Journal of Linguistics (2026), **62**, 79–111
doi:10.1017/S0022226723000385

Journal of Linguistics
CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Parsability revisited and reassessed

Sergei Monakhov

Friedrich-Schiller-Universität Jena, Germany
Email: sergei.monakhov@uni-jena.de

Received: 12 September 2022; **Revised:** 10 November 2023; **Accepted:** 18 November 2023

Keywords: complex words, compositionality, parsability, productivity, semantic transparency

Abstract

This paper provides evidence that the inveterate way of assessing linguistic items' degrees of analysability by calculating their derivation to base frequency ratios may obfuscate the difference between two meaning processing models: one based on the principle of compositionality and another on the principle of parsability. I propose to capture the difference between these models by estimating the ratio of two transitional probabilities for complex words: P (affix | base) and P (base | affix). When transitional probabilities are comparably low, each of the elements entering into combination is equally free to vary. The combination itself is judged by speakers to be semantically transparent, and its derivational element tends to be more linguistically productive. In contrast, 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 are also considered to be semantically complex but appear less transparent because the 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.

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 & 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 & Culbertson 2021; Zee et al. 2021). However, many researchers who have examined

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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 that 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 & Ben Hedia 2018), other studies reported no such effect or even an effect in the opposite direction (Pluymaekers, Ernestus & Baayen 2005; Schuppler et al. 2012; Zimmerer, Scharinger & Reetz 2014; Ben Hedia & Plag 2017; Stein & 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 & 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 & 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 & 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 that 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 Hay, 'the chances of a high-frequency derived form (…) being more frequent than its base are much higher than the chances of a low-frequency derived form being more frequent than its base' (Hay 2001: 1052). 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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 & Segui 1989; Schreuder & Baayen 1995; De Jong, Schreuder & 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 hapaxes: *sub-measles* and *sub-banksit.* Both of them have frequencies of one 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). 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. Monakhov 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 & Schreuder 1992; Clahsen 1999; Pinker & Ullman 2003; Ullman 2004; Silva & 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 2010), 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 (2010), 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 & 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 combination types: type (1), both elements are fixed (e.g. *con-tact*); type (2), both elements are variables (e.g. *non-linear*); type (3), the first element is a variable and the second element is fixed (e.g. *em-power*); and type (4), the first element is fixed and the second element is a

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

variable (e.g. *un-couth*). Note that under this approach, both productive and unproductive affixes and both free and bound bases can be either fixed elements or slot fillers.

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 because 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), 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, because 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). 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, yet the meanings of the bases might have nothing in common (Stiebels 1996; Zeller 2001; Monakhov 2023a).

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; Dressler 2005; Varvara, Lapesa & Padó 2021). Because 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 & 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 & 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, 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

concept (semantically) parsable (although, of course, this combination remains perfectly morphologically analysable).

The pole of the combination of words is a continuum of its own. 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 sense.

Finally, this logic can be extended to the sub-continuum of multi-morphemic words. Among them, one can easily distinguish between i) words of type (1) that behave like idioms, where both elements are fixed; ii) words of types (3) and (4) that resemble collocations, with one element fixed and another one free to vary; and iii) words of type (2) that can be thought of as free combinations of morphemes, where both elements are slots.

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). This hierarchy is visualised in Figure 1.

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

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

or

$$Meaning_{ITEM} = X + Meaning_{COMPONENT\ 2}$$

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.

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.

The rest of the paper is structured as follows. In the next section, I show that the construction type of a given complex word (and the meaning processing model associated

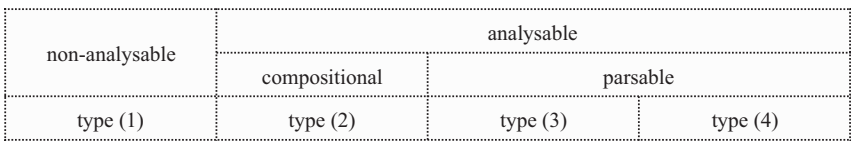


Figure 1 Schema of analysability/compositionality/parsability relationship.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

with this type) can be inferred by calculating the log ratio of the transitional probabilities P (affix | base) and P (base | affix).

Study 1 is dedicated to probing into the cognitive reality of the four conjectured construction types. Specifically, I am interested in whether speakers perceive complex words of types (3) and (4) differently than complex words of types (1) and (2) with regard to their morphological analysability and semantic transparency. 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 be 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.

2. Complex words' construction types and transitional probabilities

One might hypothesise that under the dual-route model's 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 non-analysable expression 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; Monakhov 2023a).

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. Because one fixed element normally appears

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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', and *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-polieren* 'polish up', *auf-motzen* 'pimp up', *auf-polstern* 'pad up', and *auf-putschen* 'pump 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 & 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 & Marantz 2011). This is yet another illustration of the aforementioned conflation problem, because complex words of types (1) and (4) reveal comparable 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 2).

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,

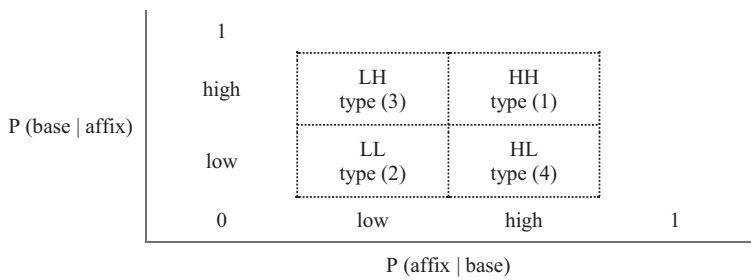


Figure 2 Schema of complex words' transitional probabilities' patterns.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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, we expect to find that the fixed element communicates less information about the filler than the filler about the fixed element (cf. Gries & Stefanowitsch 2004).

