rules, Patterns, Schemas, and Analogy: MMM12 Online Proceedings.* Patras: University Patras, pp. 77–91.

Ramchand, G. C. (2004). Time and the event: The semantics of Russian prefixes. *Nordlyd 32*(2): 323–366.

Rehurek, R., & Sojka, P. (2011). *Gensim–python framework for vector space modelling*. Brno: Masaryk University, Faculty of Informatics, NLP Centre.

Richardson, K. (2007). *Case and aspect in Slavic*. Oxford: Oxford University Press.

Romanova, E. (2004). Superlexical vs. lexical prefixes. *Nordlyd 32*(2): 255–278.

Rousseau, A. (Ed.) (1995). *Les préverbes dans les langues d'Europe. Introduction à l'étude de la préverbation*. Lille: Presses Universitaires du Septentrion.

Rychlý, P. (2008). A Lexicographer-Friendly Association Score. In P. Sojka, A. Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN2008*. Brno: Masaryk University, pp. 6–9.

Saldana, C., Oseki, Y., & Culbertson, J. (2021). Cross-linguistic patterns of morpheme order reflect cognitive biases: An experimental study of case and number morphology. *Journal of Memory and Language 118*: 104204. https://doi.org/10.1016/j.jml.2020.104204.

Šarić, L. (2003). Prepositional categories and prototypes: Contrasting some Russian, Slovenian, Croatian and Polish examples. *Jezikoslovlje 4(2)*: 187–204.

Schlechtweg, D., Eckmann, S., Santus, E., Schulte im Walde, S., & Hole, D. (2017). German in Flux: Detecting Metaphoric Change via Word Entropy. In *Proceedings of the 21st Conference on Computational Natural Language Learning*, Vancouver, pp. 354–367.

Schmid, H.-J. (2017). A framework for understanding linguistic entrenchment and its psychological foundations. In Hans-Jörg Schmid (Ed), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Berlin: De Gruyter, pp. 9–35.

Schreuder, R., & Baayen, H. (1995). Modeling morphological processing. In L. Feldman (Ed.), *Morphological Aspects of Language Processing*. Hillsdale, NJ: Erlbaum, pp. 131–156.

Schuppler, B., van Dommelen, W. A., Koreman, J., & Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics 40(4):* 595–607. https://doi.org/10.1016/j.wocn.2012.05.004.

Schwieter, J. W., & Wen, Zh. (Eds) (2022). *The Cambridge Handbook of Working Memory and Language*. Cambridge: Cambridge University Press.

Selkirk, E. (1982). *The syntax of words*. Cambridge, MA: MIT Press.

Sheskin, D. J. (2011). *Handbook of Parametric and Non-parametric Statistical Procedures*. Boca Raton, FL: Chapman and Hall/CRC Press.

Shvedova, N. (Ed.) (1980). *Russkaya Grammatika* [Russian Grammar]. Moscow: Nauka. http://www.rusgram.narod.ru. [September 2022].

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition 11(2)*: 245–260.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Smolka, E., Preller, K., & Eulitz, C. (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(1)*: 16–36.

Smolka, E., Khader, P., Wiese, R., Zwitserlood, P., & Rösler, F. (2013). Neurobiological evidence for the continuous processing of linguistic categories of regular and irregular verb inflection in German. *Journal of Cognitive Neuroscience 25(8)*: 1284–1304.

Sokal, R., & Rohlf, F. (1995). *Biometry* (3rd ed). New York: Freeman.

Spedicato, G. (2017). Discrete Time Markov Chains with R. *The R Journal 9(2)*: 84–104. https://doi.org/10.32614/RJ-2017-036.

Spencer, A., & Zaretskaya, M. (1998). Verb prefixation in Russian as lexical subordination. *Linguistics 36*: 1-40.

Stefanowitsch, A. (2008). Negative entrenchment. A usage-based approach to negative evidence. *Cognitive Linguistics 19*: 513–531.

Stein, S. D., & Plag, I. (2022). How relative frequency and prosodic structure affect the acoustic duration of English derivatives. *Laboratory Phonology 13(1)*. https://doi.org/10.16995/labphon.6445.

Stiebels, B. (1998). Complex denominal verbs in German and the morphology-semantics interface. In: G. Booij, J. Van Marle (Eds), *Yearbook of Morphology 1997*. Dordrecht: Springer. https://doi.org/10.1007/978-94-011-4998-3_10.

Stiebels, B. (1996). *Lexikalische Argumente und Adjunkte: Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Berlin, Boston: Akademie Verlag. https://doi.org/10.1515/9783050072319.

Swan, M. (2017). *Practical English usage*. (4th ed). Oxford : Oxford University Press.

Tatevosov, S. G. (2008). Intermediate prefixes in Russian. In A. Antonenko, C. Bethin, J. Baylin (Eds.), *Formal approaches to Slavic linguistics*. New York: Ann Arbor, MI: University of Michigan Press, pp. 423–442.

Taylor, J. (2012). *The Mental Corpus: How Language is Represented in the Mind*. Oxford: Oxford University Press.

Tikhonov, A. N. (1998). *Russkij glagol: Problemy teorii i leksikografirovanija* [Russian Verb: Problems of Theory and Lexicography]. Moscow: Academia.

Tikhonov, A. N. (1985). *Slovoobrazovatel'ny slovar russkogo jazyka* [Word-Formation Dictionary of Russian Language]. Moscow: Russkij Jazyk.

Tremblay, A., & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In *Perspectives on Formulaic Language: Acquisition and Communication*, pp. 151–173.

Turney, P. D., Neuman, Y., Assaf, D., & Cohen, Y. (2011). Literal and Metaphorical Sense Identification through Concrete and Abstract Context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh, pp. 680–690.

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research 37*: 141–188.

Ullman, M. T. (2004). Contributions of neural memory circuits to language: The declarative/procedural model. *Cognition 92(1–2)*: 231–270.

Varaksin, L. A. (1996). *Semantičeskij aspekt russkoj glagol'noj prefiksacii* [Semantic Aspect of Russian Verb Prefixation]. Ekaterinburg.

Varvara, R., Lapesa, G., & Padó, S. (2021). Grounding semantic transparency in context. *Morphology 31*: 409–446. https://doi-org.libproxy.viko.lt/10.1007/s11525-021-09382-w.

Vincent, N. (1999). The Evolution of C-Structure: Prepositions and PPs from Indo-European to Romance. *Linguistics 37(6)*: 1111–1153.

Volokhina, G. A., & Popova, Z. D. (1993). *Russkie glagol'nye pristavki: semantičeskoe ustrojstvo, sistemnye otnošenija* [Russian Verbal Prefixes: Semantics and Systematicity]. Voronezh.

Wang, S., Durrett, G., & Erk, K. (2018). Modeling Semantic Plausibility by Injecting World Knowledge. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. New Orleans, Louisiana, pp. 303–308.

Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software 6(60)*: 3021. https://doi.org/10.21105/joss.03021.

Wiegand, H. E. (1992). Elements of a theory towards a so-called lexicographic definition. *Lexicographica 8*: 175–289.

Wiland, B. (2012). Prefix stacking, syncretism, and the syntactic hierarchy. In M. Ziková & M. Dočekal (Eds.), *Slavic languages in formal grammar. Proceedings of FDSL 8.5, Brno 2010*. Frankfurt am Main: Peter Lang, pp. 307–324.

Zee, T., Ten Bosch, L., Plag, I., & Ernestus, M. (2021). Paradigmatic relations interact during the production of complex words: Evidence from variable plurals in Dutch. *Frontiers in Psychology 12*: 720017. https://doi.org/10.3389/fpsyg.2021.720017.

Zeller, J. (2001). Particle Verbs and Local Domains. Amsterdam: John Benjamins.

Zimmerer, F., Scharinger, M., & Reetz, H. (2014). Phonological and morphological constraints on German /t/-deletions. *Journal of Phonetics 45:* 64–75. https://doi.org/10.1016/j.wocn.2014.03.006.

Zirkel, L. (2010). Prefix combinations in English: structural and processing factors. *Morphology 20*: 239–266.

Appendix 1. Target and priming verbs in the nonce-base experiment

#	prefix	prefix meaning	target verb	priming verb (Condition 1)	priming verb (Condition 2)
prepositional prefixes					
1	*v-*	‘to place something somewhere by means of an action identified by the base verb’	*v-litj* ‘pour in’	*v-katitj* ‘roll in’	*do-bavitj* ‘add’
2	*ob-*	‘to surpass another performer of an action identified by the base verb’	*ob-igratj* ‘outplay’	*ob-skakatj* ‘outdo’	*po-beditj* ‘win’
3	*na-*	‘to accumulate in a certain amount by means of a surface-oriented action identified by the base verb’	*na-soritj* ‘litter on’	*na-lipnutj* ‘stick to’	*iz-gvazdatj* ‘make a mess of’
4	*ot-*	‘to perform an action identified by the base verb intensively, completely, and finally’	*ot-repetirovatj* ‘rehearse’	*ot-delatj* ‘decorate’	*raz-učitj* ‘prepare, read through’
5	*na-*	‘to accumulate in a certain amount by means of an action identified by the base verb’	*na-lovitj* ‘catch’	*na-ceditj* ‘pour in slowly’	*po-jmatj* ‘take hold of’
6	*pod-*	‘to perform (bring to fruition) an action identified by the base verb’	*pod-sčitatj* ‘calculate’	*pod-mesti* ‘sweep’	*pri-kinutj* ‘figure outj
7	*pro-*	‘to miss something while performing an action identified by the base verb'	*pro-karaulitj* ‘miss while watching outj	*pro-spatj* ‘oversleep’	*u-pustitj* ‘fail to catch’
8	*za-*	‘to perform (bring to fruition) an action identified by the base verb’	*za-gustetj* ‘thicken’	*za-minirovatj* ‘lay mines’	*o-krepnutj* ‘get stronger’
9	*ot-*	‘to separate something that was previously attached as a result of an action identified by the base verb; to annul of the result of such action’	*ot-lepitj* ‘detach’	*ot-kolotj* ‘break off; come up with’	*u-bratj* ‘remove’
10	*pro-*	‘to direct an action identified by the base verb through something inward’	*pro-gryztj* ‘gnaw through’	*pro-leztj* ‘crawl through’	*iz-rešetitj* ‘riddle’
11	*pod-*	‘to get closer, to join something with an action identified by the base verb’	*pod(o)-dvinutj* ‘move closer’	*pod-sestj* ‘sit near’	*pri-blizitsja* ‘come near to’
12	*nad-*	‘to additionally increase the size of the object by adding something to it (sometimes to its upper part) with the help of an action identified by the base verb’	*nad-vjazatj* ‘tie on’	*nad-stroitj* ‘build upon’	*pri-krepitj* ‘attach, fasten’
13	*do-*	‘to bring to an undesirable state by an action identified by the base verb’	*do-ezditj* ‘exhaust somebone’	*do-kanatj* ‘finish off someone’	*iz-mučitj* ‘overtire, enfeeble’

14	*pod-*	‘to direct an action identified by the base verb downwards, under something’	*pod-plytj* ‘swim under’	*pod(o)-stlatj* ‘lay under’	*pri-dvinutj* ‘move closer to’
15	*pod-*	‘to clean up something, remove all residues with an action identified by the base verb’	*pod-lizatj* ‘lick up’	*pod-edatj* ‘eat up’	*vy-draitj* ‘scour’
16	*ot-*	‘to head somewhere by means of an action identified by the base verb’’	*ot-vezti* ‘drive to'	*ot-tasčitj* ‘drag away’	*u-slatj* ‘send to’
17	*iz-*	‘to remove something from somewhere by means of an action identified by the base verb’	*iz-litj* ‘pour out (words, feelings)’	*iz-gnatj* ‘drive off’	*vy-razitj* ‘express’
18	*s-*	‘to perform an action identified by the base verb once’	*s-glupitj* ‘make a stupid thing’	*s-xoditj* ‘go to’	*o-šibitsja* ‘make a mistake'
19	*po-*	‘a repeated, sometimes also sequential, action identified by the base verb, which has been applied to all or many objects, or committed by all or many subjects’	*po-sažatj* ‘imprison’	*po-tajatj* ‘thaw outj	*arestovatj* ‘arrestj
20	*ob-*	‘to extend an action identified by the base verb to many objects (or to many places within a single space)’	*ob-ezditj* ‘go everywhere’	*ob-letetj* ‘fly around’	*na-vestitj* ‘drop in, come to see’
21	*pod-*	‘to perform an action identified by the base verb additionally and, as a rule, with insignificant intensity'	*pod-copitj* ‘save up’	*pod-mešatj* ‘mix in’	*po-berečj* ‘retain, store’
22	*za-*	‘to perform an action identified by the base verb in advance, beforehand, pre-emptively’	*za-stolbitj* ‘stake outj	*za-gotovitj* ‘prepare’	*po-metitj* ‘mark’
23	*na-*	‘to perform an action identified by the base verb intensively’	*na-bezobrazničatj* ‘mess up’	*na-gladitj* ‘iron’	*po-šalitj* ‘misbehave’
24	*po-*	‘to start an action identified by the base verb’	*po-bežatj* ‘start running’	*po-gnatsja* ‘start chasing’	*na-pravitsja* ‘head to’
25	*do-*	‘to bring to an end or to a limit an action identified by the base verb’	*do-letetj* ‘reach by flying’	*do-čitatj* ‘read through’	*pri-bytj* ‘arrive atj
26	*pro-*	‘to move forward, to overcome some distance by means of an action identified by the base verb’	*pro-exatj* ‘drive through’	*pro-plytj* ‘swim through’	*pri-xoditj* ‘reach, come to’
27	*iz-*	‘to extend an action identified by the base verb to many places, to many objects’	*iz-ranitj* ‘inflict wounds’	*iz-ezditj* ‘ride along and across’	*raz(o)-dratj* ‘shred, tear up’
28	*na-*	‘to teach someone something by means of an action identified by the base verb’	*na-muštrovatj* ‘train, prime’	*na-učitj* ‘teach’	*vy-dressirovatj* ‘train, prime’
29	*za-*	‘to perform an action identified by the base verb in passing; to deviate briefly from the main course of action’	*za-nesti* ‘bring in’	*za-jti* ‘come in’	*pri-voločj* ‘drag along with’
30	*s-*	‘to deliver from different places to the same place, to connect by means of an action identified by the base verb’	*s-tolknutj* ‘push againstj	*s-kleitj* ‘glue together’	*po-ssoritj* ‘sow discord’