The expected pattern is schematised in Figure 2 (the two-letter abbreviations proposed therein will be used as shorthand for respective construction types throughout the rest of the paper). 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.

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 positioning of HH and LL on this number line may raise questions. For both constructions, the transitional probabilities are assumed to be the same, that is, their ratio should be unity, and log ratio should be equal to zero. However, given the way I suggest calculating transitional probabilities, one might never find a word for which they will be exactly equal (for example, for HH type, this will require a word whose affix only combines with its base and whose base only combines with its affix). Besides, I needed a cut-off point to create two groups of experimental stimuli. So, the best way to think about the distribution of measures above is in terms of HH and LL log ratios approaching the limit of zero from two different poles. Thus, some LH words may become more HH-like as their P (affix | base) → 1 while P (base | affix) remains high. In this scenario, a certain base loses its ability to accommodate different affixes and is constrained to occur with one affix. Conversely, some HL words may become more LL-like as their P (affix | base) → 0 while P (base | affix) remains low. In this scenario, a certain base broadens the scope of its applicability and can be combined with multiple affixes rather than with just one.

3. Study 1: Perceived complexity of the different types of complex words

3.1. 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 speakers 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

more complex than HH words, it naturally follows that LH and HL words should be placed somewhere in between in this hierarchy.

The experiments described below were conducted on English and Russian data. 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, and 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, because in Russian, LH and HL constructions are semantically differentiated (see below), controlling for their meanings seemed infeasible, so I did not do this for English data either.

Words of both languages were assigned to construction types based on the values of their transitional probabilities' log ratios: (i) LH: log ratio < -1 (with the exception of *out-*, for which -0.73 was the lowest value in my data); (ii) HH: $-1 <$ log ratio < 0; (iii) LL: $0 <$ log ratio < 1; and (iv) 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 nine billion words). English stimuli can be found in Appendix 1 and Russian stimuli in Appendix 2.

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 lead 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.*

One can also estimate the base frequency of a particular derivation in a different way, calculating first the cumulative root frequency (Cole, Beauvillain & 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 & Baayen 2000). This method has some desirable properties. First, it allows stabilising 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 (*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 3, 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

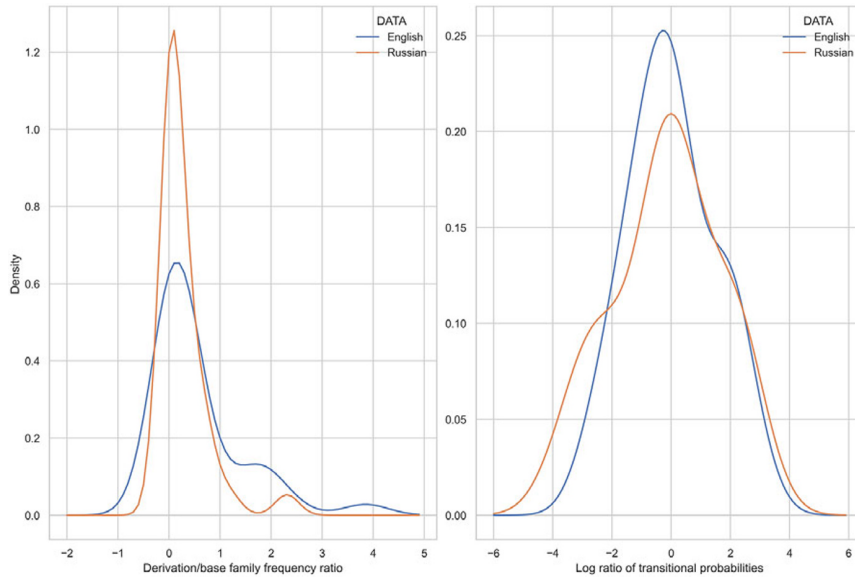


Figure 3 (Colour online) Densities of derivation to base family frequency ratios (left panel) and transitional probabilities' log ratios (right panel), English and Russian.

Table 1 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	$D = 0.18$, $p = 0.19$	$D = 0.20$, $p = 0.10$

ratios are, in contrast, normally distributed (Figure 3, right panel) as verified by the Kolmogorov-Smirnov test for goodness of fit in Table 1.

3.2. Experimental design

Both English and Russian experiments drew on the design proposed by Hay (2001). 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 grants 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

Table 2 Participants' matrix for each word pair.

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

respective language, and each set of participants was constructed to conform to the matrix in Table 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.

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:

$$\binom{5}{2} \textit{word combinations} \cdot 8 \textit{ prefixes} \cdot 24 \textit{ subjects} = 1{,}920.$$

3.3. Results

Each word pair in my data was evaluated by 24 different people, 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 $Y \sim Binomial(n, \theta)$. 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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 4, and the Russian results are in Figure 5. The means and highest-density intervals of θs are provided in Table 3. The results for the individual construction pairings can be found in Table 4 (English) and Table 5 (Russian).