31	*s-*	‘to perform (bring to fruition) an action identified by the base verb’	*s-komkatj* ‘crumple’	*s-mjagčitj* ‘soften’	*iz-lomatj* ‘break, crumble’
32	*po-*	‘to perform (bring to fruition) an action identified by the base verb’	*po-zavtrakatj* ‘eat breakfastj	*po-gibnutj* ‘perish’	*na-sytitsja* ‘feel full’
33	*po-*	‘to perform an action identified by the base verb within a certain period of time (often for a short time)’	*po-rabotatj* ‘work for a while’	*po-kuritj* ‘have a smoke’	*na-lomatsja* ‘work until exhaustion’
34	*s-*	‘to destroy, damage, deplete as a result of an action identified by the base verb’	*s-goretj* ‘burn down’	*s-ževatj* ‘chew up’	*vs-pyxnutj* ‘flare up’
35	*za-*	‘to apply an action identified by the base verb to a part of the objectj	*za-stiratj* ‘wash up’	*za-tesatj* ‘trim a log’	*po-čistitj* ‘clean’
36	*za-*	‘to perform an action identified by the base verb immediately after another action’	*za-ževatj* ‘chew something to get rid of an aftertaste’	*za-njuxatj* ‘sniff something to get rid of an aftertaste’	*pere-bitj* ‘get the taste out of one’s mouth’
37	*ot-*	‘to perform an action identified by the base verb in response to another action’	*ot-blagodaritj* ‘give creditj	*ot-reagirovatj* ‘reactj	*voz-nagraditj* ‘reward’
38	*na-*	‘to direct an action identified by the base verb to a surface of something; place something on the surface, bump into something’	*na-kleitj* ‘glue on’	*na-xlynutj* ‘inundate, overwhelm’	*pri-delatj* ‘attach, join’
39	*s-*	‘to remove something by means of an action identified by the base verb’	*s-mesti* ‘sweep away’	*s-britj* ‘shave off’	*pod(o)-rvatsja* ‘go off’
40	*pro-*	‘to perform an action identified by the base verb for some time (often for a long time)’	*pro-ždatj* ‘wait for a while’	*pro-voročatsja* ‘shift in bed for a while’	*sleditj* ‘watch closely, follow’
41	*ob-*	‘to direct an action identified by the base verb around an object in the path of movementj	*ob-exatj* ‘drive around something’	*ob(o)-jti* ‘bypass something’	*minovatj* ‘elude’
42	*pod-*	‘to perform an action identified by the base verb during another action or immediately after it, adapting to someone or something’	*pod-petj* ‘sing along’	*pod-igratj* ‘play along’	*za-skripetj* ‘squeak, screech’
43	*ob-*	‘to perform (bring to fruition) an action identified by the base verb’	*ob-venčatj* ‘wed’	*ob-vetšatj* ‘decay, deteriorate’	*po-ženitj* ‘marry’
44	*v-*	‘to fit in, to get inside something by means of an action identified by the base verb’	*v-polzti* ‘crawl into’	*v-letetj* ‘fly into’	*za-pastj* ‘fall into’
45	*ot-*	‘to refuse or to force the refusal of something by performing an action identified by the base verb’	*ot-govoritj* ‘talk out of something’	*ot-učitj* ‘wean off’	*raz-ubeditj* ‘dissuade’
46	*iz-*	‘to destroy, deplete, use up everything accessible through an action identified by the base verb’	*is-pisatj* ‘use up all writing utensils’	*is-streljatj* ‘shoot all the bullets’	*po-portitj* ‘spoil’
47	*pod-*	‘to direct an action identified by the base verb upwards’	*pod-brositj* ‘throw up’	*pod-djornutj* ‘jerk up’	*švyrnutj* ‘toss, fling’

48	*pred-*	‘to perform an action identified by the base verb in advance’	*pred-videtj* ‘foresee’	*pred(o)-steregatj* ‘warn’	*do-gadyvatsja* ‘guess’
49	*za-*	‘to get, earn, grab something through an action identified by the base verb’	*za-voevatj* ‘conquer’	*za-rabotatj* ‘earn’	*s-xvatitj* ‘seize, catch hold of’
50	*pro-*	‘to perform an action identified by the base verb intensively, thoroughly ’	*pro-ževatj* ‘chew and swallow’	*pro-dumatj* ‘think through’	*s-estj* ‘eat up’
51	*pro-*	‘to spend, exhaust, lose anything through an action identified by the base verb’	*pro-pitj* ‘exchange something for alcohol’	*pro-žitj* ‘spend a part of life’	*po-xititj* ‘steal’
52	*ob-*	‘to direct an action identified by the base verb around something or towards all sides of something’	*ob-žaritj* ‘fry’	*ob-lepitj* ‘cling to’	*za-peč’* ‘bake’
53	*nad-*	‘to apply an action identified by the base verb to a small part of the surface of an objectj	*nad-pilitj* ‘make a cut with a saw’	*nad-rezatj* ‘make a cut with a knife’	*pro-nizatj* ‘pierce’
54	*ot-*	‘to bring to an undesirable state (of damage, fatigue) as a result of an action identified by the base verb’	*ot-davitj* ‘tread on one’s footj	*ot-ležatj* ‘stay in bed until one’s limbs go numb’	*pri-ščemitj* ‘pinch’
55	*pod-*	‘to perform an action identified by the base verb in a secret, covert manner’	*pod-slušatj* ‘eavesdrop’	*pod-brositj* ‘plant (drugs, weapon)’	*raz(o)-bratj* ‘hear and understand’
56	*na-*	‘to perform an action identified by the base verb in a gentle, unobtrusive manner’	*na-igratj* ‘play music a bitj	*na-petj* ‘sing a bitj	*is-polnitj* ‘perform’
57	*na-*	‘to perform (bring to fruition) an action identified by the base verb’	*na-smešitj* ‘make someone laugh’	*na-močitj* ‘make something wetj	*raz-veselitj* ‘cheer up’
58	*do-*	‘to perform an action identified by the base verb as an addition to the previous action, which is necessary to meet the requirements’	*do-platitj* ‘pay in addition’	*do-slatj* ‘send in addition’	*pri-pljusovatj* ‘plus, add up’
59	*ob-*	‘to harm someone (sometimes, cheat someone) through an action identified by the base verb’	*ob-vorovatj* ‘rob of’	*ob-delitj* ‘deprive of’	*raz-grabitj* ‘plunder’
60	*iz-*	‘to perform an action identified by the base verb with a high degree of intensity’	*is-soxnutj* ‘get shallow’	*iz-zjabnutj* ‘get cold’	*za-čaxnutj* ‘languish, fade in’
61	*ot-*	‘to perform (bring to fruition) an action identified by the base verb’	*ot-iskatj* ‘find after some searching’	*ot-regulirovatj* ‘tune in, finesse’	*na-dybatj* ‘obtain, search outj
62	*za-*	‘to begin an action identified by the base verb’	*za-meljkatj* ‘start moving’	*za-gremetj* ‘start rattling’	*po-bežatj* ‘start moving’
63	*ot-*	‘to remove, to separate from something by means of an action identified by the base verb’	*ot-brositj* ‘throw away’	*ot-gryztj* ‘gnaw off’	*u-bratj* ‘take away’
64	*za-*	‘to move to a place (sometimes, remote) by means of an action identified by the base verb’	*za-brositj* ‘hurl to a great distance’	*za-prygnutj* ‘jump in, jump on’	*metnutj* ‘dartj

65	*za-*	'to cover up, close with something by means of an action identified by the base verb'	*za-pudritj* 'powder something'	*za-pjatnatj* 'blot one's reputation'	*pri-krytj* 'cover'
66	*pod-*	'to perform an action identified by the base verb with low intensity'	*pod-zabytj* 'alomost forgetj	*pod-bodritj* 'cheer up a bitj	*za-pamjatovatj* 'forgetj
67	*ot-*	'to end an action identified by the base verb that has lasted for a certain period of time'	*ot-gremetj* 'stop rumbling'	*ot-tsvesti* 'stop blossoming'	*pro-trezvonitj* 'chime'
68	*pred-*	'to find something in front of oneself as a result of an action identified by the base verb'	*pred-stavitj* 'imagine'	*pred-stojatj* 'awaitj	*v(o)-obrazitj* 'envisage'
69	*za-*	'to bring someone to an undesirable state (of unfitness, fatigue, exhaustion) through an action identified by the base verb'	*za-draznitj* 'tease someone'	*za-moročitj* 'make a fool out of'	*iz-vesti* 'bring to the end of one's tether'
70	*pro-*	'to move forward, to overcome some distance by means of an action identified by the base verb'	*pro-šagatj* 'pace'	*pro-nesti* 'pass (about danger)'	*ot-maxatj* 'cover a great distance'
71	*po-*	'to perform an action identified by the base verb with low intensity, sometimes also gradually'	*po-portitj* 'spoil a bitj	*po-otstatj* 'lag behind a bitj	*u-grobitj* 'ruin'
72	*iz-*	'to perform (bring to fruition) an action identified by the base verb'	*iz-lečitj* 'cure'	*is-pugatj* 'frighten'	*vos-kresitj* 'resurrectj
73	*pro-*	'to perform (bring to fruition) an action identified by the base verb'	*pro-demonstrirovatj* 'demonstrate'	*pro-zvučatj* 'sound'	*po-kazatj* 'show'
74	*o-*	'to direct an action identified by the base verb around something, on all sides of something'	*o-kutatj* 'envelop'	*o-ledenetj* 'freeze, get covered with ice'	*ob-voloč'* 'encapsulate'
75	*o-*	'to direct an action identified by the base verb past an object in the path of movementj	*o-bežatj* 'run around'	*o-plytj* 'swim around'	*ob-ognutj* 'circle, detour'
76	*o-*	'extend an action identified by the base verb to many objects (or to many places within a single object)'	*o-prositj* 'question, survey'	*o-delitj* 'endow'	*pro-intervjuirovatj* 'interview'
77	*o-*	'to perform (bring to fruition) an action identified by the base verb'	*o-bespokoitj* 'worry, raise concerns'	*o-čistitj* 'clean up'	*ras-stroitj* 'upset, unsettle'
78	*pri-*	'to reach a certain place, to arrive or to be delivered to a certain place by means of an action identified by the base verb, to join something'	*pri-bresti* 'reach some place while hobbling'	*pri-parkovatj* 'park (a car)'	*pod-ojti* 'come near'
79	*pri-*	'to perform an action identified by the base verb with little intensity, not completely'	*pri-vstatj* 'raise oneself a bitj	*pri-tormozitj* 'slow down, pull over'	*pod-njatsja* 'stand up'
80	*pri-*	'to perform an action as an addition to the action identified by the base verb; add something to what is already there'	*pri-kupitj* 'buy additionally'	*pri-sočinitj* 'lie a bit, decorate a story'	*pod-iskatj* 'seek outj
81	*pri-*	'to perform an action identified by the base verb during or immediately after another action'	*pri-svistnutj* 'whistle'	*pri-stuknutj* 'clatter'	*uljuljukatj* 'hootj

82	*pri-*	‘to perform (bring to fruition) an action identified by the base verb’	*pri-styditj* ‘shame’	*pri-laskatj* ‘caress’	*u-sovestitj* ‘reprobate’
83	*so-*	‘to perform an action identified by the base verb jointly'	*so-učastvovatj* ‘participate’	*so-suščestvovatj* ‘coexistj	*po-sobničatj* ‘abetj
84	*u-*	‘to move away from somewhere, to leave (force to leave) some place with the help of an action identified by the base verb’	*u-gnatj* ‘hijack’	*u-polzti* ‘crawl away’	*po-xititj* ‘steal’
85	*u-*	‘to get completely covered in something by means of an action identified by the base verb’	*u-stavitj* ‘set up’	*u-kutatsja* ‘wrap oneself up’	*za-xlamitj* ‘clutter’
86	*u-*	‘bring someone or something to an undesirable state (extreme fatigue, powerlessness, exhaustion) by means of an action identified by the base verb’	*u-ezditj* ‘wear down’	*u-kačatj* ‘to cause motion sickness’	*do-kanatj* ‘finish off’
87	*u-*	‘to get reduced by means of an action identified by the base verb’	*u-žatj* ‘reduce by squeezing’	*u-šitj* ‘stitch up’	*so-kratitj* ‘shorten’
88	*u-*	‘to destroy, deplete something by means of an action identified by the base verb’	*u-xlopatj* ‘spend a large amount of something in vain’	*u-plesti* ‘eat everything up’	*po-tratitj* ‘spend’
89	*u-*	‘to make something fit in somewhere by means of an action identified by the base verb’	*u-mestitj* ‘fit something in some place’	*u-pisatj* ‘use all provided space for writing’	*v-tisnutj* ‘squeeze in’
90	*u-*	‘to keep the posture identified by means of the base verb’	*u-terpetj* ‘keep patience’	*u-stojatj* ‘withstand’	*s-deržatsja* ‘hold back’
91	*u-*	‘to perform (bring to fruition) an action identified by the base verb’	*u-styditj* ‘make ashamed’	*u-žalitj* ‘sting’	*po-sramitj* ‘put someone to shame’
Unprepositional prefixes					
1	*vz-*	‘to get upwards by means of an action identified by the base verb’	*vz-loxmatitj* ‘dishevel’	*vz-letetj* ‘take off’	*ras-trepatj* ‘dishevel’
2	*vz-*	‘to perform an action identified by the base verb intensely or abruptly, suddenly’	*vz-vizgnutj* ‘screech’	*vz-dorožatj* ‘become expensive’	*pro-piščatj* ‘squeal’
3	*vz-*	‘to start an action identified by the base verb intensely or abruptly, suddenly’	*vz-revetj* ‘roar’	*vz-volnovatsa* ‘feel uneasy’	*rjavknutj* ‘bark'
4	*vz-*	‘to perform (bring to fruition) an action identified by the base verb’	*vz-besitj* ‘outrage'	*vs-potetj* ‘sweatj	*raz-jaritj* ‘infuriate’
5	*voz-*	‘to get upwards by means of an action identified by the base verb’	*voz-vesti* ‘erectj	*vos-paritj* ‘soar’	*po-stroitj* ‘build’
6	*voz-*	‘to perform of an action identified by the base verb once again’	*voz-roditj* ‘resurrectj	*vos-soedinitj* ‘reunite’	*o-živitj* ‘revive’
7	*voz-*	‘to start an action identified by the base verb’	*voz-likovatj* ‘rejoice’	*voz-nenavidetj* ‘start hating’	*ob-radovatsja* ‘be delighted’
8	*voz-*	‘to perform (bring to fruition) an action identified by the base verb’	*voz-mužatj* ‘come of age’	*vos-prepjatstvovatj* ‘hinder’	*za-materetj* ‘mature’