Some remarks on notation include the following. (i) Tables 4 and 5 should be read by rows. For example, 0.61 in cell LH–HH of Table 4 means that 61% of participants judged LH words to be more complex compared with HH words. (ii) 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. (iii) Finally, PA stands for the pseudo-affixed type.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

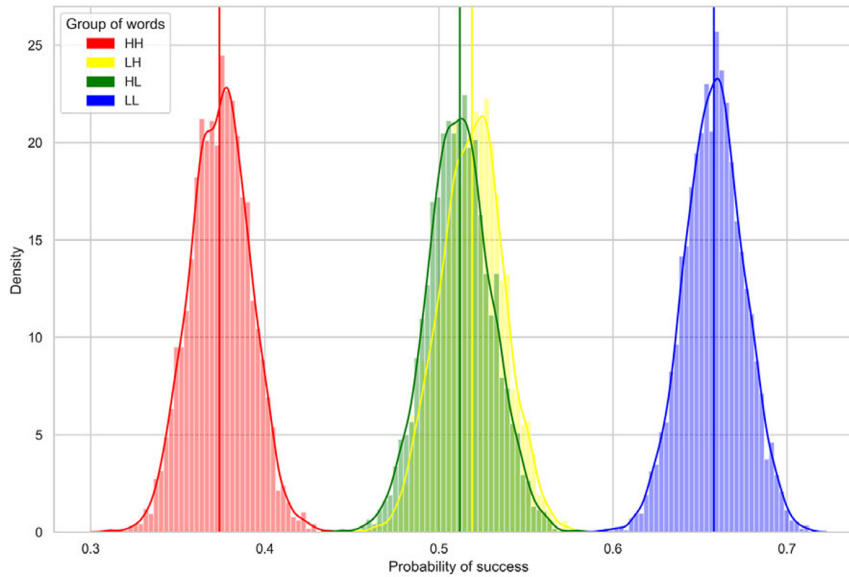


Figure 4 (Colour online) Posterior distributions of the English experiment results.

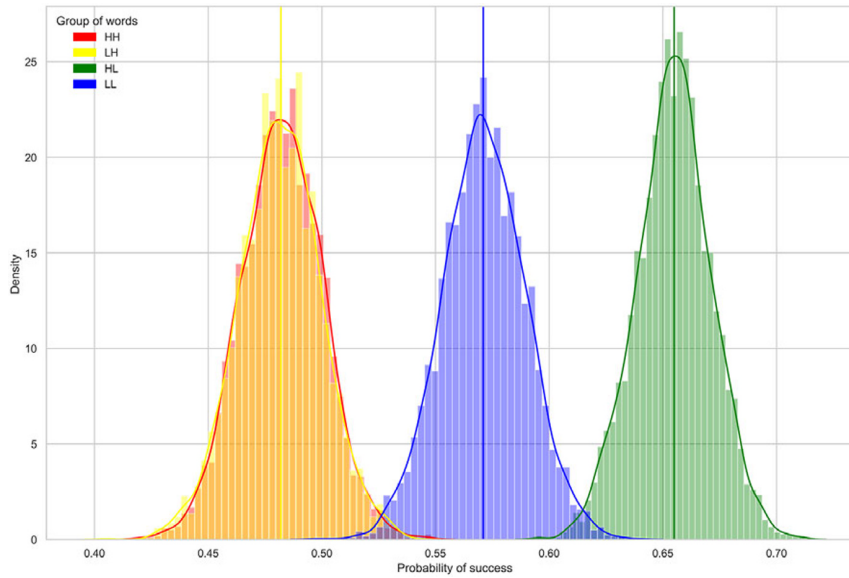


Figure 5 (Colour online) Posterior distributions of the Russian experiment results.

3.4. 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

Table 3 Means and highest-density intervals of θs (English and Russian).

	Probability of success		
Construction type	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 4 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 5 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

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 4), suggesting that under these experimental conditions, participants exhibited no

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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 4, (i) no construction type has a greater proportion of successes than LL, (ii) the LH and HL contest ended in a tie, and (iii) 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 words merged with LL words and HL words with HH words (HH ⋃ HL < LH ⋃ LL).

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 while 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. With LH words, the model takes the following form: $Meaning_{ITEM} = X_{AFFIX} + Meaning_{BASE}$. 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 ($X_{AFFIX} = Meaning_{ITEM} - Meaning_{BASE}$). By contrast, in HL words, $Meaning_{ITEM} = Meaning_{AFFIX} + X_{BASE}$, and the participants had to evaluate the contribution of the base while holding the meaning of the affix fixed ($X_{BASE} = Meaning_{ITEM} - Meaning_{AFFIX}$).

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).

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 & 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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.

I used identical Continuous-Bag-of-Words 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 (Bojanowski et al. 2017). 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 6.

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

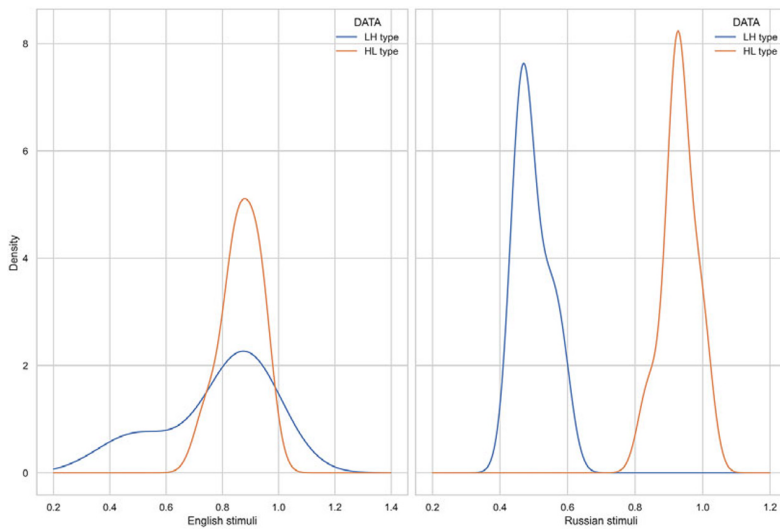


Figure 6 (Colour online) Densities of the cosine distances for English (left panel) and Russian (right panel) stimuli; LH and HL construction types.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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 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 & Carroll 2003).