9	*vy-*	‘to move away, to stand out from something, to direct out by means of an action identified by the base verb’	*vy-lomatj* ‘break outj	*vy-karabkatsja* ‘make it through’	*ot-odratj* ‘rip off’
10	*vy-*	‘to perform an action identified by the base verb intensively and/or thoroughly’	*vy-belitj* ‘make white’	*vy-lizatj* ‘lick outj	*o-svetitj* ‘shed light on’
11	*vy-*	‘to get, obtain, find something by means of an action identified by the base verb’	*vy-lovitj* ‘catch, fish outj	*vy-stradatj* ‘achieve through suffering’	*po-jmatj* ‘catch’
12	*vy-*	‘to endure something or wait for something for some time while performing an action identified by the base verb’	*vy-sidetj* ‘wait for something inactively’	*vy-žitj* ‘survive’	*do-ždatsja* ‘receive after long waiting’
13	*vy-*	‘to perform (bring to fruition) an action identified by the base verb’	*vy-kopatj* ‘dig outj	*vy-lečitj* ‘cure’	*do-statj* ‘getj
14	*de-*	‘to perform an action contrary to the action identified by the base verb, thus annulling the result of the former action’	*de-šifrovatj* ‘decipher’	*de-mobilizovatj* ‘demobilise’	*ras-kodirovatj* ‘decode’
15	*dis-*	‘to perform an action contrary to the action identified by the base verb, thus annulling the result of the former action’	*dis-kvalificirovatj* ‘disqualify’	*dis-garmonirovatj* ‘be in disharmony’	*za-banitj* ‘ban’
16	*nedo-*	‘to perform an action identified by the base verb incompletely, fail to achieve the necessary standard’	*nedo-ocenitj* ‘underestimate’	*nedo-žaritj* ‘underbake’	*pre-umenjšitj* ‘belittle’
17	*niz-*	‘to direct an action identified by the base verb downwards’	*niz-vergnutj* ‘overthrow’	*niz-ojti* ‘descend’	*s-brositj* ‘drop’
18	*pere-*	‘to direct an action identified by the base verb from one place to another through an object or space’	*pere-brositj* ‘overthrow’	*pere-pilitj* ‘saw through’	*kinutj* ‘throw’
19	*pere-*	‘to place something between different objects or parts of one object by means of an action identified by the base verb’	*pere-sypatj* ‘sprinkle with something’	*pere-vitj* ‘join by twisting’	*s-mešatj* ‘mix’
20	*pere-*	‘to perform an action identified by the base verb repeatedly, anew, sometimes in a new manner’	*pere-delatj* ‘redo’	*pere-kroitj* ‘reshape’	*iz-menitj* ‘change’
21	*pere-*	‘to perform an action identified by the base verb repeatedly or sequentially, distribute it to all or many objects’	*pere-buditj* ‘wake up everyone’	*pere-byvatj* ‘have many people as visitors’	*pod-njatj* ‘raise’
22	*pere-*	‘to perform an action identified by the base verb with an undesirable duration and/or intensity’	*pere-xvalitj* ‘overpraise’	*pere-gretj* ‘overwarm’	*slavoslovitj* ‘sing the praises of’
23	*pere-*	‘to perform an action identified by the base verb intensively’	*pere-koverkatj* ‘twist up’	*pere-trusitj* ‘chicken outj	*iz-vratitj* ‘distortj
24	*pere-*	‘by means of an action identified by the base verb, to surpass another performer of the same action’	*pere-pljasatj* ‘win in a dancing contestj	*pere-sporitj* ‘get the upper hand in dispute’	*ob-skakatj* ‘outdo’

25	*pere-*	'to extend an action identified by the base verb to a specific, usually necessary or predetermined period of time'	*pere-ždatj* 'wait till the end of something'	*pere-zimovatj* 'live through the winter'	*po-vremenitj* 'wait, hold off'
26	*pere-*	'to cease an action identified by the base verb, usually after a long or intensive performing of the action'	*pere-xotetj* 'stop wanting'	*pere-broditj* 'stop brewing'	*raz-dumatj* 'change one's mind'
27	*pere-*	'to perform an action identified by the base verb in a brief, non-intensive manner'	*pere-kuritj* 'stop doing something for a smoke'	*pere-dohnutj* 'have a short respite'	*po-dymitj* 'puff out smoke'
28	*pre-*	'to perform an action identified by the base verb fully, intensively, sometimes excessively'	*pre-uveličitj* 'exaggerate'	*pre-ispolnitj* 'fill up'	*pri-sočinitj* 'elaborate'
29	*raz-*	'to spread in different directions, disengage by means of an action identified by the base verb'	*ras-krošitj* 'crumble'	*raz-oslatj* 'send outj	*iz-meljčitj* 'shred'
30	*raz-*	'to annul the result of an action identified by the base verb; to refuse or force to refuse to do something'	*raz-morozitj* 'defreeze'	*raz-minirovatj* 'demine'	*ot-tajatj* 'thaw outj
31	*raz-*	'to perform an action identified by the base verb with high inensity'	*ras-tolstetj* 'gain weightj	*ras-kritikovatj* 'chastise'	*o-žiretj* 'become obese'
32	*raz-*	'to perceive or explain something in detail by means of an action identified by the base verb'	*raz-gljadetj* 'discern'	*ras-tolkovatj* 'spell outj	*pri-metitj* 'notice'
33	*raz-*	'to perform (bring to fruition) an action identified by the base verb'	*ras-cvesti* 'bloom'	*raz-veselitj* 'make someone laugh'	*po-xorošetj* 'flourish, become more beautiful'
34	*re-*	'to perform an action identified by the base verb repeatedly, anew, sometimes in a new manner'	*re-organizovatj* 'reorganise'	*re-interpretirovatj* 'reinterpretate'	*pere-stroitj* 'rebuild'