4. Study 2: Disentangling parsability and compositionality

4.1. A probabilistic model of complex words' perceived complexity

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 3 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 observation can be made 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 predict a most likely winner in the complexity assessment contest for each of the 17 pairs of words in Hay's data, drawing on the evidence obtained during my English experiment. The model, a fragment of which is visualised in Figure 7, is a Bayesian network with three types of nodes: (i) 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-*, and *un-*) affect complexity judgements; (ii) five construction type nodes, that is, nodes that encode how four construction types and one pseudo-affixed type (*LH*, *HH*, *HL*, *LL*, and *PA*) affect complexity judgements; and (iii) 110 contest nodes, that is, nodes that

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

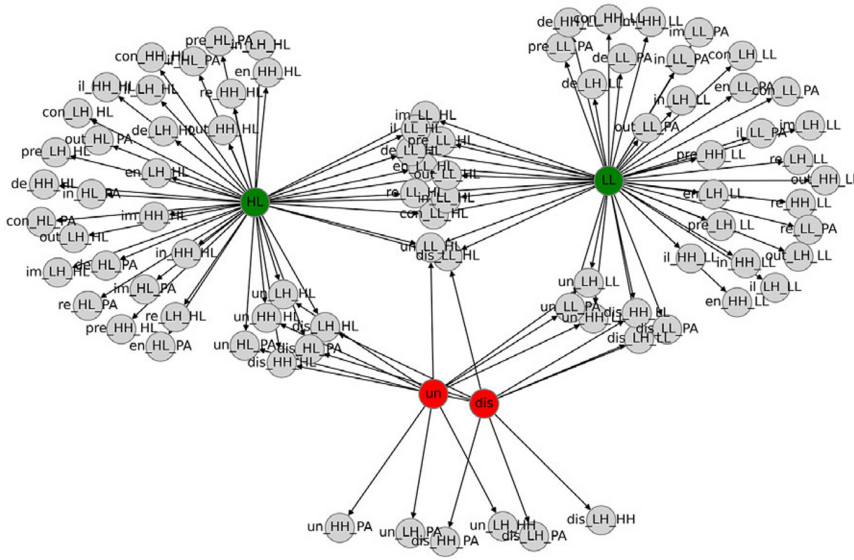


Figure 7 (Colour online) Fragment of the Bayesian network for predicting the outcomes of words' complexity contests.

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.).

Despite its fairly complicated global structure, on the local level, this Bayesian network is a very simple, state-observational model (Koller & 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: (i) 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 (ii) 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 & 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).

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

4.2. Partial replication of Hay's experiment (2001)

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 × 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 2. 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, although 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. Although 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.

4.3. Semantic vector space of complex words

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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

example, the following set of operations will be needed: (i) subtract the vector of the filler element (*canny* in *uncanny* and *un* in *uncommon*) from the vector of the complex word $\vec{W}$, (ii) check whether the subtrahend vector $\vec{S}$ encodes something meaningful and relatable to the meaning of the derivation, and (iii) check whether the difference vector $\vec{D}$ encodes something meaningful and relatable to the meaning of the derivation. Operation (i) is straightforward. Operations (ii) and (iii) may be performed by finding the nearest neighbours of the vectors $\vec{S}$ and $\vec{D}$ and calculating the average cosine similarity of these neighbours' embeddings to the vector representation of the complex word $\vec{W}$.

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 & Polguère 2018).

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 because 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: (i) obtain the vectors $\vec{E}_1$ and $\vec{E}_2$ of the first and second elements of the complex word $\vec{W}$, (ii) check whether the addend vector $\vec{E}_1$ encodes something meaningful and relatable to the meaning of the derivation, and (iii) check whether the addend vector $\vec{E}_2$ encodes something meaningful and relatable to the meaning of the derivation. Again, operations (ii) and (iii) may be performed by finding the nearest neighbours of the vectors $\vec{E}_1$ and $\vec{E}_2$ and calculating the average cosine similarity of these neighbours' embeddings to the vector representation of the complex word $\vec{W}$.

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, although slightly above the conventional significance level ($M_{\text{Group A}} = 0.72$, $M_{\text{Group B}} = 0.60$, $t = 1.87$, $p = 0.07$). 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.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