Appendix 2. Hay and Baayen's (2002) data with productivity measures*

affix	hapaxes	tokens-P	token-PR	tokens	productivity
anti	48	259	0.44	589	0.082
be	26	1017	0.05	20340	0.001
con	20	790	0.15	5267	0.004
counter	29	491	0.92	534	0.054
cross	14	320	0.99	323	0.043
de	40	469	0.05	9380	0.004
dis	15	2187	0.16	13669	0.001
em	5	344	0.27	1274	0.004
en	21	1350	0.12	11250	0.002
fore	12	1683	0.75	2244	0.005
im	6	463	0.08	5788	0.001
in	58	1084	0.08	13550	0.004
inter	20	617	0.30	2057	0.010
mid	35	1065	0.94	1133	0.031
mis	12	1791	0.84	2132	0.006
non	56	264	0.33	800	0.070
out	29	2204	0.30	7347	0.004
over	60	3283	0.82	4004	0.015
pre	37	488	0.20	2440	0.015
re	76	6542	0.20	32710	0.002
self	19	393	0.52	756	0.025
sub	23	258	0.17	1518	0.015
super	46	364	0.67	543	0.085
trans	4	733	0.7	1047	0.004

un	61	4417	0.39	11326	0.005
under	27	1182	0.59	2003	0.013
able	57	3161	0.19	16637	0.003
age	31	1285	0.09	14278	0.002
al	40	2430	0.04	60750	0.001
an	47	398	0.03	13267	0.004
ance	6	671	0.05	13420	0.000
ant	17	816	0.09	9067	0.002
ary	8	931	0.16	5819	0.001
ate	15	385	0.07	5500	0.003
ation	28	1695	0.06	28250	0.001
dom	6	74	0.02	3700	0.002
ee	22	207	0.05	4140	0.005
eer	4	76	0.10	760	0.005
en	63	2443	0.12	20358	0.003
ence	7	167	0.01	16700	0.000
ent	17	354	0.01	35400	0.000
er	251	19872	0.21	94629	0.003
ery	21	542	0.10	5420	0.004
ese	4	20	0.01	2000	0.002
ess	18	249	0.18	1383	0.013
ette	10	80	0.05	1600	0.006
fold	9	162	0.99	164	0.055
ful	43	4391	0.25	17564	0.002
hood	8	1441	0.67	2151	0.004
ian	29	185	0.04	4625	0.006
ic	39	550	0.03	18333	0.002
ier	2	184	0.64	288	0.007
ify	7	1701	0.38	4476	0.002
ish	59	1286	0.10	12860	0.005
ism	16	1540	0.27	5704	0.003
ist	39	1001	0.13	7700	0.005

itis	4	2	0.01	200	0.020
ity	33	1916	0.06	31933	0.001
ive	19	857	0.12	7142	0.003
ize	13	1664	0.15	11093	0.001
less	119	5313	0.74	7180	0.017
let	19	305	0.23	1326	0.014
like	270	539	0.76	709	0.381
ling	0	110	0.10	1100	0.000
ly	198	16347	0.10	163470	0.001
ment	21	370	0.01	37000	0.001
most	7	270	0.73	370	0.019
ness	128	3845	0.23	16717	0.008
oid	4	21	0.11	191	0.021
or	62	2013	0.12	16775	0.004
ory	8	2239	0.57	3928	0.002
ous	16	1142	0.07	16314	0.001
proof	14	183	0.71	258	0.054
ry	25	537	0.11	4882	0.005
ship	24	1000	0.36	2778	0.009
some	11	862	0.74	1165	0.009
ster	12	586	0.21	2790	0.004
th	1	1959	0.10	19590	0.000
ward	12	2932	0.89	3294	0.004
y	244	6522	0.13	50169	0.005

* Some columns are omitted, one column is added.

Appendix 3. English stimuli, their frequencies, frequency ratios, transitional probabilities*

word	type	BFA	BFF	DF	DBRF	DBRA	PAB	PBA	TPR
				con-					
conduct	LH	174273	46843243	3318625	0.071	19.043	0.018	0.087	-1.54
contact	HH	28865	1387217	5346901	3.854	185.23	0.075	0.141	-0.63
confuse	LL	239575	2294382	556100	0.242	2.321	0.017	0.014	0.15
confess	HL	6261	25481613	262738	0.010	41.964	0.057	0.006	2.10
contrast	PA	0	0	1483715	0	0	0	0	0
				en-					
enable	LH	7344053	10239826	2982423	0.291	0.406	0.006	0.098	-2.75
enjoy	HH	892144	2480430	4761271	1.920	5.337	0.113	0.157	-0.32
entrust	LL	2835392	3611962	120043	0.033	0.042	0.005	0.003	0.25
engrave	HL	463984	539394	115700	0.214	0.249	0.023	0.003	1.83
enhance	PA	0	0	2481571	0	0	0	0	0
				de-					
decrease	LH	68429	11269236	1797654	0.160	26.270	0.004	0.025	-1.64
describe	HH	83675	2274956	5142566	2.261	61.459	0.028	0.073	-0.93
deforest	LL	1782270	2150472	7097	0.003	0.004	0.0001	0.0001	0.26
debunk	HL	69164	172653	60517	0.351	0.875	0.016	0.0008	2.95
devise	PA	0	0	379154	0	0	0	0	0
				dis-					
disease	LH	825145	4853308	4328576	0.892	5.246	0.011	0.126	-2.43
discuss	HH	7509	9231859	3750241	0.406	499.43	0.082	0.110	-0.28
distaste	LL	1660540	1796155	18916	0.011	0.011	0.0009	0.0005	0.51
disguise	HL	77689	101484	172349	1.698	2.218	0.057	0.005	2.42
disdain	PA	0	0	98376	0	0	0	0	0

in-									
intend	LH	1933371	17355833	1823236	0.105	0.943	0.004	0.021	-1.54
install	HH	304580	5153941	2312644	0.449	7.593	0.018	0.026	-0.35
inbreed	LL	819073	1432468	9030	0.006	0.011	0.0002	0.0001	0.79
insane	HL	84260	119268	191250	1.604	2.270	0.014	0.002	1.90
invite	PA	0	0	2766631	0	0	0	0	0
pre-									
prevent	LH	264828	38943255	3154308	0.081	11.911	0.039	0.141	-1.27
prepare	HH	35315	13553470	3425442	0.253	96.997	0.071	0.153	-0.77
preheat	LL	2599056	3833869	38928	0.010	0.015	0.001	0.001	0.08
precede	HL	48493	763409	330495	0.433	6.815	0.075	0.014	1.63
premier	PA	0	0	1304504	0	0	0	0	0
re-									
report	LH	1747225	126414613	12245124	0.097	7.008	0.009	0.068	-1.99
resource	HH	6225845	6369149	4952968	0.778	0.796	0.012	0.027	-0.79
redraw	LL	3433973	4272393	16979	0.004	0.005	0.0001	0.00009	0.23
relax	HL	54484	578176	995732	1.722	18.276	0.025	0.005	1.51
regime	PA	0	0	1417623	0	0	0	0	0
out-									
outright	LH	16740346	17942002	165396	0.009	0.010	0.005	0.011	-0.73
outfit	HH	3112674	8090238	334775	0.041	0.108	0.018	0.022	-0.20
outgrow	LL	6609979	6981134	38336	0.005	0.006	0.003	0.002	0.36
outcrop	HL	1121673	1164821	34839	0.030	0.031	0.019	0.002	2.10
outage	PA	0	0	94366	0	0	0	0	0

* BFA—base frequency alone, i.e. frequency of *able* in *enable*; BFF—base family frequency, i.e. combined frequency of *able*, *disable*, etc.; DF—derivation frequency; DBRF—derivation/base ratio (family), i.e. DF/BFF; DBRA—derivation/base ratio (alone), i.e. DF/BFA; PAB—transitional probability P (affix | base); PBA—transitional probability P (base | affix); TPR—log ratio PAB/PBA.