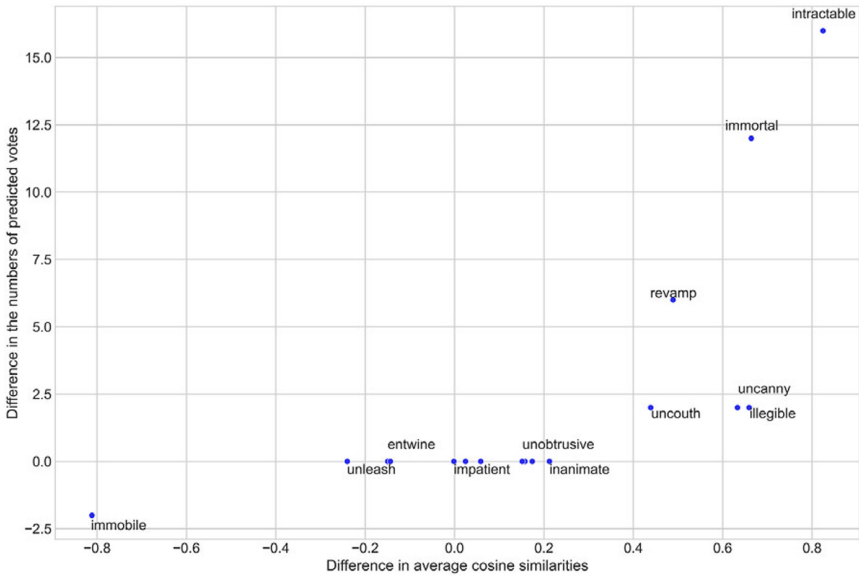


Figure 8 (Colour online) Differences (group A – group B) in cosine similarities and predicted votes.

In Figure 8, the dots represent the complex words from group A (they are labelled selectively to avoid overplotting), the X-axis coordinates correspond to the differences between these words' average cosine similarities and those of their counterparts from group B, and the Y-axis 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.

The choice and pairing of 34 prefixed stimuli in this dataset were such that, given the probabilities of success presented in Table 3, 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 6). 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 the 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.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

Table 6 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	0.00	0.027
Immoral	HH	6	9	0.00	0.043
Retool	HH	9	12	0.00	0.002

4.4. Two types of complexity and two models of meaning processing

The presentation of the examples in Table 6 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, although 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. We 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 $\vec{D}$: *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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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*,

$$P(\text{affix} \mid \text{base})_{Büchlein} = 47{,}121/539{,}800 = 0.08$$
$$P(\text{base} \mid \text{affix})_{Büchlein} = 47{,}121/721{,}543 = 0.06$$
$$\varepsilon_{Büchlein} = \log(0.08/0.06) = 0.28,$$

and for *Bücherei*,

$$P(\text{affix} \mid \text{base})_{Bücherei} = 67{,}174/539{,}800 = 0.12$$
$$P(\text{base} \mid \text{affix})_{Bücherei} = 67{,}174/3{,}329{,}967 = 0.02$$
$$\varepsilon_{Bücherei} = \log(0.12/0.02) = 1.79.$$

Given that $0 < \varepsilon_{Büchlein} < 1 < \varepsilon_{Bücherei}$, 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 (base *Buch*) serves as a filler for the construction's empty slot.

5. Study 3: Bringing productivity and parsability together

5.1. 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 & Baayen (2002). As a way to evaluate affixes' productivity, they used Baayen's hapax-based measure (Baayen 1991, 1992, 1994, 2009; Baayen & Lieber 1991; Baayen & 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 & Baayen 2003). Mathematically, parsing ratios are defined as the proportions of forms (types or tokens) that fall above the so-called parsing line given by the following equation: log(base frequency) = 3.76 + .76 * log(derivation frequency) (Hay & Baayen 2002).

Using this set of measures, Hay and Baayen found i) a significant inverse relationship between token frequency and the proportion of tokens that are parsed and ii) a significant

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

positive relationship between the proportion of tokens that are parsed and Baayen's productivity measure. Based on these results, authors claimed that i) 'the more often you encounter an affix, (…) the less productive that affix is likely to be' and that ii) 'the more often we encounter an affix (…), the less likely we are to parse words containing it' (Hay & 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 as 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, although easily analysable, can be completely disregarded by speakers 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.

5.2. 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) is 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 the 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.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

As for the linguistic productivity of a prefix, I did not want to use Baayen's hapax-based measure, because, as has been pointed out in the literature, this measure is ill-suited for the comparison of affixes with different token numbers (Gaeta & Ricca 2006; Pustylnikov & Schneider-Wiejowski 2010; cf. Bauer 2001). 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 used the algorithm suggested by Monakhov (2023b) and evaluated the morphological productivity of the prefixes in my data as their probability to combine with a random base.

The productivity values for English and Russian prefixes to be found in Table 7 (alongside parsability ratios) stand to reason. As a sanity check for the English data, I did the following. First, 995 random content words (nouns, verbs, and adjectives) without

Table 7. 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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

prefixes were sampled from the *ententen18* internet corpus, their frequencies ranging from 48,421,599 to 54. Each of 24 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. The proportions of non-zero hits among all requests were found to be almost perfectly correlated with the probabilistic productivity measures of the respective prefixes in Table 7 ($r = 0.94, p < 0.001$). Therefore, it appears that the estimated probabilities are not too far off the mark.

5.3. Analysis of English results

The results presented in Table 7 for the English data are plotted in Figure 9. The observable distribution of dots here has a characteristic U shape and is reasonably well modelled by a polynomial regression with two terms. This suggests that parsability ratio, calculated as I propose in this paper, 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 or 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, because 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 9. 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

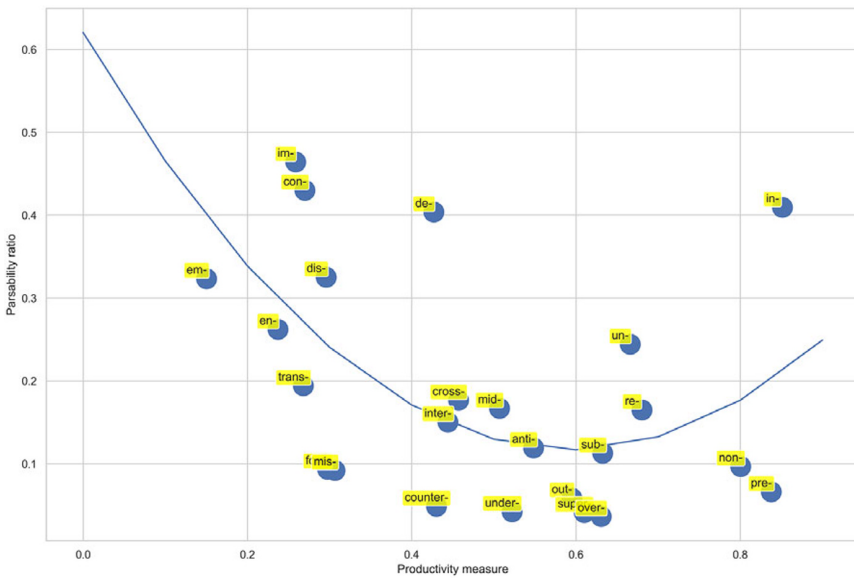


Figure 9 (Colour online) English prefixes' parsability and productivity.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

Table 8. Regression model summary (English prefixes).

Term	5% CI	Coefficient	95% CI	SE	*p*
Intercept	−2.12	−1.60	−1.08	0.25	< 0.01
Parsability	−0.68	−0.40	−0.12	0.13	< 0.01
Preposition	0.01	0.08	0.14	0.02	0.01
Interaction	0.005	0.03	0.05	0.01	0.02

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

hypothesise that the third variable of interest is the frequency of respective prepositions or particles.

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 8). The obtained coefficients show that for the prefixes that 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.

5.4. Analysis of Russian results

The Russian language provides many more possibilities for this type of analysis. Of the 27 prefixes in my data, 17 not only are 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 7 are visualised in Figure 10. The picture, overall, bears a remarkable resemblance to the U-shape distribution of English prefixes observed in Figure 9. Again, this distribution can be reasonably well approximated by a polynomial regression with two terms.

More telling is the distribution of prefixes if one takes into account their prepositional or non-prepositional natures. For non-prepositional 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.

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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

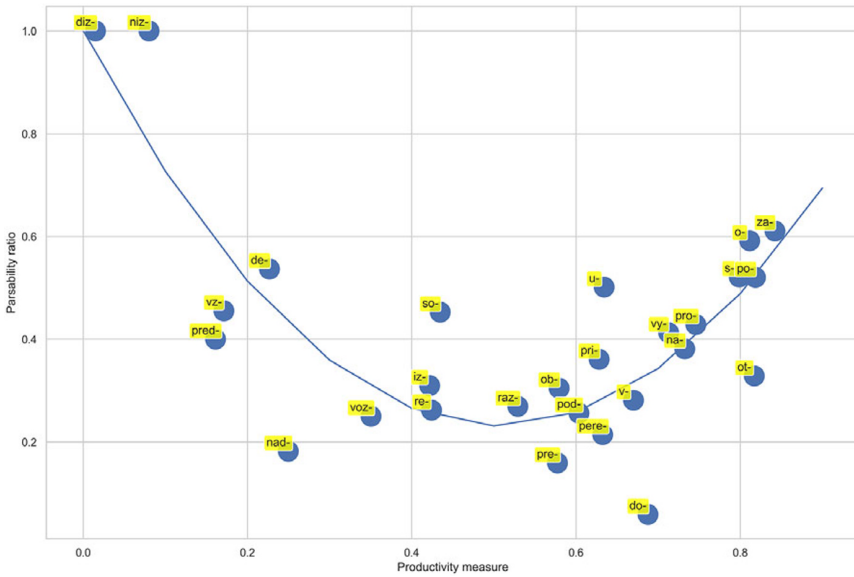


Figure 10 (Colour online) Russian prefixes' parsability and productivity.

frequency of the respective preposition – explains a significant amount of the total variation (Table 9).

Thus, Russian prefixes, corresponding to low-frequency prepositions, behave exactly like non-prepositional prefixes: the lower the parsability ratio, the greater the linguistic productivity. On the other hand, those prefixes whose prepositional counterparts are 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 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, yet its linguistic productivity is 0.81.

I would argue that the difference here is because the corresponding preposition *o* is 6.75 times more frequent than the preposition *pred.* Thus, even though one would expect the

Table 9 Regression model summary (Russian prefixes).

Term	5% CI	Coefficient	95% CI	SE	*p*
Intercept	−3.63	−2.95	−2.27	0.32	< 0.01
Parsability	−2.18	−1.59	−0.99	0.28	< 0.01
Preposition	0.09	0.14	0.19	0.02	< 0.01
Interaction	0.05	0.09	0.14	0.02	< 0.01

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

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

prefix *o-* to be unproductive given its high parsability ratio, the presence of a free element coinciding with it in form and partially overlapping in meaning may facilitate the production of new items.

6. 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 words 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 speakers 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 and another for the base. The words that are characterised by a greater discrepancy between transitional probabilities from affix

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

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. Although 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; thus, this issue requires further investigation.

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/S0022226723000385.

Data availability statement. All data necessary to replicate the study's findings are publicly available at: https://doi.org/10.5281/zenodo.7883848.

References

Baayen, R. Harald. 1991. Quantitative aspects of morphological productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology* 1991, 109–150. Dordrecht: Kluwer Academic Publishers.

Baayen, R. Harald. 1992. On frequency, transparency and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology* 1992, 181–208. Dordrecht: Kluwer.

Baayen, R. Harald. 1994. Productivity in language production. *Language and Cognitive Processes* 9, 447–469.