Appendix 4. Russian stimuli, their frequencies, frequency ratios, transitional probabilities*

word	type	BFA	BFF	DF	DBRF	DBRA	PAB	PBA	TPR
				pod-					
podpisatj	LH	2959970	5961378	226459	0.038	0.077	0.008	0.05	-1.75
podgotovitj	HH	882285	1088924	224561	0.206	0.255	0.040	0.049	-0.20
podbrositj	LL	310894	524267	14436	0.028	0.046	0.006	0.003	0.68
podmetatj	HL	53272	162798	11055	0.068	0.208	0.015	0.002	1.81
podlichatj	PA	0	0	565	0	0	0	0	0
				za-					
zabratj	LH	1655153	3504961	184273	0.053	0.111	0.001	0.011	-2.10
zakonchitj	HH	100100	387068	290630	0.751	2.903	0.011	0.017	-0.41
zaplakatj	LL	273468	291132	42680	0.147	0.156	0.003	0.002	0.29
zamorozitj	HL	6592	15725	11416	0.726	1.732	0.011	0.0006	2.80
zatevatj	PA	0	0	28334	0	0	0	0	0
				ot-					
othoditj	LH	1333752	25153753	197690	0.008	0.148	0.001	0.033	-3.35
otkrytj	HH	13691	649390	762615	1.174	55.702	0.079	0.127	-0.47
otkinutj	LL	87099	540337	17332	0.032	0.199	0.004	0.002	0.45
otklonitj	HL	14313	44159	27183	0.616	1.899	0.056	0.004	2.51
otrinutj	PA	0	0	6542	0	0	0	0	0
				na-					
nazhatj	LH	92268	20031758	251846	0.013	2.730	0.001	0.016	-2.57
nazvatj	HH	458756	1314836	619185	0.471	1.350	0.032	0.040	-0.22
nadelitj	LL	280395	1106203	40569	0.037	0.145	0.003	0.002	0.29
navreditj	HL	61301	117949	67561	0.573	1.102	0.037	0.004	2.11
nagletj	PA	0	0	4553	0	0	0	0	0

pere-									
peredatj	LH	2262028	24586215	266012	0.011	0.118	0.003	0.053	-2.79
pereehatj	HH	840977	2736934	136905	0.050	0.163	0.014	0.027	-0.63
perezvonitj	LL	454236	941034	28612	0.030	0.063	0.009	0.005	0.44
pereputatj	HL	122983	159399	30813	0.193	0.251	0.049	0.006	2.07
perechitj	PA	0	0	8238	0	0	0	0	0
vy-									
vyvoditj	LH	149030	13411577	523589	0.039	3.513	0.003	0.046	-2.46
vygljadetj	HH	392996	587726	1358391	2.311	3.457	0.073	0.121	-0.49
vyslushatj	LL	764824	1006671	49452	0.049	0.065	0.004	0.004	0.11
vyrastitj	HL	45883	65852	32193	0.489	0.702	0.034	0.002	2.48
vychestj	PA	0	0	4305	0	0	0	0	0
vz-									
vzdumatj	LH	3599138	4991554	29574	0.006	0.008	0.001	0.020	-2.52
vzletetj	HH	341725	845882	40412	0.048	0.118	0.012	0.027	-0.78
vzlomatj	LL	139986	379288	12283	0.032	0.088	0.008	0.008	0.03
vzmahnutj	HL	60139	69706	13206	0.189	0.220	0.043	0.008	1.58
vzorvatj	PA	0	0	24309	0	0	0	0	0
ob-									
obhoditj	LH	1333752	25105143	246300	0.010	0.185	0.003	0.144	-3.71
obsuditj	HH	778254	886790	181891	0.205	0.234	0.061	0.106	-0.54
obmytj	LL	194595	371315	2634	0.007	0.014	0.002	0.001	0.50
oblizatj	HL	64131	81559	13252	0.162	0.207	0.050	0.007	1.87
obidetj	PA	0	0	44681	0	0	0	0	0

* BFA—base frequency alone, i.e. frequency of *pisatj* ‘write’ in *podpisatj* ‘subscribe’; BFF—base family frequency, i.e. combined frequency of *pisatj*, *perepisatj* ‘rewrite’, etc.; DF—derivation frequency; DBRF—derivation/base ratio (family), i.e. DF/BFF; DBRA—derivation/base ratio (alone), i.e. DF/BFA; PAB—transitional probability P (affix | base); PBA—transitional probability P (base | affix); TPR—log ratio PAB/PBA.

Appendix 5. English prefixed stimuli from Hay's paper (2001), their frequencies, frequency ratios, and transitional probabilities*

word	type	BFA	BFF	DF	DBRF	DBRA	PAB	PBA	TPR
group A									
refurbish	HL	1175	75635	98361	1.300	83.711	0.002	0.00008	3.38
inaudible	HL	76289	88466	44635	0.505	0.585	0.0009	0.00008	2.44
incongruous	HL	1648	4682	15106	3.226	9.166	0.002	0.00002	4.42
uncanny	HL	9494	9988	44325	4.438	4.669	0.0009	0.00005	2.77
unleash	HL	80300	90822	154652	1.703	1.926	0.0007	0.0002	1.32
immutable	HL	12009	19044	34789	1.827	2.897	0.0008	0.00004	2.85
unobtrusive	HL	7072	15162	21384	1.410	3.024	0.0008	0.00002	3.44
entwine	HL	27215	99990	21462	0.215	0.789	0.0002	0.00002	2.09
immortal	HL	190496	821057	128506	0.157	0.675	0.0004	0.0001	1.02
illegible	LL	24693	29523	28228	0.956	1.143	0.0006	0.0004	0.39
intractable	HL	11063	52782	33846	0.641	3.059	0.001	0.00006	2.8
uncouth	HL	804	2036	7879	3.870	9.800	0.001	0.00001	4.64
impatient	LH	7861337	8362368	61534	0.007	0.008	0.000008	0.00008	-2.23
revamp	HL	18951	262112	83375	0.318	4.400	0.002	0.00007	3.33
inanimate	HL	118688	821005	30230	0.037	0.255	0.0005	0.00005	2.18
reiterate	HL	39947	2018307	138395	0.069	3.464	0.001	0.0001	2.13
immobile	LH	1907792	3313397	19921	0.006	0.010	0.000008	0.00002	-1.14
group B									
rekindle	HL	47611	52700	27671	0.525	0.581	0.0004	0.00002	2.85
inadequate	LL	694688	950866	281867	0.296	0.406	0.0007	0.0005	0.36
invulnerable	LL	589805	595060	11346	0.019	0.019	0.00004	0.00002	0.84

uncommon	LH	5060568	6492455	198201	0.031	0.039	0.00004	0.0002	-1.79
unscrew	LL	518127	632215	18148	0.029	0.035	0.00003	0.00002	0.48
immoderate	LL	668640	1053011	3822	0.004	0.006	0.000006	0.000006	0.27
unaffected	LL	298282	340039	76823	0.226	0.258	0.0002	0.0001	0.79
enshrine	HL	172864	177265	50010	0.282	0.289	0.0002	0.00006	1.49
immoral	HH	1223139	1970668	86060	0.044	0.070	0.00007	0.0001	-0.49
illiberal	LH	791773	1158030	9026	0.008	0.011	0.00001	0.0001	-2.52
impractical	HH	1517659	1907834	52096	0.027	0.034	0.00004	0.00007	-0.56
unkind	LH	5491034	6976852	17872	0.003	0.003	0.000003	0.00002	-1.89
imperfect	LH	3213526	4603007	102911	0.022	0.032	0.00003	0.0001	-1.36
retool	HH	4427230	5013733	13657	0.003	0.003	0.000007	0.00001	-0.5
inaccurate	LL	1106796	1545201	125936	0.082	0.114	0.0002	0.0002	0.12
reorganise	LL	1999774	10719601	66094	0.006	0.033	0.0001	0.00005	0.67
immodest	HL	316905	406756	4167	0.010	0.013	0.00001	0.000005	1