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kyto (eds.), *Corpus linguistics. An international handbook*, 900–919. Berlin: Mouton De Gruyter.

Baayen, R. Harald & Rochelle Lieber. 1991. Productivity and English word-formation: A corpus-based study. *Linguistics* 29, 801–843.

Baayen, R. Harald & Antoinette Renouf. 1996. Chronicling *The Times*: Productive lexical innovations in an English newspaper. *Language* 72, 69–96.

Bannard, Colin. 2005. Learning about the meaning of verb–particle constructions from corpora. *Computer Speech and Language* 19, 467–478.

Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.

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

Ben Hedia, Sonia & Ingo Plag. 2017. Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics* 62, 34–49.

Berg, Kristian. 2013. Graphemic alternations in English as a reflex of morphological structure. *Morphology* 23: 387–408.

Biskup, Petr. 2019. *Prepositions, case and verbal prefixes: The case of Slavic*. Amsterdam: John Benjamins.

Blom, Corrien. 2005. *Complex predicates in Dutch*. Utrecht: LOT.

Bojanowski, Piotr, Edouard Grave, Armand Joulin & Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5, 135–146.

Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6, 213–234.

Booij, Geert. 2010. Construction morphology. *Language and Linguistics Compass* 4/7, 543–555.

Bott, Stefan & Sabine Schulte im Walde. 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)*, 509–516.

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

Clahsen, Harald. 1999. Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Studies* 22, 991–1060.

Cole, P., C. Beauvillain & J. Segui. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language* 28: 1–13.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

Cordeiro, Silvio, Aline Villavicencio, Marco Idiart & Carlos Ramisch. 2019. Unsupervised compositionality prediction of nominal compounds. *Computational Linguistics* 45, 1–57.
Croft, William. 2001. *Radical construction grammar. syntactic theory in typological perspective*. Oxford: Oxford University Press.
Culicover, Peter & Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.
De Jong, Nivja H., Robert Schreuder & R. Harald Baayen. 2000. The morphological family size effect and morphology. *Language and Cognitive Processes* 15, 329–365.
Deaver, Guinevere J. 2013. *The effects of frequency on dual-route versus single-route processing of morphologically complex terms: A usage-based experiment.* Ph.D. dissertation, Brigham Young University.
Diessel, Holger. 2019. *The grammar network. How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.
Dressler, Wolfgang U. 2005. Word-formation in natural morphology. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 267–284. Berlin: Springer.
Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
Firth, J. R. 1957. A synopsis of linguistic theory 1930–1955. In J. R. Firth (ed.), *Studies in linguistic analysis*, 1–32. Oxford: Blackwell.
Frauenfelder, Ulrich Hans & Robert Schreuder. 1992. Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In Booij & Marle (eds.), 165–183.
Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44, 57–89.
Goldberg, Adele. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.
Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis. A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9, 97–129.
Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39, 1041–1070.
Hay, Jennifer. 2002. From speech perception to morphology: Affix-ordering revisited. *Language* 78, 527–555.
Hay, Jennifer. 2003. *Causes and consequences of word structure*. London: Routledge.
Hay, Jennifer. 2007. The phonetics of *un*. In Judith Munat (ed.), *Studies in functional and structural linguistics*, vol. 58: *Lexical creativity, texts and contexts*, 39–57. Amsterdam: John Benjamins.
Hay, Jennifer & R. Harald Baayen. 2002. Parsing and productivity. In Geert Booij and Jaap van Marle (eds.), *Yearbook of morphology* 2001, 203–236. Dordrecht: Kluwer.
Hay, Jennifer & R. Harald Baayen. 2003. Phonotactics, parsing and productivity. *Rivista di Linguistica* 15.1, 99–130.
Hay, Jennifer & Ingo Plag. 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.
Hoffmann, Thomas & Graeme Trousdale. 2013. *The Oxford handbook of construction grammar*. Oxford: Oxford University Press.
Jackendoff, Ray. 2002. English particle constructions, the lexicon, and the autonomy of syntax. In Dehé et al. (eds.), 67–94.
Jackendoff, Ray. 2008. Construction after construction and its theoretical challenge. *Language* 84, 8–28.
Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The TenTen corpus family. In Andres Hardie & Robbie Love (eds.), *Corpus Linguistics 2013: Abstract book*, 125–127. Lancaster: UCREL.
Kiparsky, Paul. 1997. Remarks on denominal verbs. In: Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Complex predicates*, 473–499. Stanford, CA: CSLI Publications.
Koller, Daphne & Nir Friedman. 2009. *Probabilistic graphical models. Principles and techniques*. Cambridge, MA: MIT Press.
Langacker, Ronald. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*, vol. I. Stanford: Stanford University Press.
Larsen, Darrell. 2014. *Particles and particle-verb constructions in English and other Germanic languages.* Ph.D. dissertation, University of Delaware.
Lewis, Gwyneth, Olla Solomyak & Alec Marantz. 2011. The neural basis of obligatory decomposition of suffixed words. *Brain and Language* 118, 118–127.
Lindquist, Hans. 2009. *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.
Lohde, Michael. 2006. *Wortbildung des modernen Deutschen*. Tübingen: Narr.
Manova, Stela. 2010. Suffix combinations in Bulgarian: Parsability and hierarchy-based ordering. *Morphology* 20, 267–296.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

Mason, Oliver. 2000. Parameters of collocation: The word in the centre of gravity. In J. M. Kirk (ed.), *Corpora galore. analyses and techniques in describing English*, 267–280. Amsterdam/Atlanta, GA: Rodopi.