* BFA—base frequency alone, i.e. frequency of *organise* in *reorganise*; BFF—base family frequency, i.e. combined frequency of *organise*, *disorganise*, etc.; DF—derivation frequency; DBRF—derivation/base ratio (family), i.e. DF/BFF; DBRA—derivation/base ratio (alone), i.e. DF/BFA; PAB—transitional probability P (affix | base); PBA—transitional probability P (base | affix); TPR—log ratio PAB/PBA.

Appendix 6. Experimental stimuli of the prefix-base entropy study

#	prefix	verb	meaning
1	de-	deshifrovatj	decipher
2	dis-	diskvalifitsirovatj	disqualify
3	do-	doezditj	ruin, spoil
4	do-	doplatitj	pay extra
5	do-	dochitatj	finish reading
6	iz-	izgnatj	drive out, expel
7	iz-	izlechitj	heal, cure
8	iz-	izranitj	hurt, wound
9	iz-	ispisatj	cover with writing
10	iz-	issohnutj	become thin, wither
11	na-	nagladitj	iron a lot
12	na-	nakleitj	glue on
13	na-	nalipnutj	stick to
14	na-	nalovitj	catch a lot
15	na-	napetj	hum
16	na-	nasmeshitj	make laugh
17	na-	nauchitj	teach
18	nad-	nadvjazatj	tie, bind above smth
19	nad-	nadpilitj	make an incision
20	nedo-	nedozharitj	undercook
21	niz-	nizvergnutj	overthrow
22	o-	obezhatj	run around
23	o-	okutatj	envelop
24	o-	oprositj	interrogate
25	o-	ochistitj	clear
26	ob-	obvenchatj	marry

27	ob-	obdelitj	deprive
28	ob-	obzharitj	fry
29	ob-	obletetj	fly around
30	ob-	obskakatj	get to windward of smb
31	ob-	objehatj	go around
32	ot-	otblagodaritj	thank
33	ot-	otbrositj	throw away
34	ot-	otgovoritj	talk out of smth
35	ot-	otgremetj	stop rumbling
36	ot-	otdavitj	press hard, crush
37	ot-	otkolotj	break off
38	ot-	otregulirovatj	adjust
39	ot-	otrepetirovatj	rehearse
40	ot-	ottaschitj	drag away
41	pere-	perebuditj	wake up many people
42	pere-	peredelatj	redo
43	pere-	perezhdatj	wait out
44	pere-	perekuritj	smoke for a while
45	pere-	perepilitj	saw through
46	pere-	perepljasatj	outdance
47	pere-	peresypatj	sprinkle with smth
48	pere-	peretrusitj	get the wind up
49	pere-	perehvalitj	overpraise
50	pere-	perehotetj	stop wanting
51	po-	pobezhatj	start running
52	po-	poobedatj	dine
53	po-	poportitj	spoil
54	po-	porabotatj	work for a while
55	po-	posazhatj	plant / put in prison
56	pod-	podbodritj	cheer smb up
57	pod-	podbrositj	toss up
58	pod-	podkopitj	save up

59	pod-	podlizatj	suck up
60	pod-	podmesti	sweep, broom
61	pod-	podpetj	sing along
62	pod-	podplytj	swim near to smth
63	pod-	podsestj	sit next to smth
64	pod-	podslushatj	overhear, eavesdrop
65	pre-	preuvelichitj	exaggerate
66	pred-	predvidetj	foresee
67	pred-	predstavitj	introduce / imagine
68	pri-	pribresti	wander into
69	pri-	prikupitj	buy additionally
70	pri-	prisvistnutj	whistle
71	pri-	pristyditj	put to shame
72	pri-	pritormozitj	slow down
73	pro-	progryztj	gnaw through
74	pro-	produmatj	think through
75	pro-	proehatj	drive through
76	pro-	prozhdatj	wait for a long time
77	pro-	prozvuchatj	resound
78	pro-	propitj	drink away
79	pro-	prospatj	oversleep
80	pro-	proshagatj	walk for some time
81	raz-	razgljadetj	discern
82	raz-	razmorozitj	unfreeze
83	raz-	raskritikovatj	chastise, reprimand
84	raz-	raskroshitj	crumble
85	raz-	rastsvesti	bloom
86	re-	reinterpretirovatj	reinterpret
87	s-	sbritj	shave off
88	s-	sgoretj	burn down
89	s-	skleitj	glue together
90	s-	smjagchitj	soften, attenuate

91	s-	shoditj	visit
92	so-	sosuschestvovatj	live together
93	u-	ugnatj	steal, hijack
94	u-	uzhatj	squeeze
95	u-	ukachatj	become seasick
96	u-	umestitj	fit, squeeze in
97	u-	uplesti	gobble
98	u-	ustavitj	clutter
99	u-	ustyditj	shame
100	u-	uterpetj	endure
101	v-	vkatitj	roll in
102	v-	vpolzti	crawl in
103	voz-	vozvesti	erect
104	voz-	vozlikovatj	rejoice
105	voz-	vozmuzhatj	mature
106	voz-	vozroditj	revive
107	vy-	vybelitj	bleach
108	vy-	vykopatj	dig out
109	vy-	vylovitj	fish out
110	vy-	vylomatj	break out
111	vy-	vysidetj	sit out
112	vz-	vzbesitj	infuriate
113	vz-	vzvolnovatjsja	get excited
114	vz-	vzdorozhatj	rise in price
115	vz-	vzletetj	take off
116	za-	zabrositj	hurl, toss
117	za-	zavoevatj	conquer
118	za-	zagotovitj	prepare, stock
119	za-	zagremetj	start rumbling
120	za-	zadraznitj	tease, bully
121	za-	zaminirovatj	mine
122	za-	zanesti	bring in

123	za-	zanjuhatj	sniff
124	za-	zapudritj	powder
125	za-	zastiratj	wash out

Erklärung

gemäß § 5 (1) Punkt 3 der Promotionsordnung der Philosophischen Fakultät

Ich erkläre,

dass mir die geltende Promotionsordnung der Fakultät bekannt ist;

dass ich die Dissertation selbst angefertigt, keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben habe;

dass mich ausschließlich die folgenden Personen bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt haben;

dass die Hilfe einer kommerziellen Promotionsvermittlung nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für die Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen;

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe;

dass ich nicht die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht habe (wenn doch, bitte Ergebnis angeben).

Jena, 07.09.2023

Sergei Monakhov