McCarthy, Diana, Bill Keller & John Carroll. 2003. Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, 73–80. Sapporo, Japan: Association for Computational Linguistics.

McIntyre, Andrew. 2001. Argument blockages induced by verb particles in English and German: event modification and secondary predication In Nicole Dehé & Anja Wanner (eds.), *Structural aspects of semantically complex verbs*, 131–164. Berlin: Peter Lang.

McIntyre, Andrew. 2002. Idiosyncrasy in particle verbs. In Nicole Dehé, Ray Jackendoff, Andrew McIntyre & Silke Urban (eds.), *Verb-particle explorations*, 95–118. Berlin: Mouton de Gruyter.

McIntyre, Andrew. (2007). Particle verbs and argument structure. *Language and Linguistics Compass* 1/4, 350–367.

McIntyre, Andrew. (2015). Denominal verbs:A overview. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe*, 434–450. Berlin: Mouton de Gruyter.

Mel'čuk, Igor & Alain Polguère. 2018. Theory and practice of lexicographic definition. *Journal of Cognitive Science* 19, 417–470.

Monakhov, Sergei. 2021. Russian prefixed verbs as constructional schemas. *Russian Linguistics* 45, 45–73.

Monakhov, Sergei. 2023a. How complex verbs acquire their idiosyncratic meanings. *Language and Speech.* doi.org/10.1177/00238309231199994. Published online by Sage Publications, 29 September 2023.

Monakhov, Sergei. 2023b. Probabilistic method of measuring linguistic productivity. *arXiv:2308.12643*.

Neapolitan, Richard E. 2004. *Learning Bayesian networks*. New York: Pearson Prentice Hall.

Pearl, Judea, Madelyn Glymour & Nicholas Jewell. 2016. *Causal inference in statistics: A primer*. Wiley.

Pelucchi, Bruna, Jessica F. Hay & Jenny R. Saffran. 2009. Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition* 113, 244–247.

Petre, Peter & Hubert Cuyckens. 2008. Bedusted, yet not beheaded: The role of be-'s constructional properties in its conservation. In Alex Bergs & Gabriele Diewald (eds.), *Constructions and language change*, 133–169. Berlin: Mouton de Gruyter.

Pinker, Steven & Michael T. Ullman. 2003. Beyond one model per phenomenon. *Trends in Cognitive Sciences* 7, 108–109.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Plag, Ingo. 2021. Productivity. In Bas Aarts, April McMahon & Lars Hinrichs (eds.), *The handbook of English linguistics*, 2nd edn., 483–499. Wiley.

Plag, Ingo & Sonia Ben Hedia. 2018. The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin & Esme Winter-Froemel (eds.), *Expanding the lexicon: Linguistic innovation, morphological productivity, and ludicity*, 93–116. Berlin: Mouton de Gruyter.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America* 118, 2561–2569.

Pustylnikov, Olga & Karina Schneider-Wiejowski. 2010. Measuring morphological productivity. *Studies in Quantitative Linguistics* 5, 1–9.

Pycha, Anne. 2013. Mechanisms for remembering roots versus affixes in complex words. *Journal of the Acoustical Society of America* 19, 060201.

Saldana, Carmen, Yohei Oseki & J. Culbertson. 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.

Schuppler, Barbara, Wim A. van Dommelen, Jacques Koreman & Mirjam Ernestus. 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, 595–607.

Schreuder, Robert & R. Harald Baayen. 1995. Modelling morphological processing. In Laurie Beth Feldman (ed.), *Morphological aspects of language processing*, 131–156. Hillsdale, NJ: Erlbaum.

Silva, Renita & Harald Clahsen. 2008. Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition* 11, 245–260.

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stein, Simon D. & Ingo Plag. 2022. How relative frequency and prosodic structure affect the acoustic duration of English derivatives. *Laboratory Phonology* 13.1.

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press

Stiebels, Barbara. 1996. *Lexikalische Argumente und Adjunkte: Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Berlin/Boston: Akademie Verlag.
Taylor, John. 2012. *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.
Tikhonov, Alexander N. 1985. *Slovoobrazovatel'ny Slovar Russkogo Jazyka*. [*Word-formation dictionary of Russian language*]. Moscow: Russkij Jazyk.
Ullman, Michael T. 2004. Contributions of neural memory circuits to language: The declarative/procedural model. *Cognition* 92, 231–270.
Varvara, Rossella, Gabriella Lapesa & Sebastian Padó. 2021. Grounding semantic transparency in context. *Morphology* 31, 409–446.
Wiegand, Herbert Ernst. 1992. Elements of a theory towards a so-called lexicographic definition. *Lexicographica* 8, 175–289.
Zee, Tim, Louis Ten Bosch, Ingo Plag & Mirjam Ernestus. 2021. Paradigmatic relations interact during the production of complex words: Evidence from variable plurals in Dutch. *Frontiers in Psychology* 12, 720017.
Zeller, Jochen. 2001. *Particle verbs and local domains*. Amsterdam: John Benjamins.
Zimmerer, Frank, Mathias Scharinger & Henning Reetz. 2014. Phonological and morphological constraints on German /t/-deletions. *Journal of Phonetics* 45, 64–75.

Cite this article: Monakhov, Sergei. 2026. "Parsability revisited and reassessed." *Journal of Linguistics*. 62, 79–111. doi:10.1017/S0022226723000385

https://doi.org/10.1017/S0022226723000385 Published online by Cambridge University Press