Francesca Strik Lievers*, Marianna Bolognesi and Bodo Winter

The linguistic dimensions of concrete and abstract concepts: lexical category, morphological structure, countability, and etymology

Abstract: The distinction between abstract and concrete concepts is fundamental to cognitive linguistics and cognitive science. This distinction is commonly operationalized through concreteness ratings based on the aggregated judgments of many people. What is often overlooked in experimental studies using this operationalization is that ratings are attributed to *words*, not to concepts directly. In this paper we explore the relationship between the linguistic properties of English words and conceptual abstractness/concreteness. Based on hypotheses stated in the existing linguistic literature we select a set of variables (part of speech, morphological structure, countability, etymology) and verify whether they are statistically associated with concreteness ratings. We show that English nouns are rated as more concrete compared to other parts of speech, but mass nouns are rated as less concrete than count nouns. Furthermore, a more complex morphological structure is associated with abstractness, and as for etymology, French- and Latin-derived words are more abstract than words of other origin. This shows that linguistic properties of words are indeed associated with the degree of concreteness that we attribute to the underlying concepts, and we discuss the implications that these findings have for linguistic theory and for empirical investigations in the cognitive sciences.

Keywords: concreteness; count nouns; etymology; mass nouns; part of speech

*Corresponding author: Francesca Strik Lievers, Dipartimento di Lingue e Culture Moderne, Università di Genova, Piazza Santa Sabina 2, 16124 Genova, Italy,
E-mail: francesca.striklievers@unige.it
Marianna Bolognesi, Dipartimento di Lingue, Letterature e Culture Moderne, Università di Bologna, Bologna, Italy, E-mail: m.bolognesi@unibo.it
Bodo Winter, Department of English Language and Applied Linguistics, University Birmingham, Birmingham, West Midlands, UK, E-mail: b.winter@bham.ac.uk

1 Introduction

Many people share the intuition that some concepts are more concrete and others are more abstract. For example, most people will agree on judging the concepts of HOPE, AFTERTHOUGHT, and HATRED as more abstract than those of CARROT, DOG, and HAMMER. The distinction between abstract and concrete concepts is one of the most fundamental ones in cognitive science, with a large body of literature having shown that abstract concepts are processed, learned, and memorized differently from concrete ones (Binder et al. 2005; Bolognesi and Steen 2019; Conca et al. 2021; Gao et al. 2019; Pexman et al. 2007; Reilly et al. 2017; Villani et al. 2019; Wang et al. 2010).

Many studies on this topic operationalize the concrete/abstract distinction via ratings collected from large numbers of native speakers. In one of the most influential large-scale rating studies for English, Brysbaert et al. (2014) asked participants to rate words on a scale from 1 (least concrete) to 5 (maximally concrete). Results showed that speakers judged the word *carrot* to have an average concreteness score of 5.0 (= maximally concrete), in contrast to the word *hope*, which was rated to be much less concrete, with an average score of only 1.19 (= very abstract). Such concreteness rating studies have been conducted in a number of languages, including Italian (Montefinese et al. 2014), Spanish (Guasch et al. 2016), Portuguese (Soares et al. 2017), Chinese (Yao et al. 2017), French (Bonin et al. 2018), and Croatian (Ćoso et al. 2019). Across these studies, concreteness is typically defined as degree of accessibility to the senses, i.e., whether a concept is something that can be seen, heard, felt, tasted, or smelled (but see Dunn 2015; see also Löhr 2021 for a discussion of issues with this type of definition). Instructions are usually provided to participants along the following lines: "any word that refers to objects, materials or persons should receive a high concreteness rating [and] any word that refers to an abstract concept that cannot be experienced by the senses should receive a low concreteness rating" (Spreen and Schulz 1966: 460).

Although rating studies are meant to operationalize the distinction between concrete and abstract concepts, it is important to keep in mind that participants are ultimately rating words that appear on the screen, not concepts directly (Löhr 2021). The fact that ratings are performed on words as stand-in for concepts suggests that the linguistic properties of these words may matter. This invites potential methodological issues for rating studies, as the properties of words — for example, their morphological structure, their part of speech, etc. — may "intrude" into the rating that is ultimately thought to represent the degree of concreteness of the corresponding concept.

In this paper, we demonstrate that linguistic variables are indeed statistically associated with the concreteness ratings, in a manner that can be predicted based

on linguistic theory. This has important methodological ramifications for any empirical investigation that relies on concreteness ratings, but it also has the potential to uncover new facts about conceptual and linguistic structure. Specifically, we use concreteness ratings as a novel way of testing cognitive linguistic approaches to linguistic categories such as part of speech and the count/mass distinction. Our results provide bottom-up evidence for the idea that these categories are, at least to some extent, justifiably characterized in notional terms. This also demonstrates the utility of concreteness ratings for addressing issues that are relevant to cognitive linguistic theorizing.

2 Linguistic variables and concreteness: theory-based hypotheses

In this study, we focus on English, and we investigate the following linguistic variables: part of speech, morphological structure, countability, and etymology. We selected these variables because they all allow making clear predictions for differences in the degree of concreteness based on existing linguistic theory, with special consideration of cognitive linguistic approaches. The following sections discuss each variable in turn, including the specific datasets that we use to operationalize each variable.

2.1 Lexical category (part of speech)

Most studies that test the behavioral, cognitive, and neuroscientific correlates of the abstract/concrete distinction only take into account nouns (see Vonk et al. 2019: 602–603 for discussion). However, different lexical categories, or "parts of speech", such as nouns, verbs, and adjectives, can be expected to differ in their degree of concreteness based on linguistic grounds.

Parts of speech can be defined according to a multitude of linguistic criteria, which have been emphasized to varying degrees by different scholars and theoretical orientations. As already clearly outlined by Hermann Paul, "[t]he usual division [into parts of speech] has been effected by the consideration of three points: the *meaning* of a word, taken by itself, its *function in the sentence*, and its *behaviour in regard to inflexion and word-formation*" (Paul 1891: 403; emphasis in original); in other words, the main criteria are of semantic, syntactic, and morphological nature (Baker 2003; Baker and Croft 2017; see Croft 1991; Rauh 2010 for additional criteria and a discussion of different theoretical perspectives). Our

data analysis reported below uses part-of-speech tags that are based on formal criteria (e.g., the function of words in a sentence), but to generate hypotheses with respect to the concreteness of different parts of speech, we need to focus on the semantic features of different lexical categories, which have been particularly emphasized within cognitive linguistics (see, e.g., Langacker 1987a, strongly advocating for a notional basis of grammatical categories).

Nouns, verbs, and adjectives, arguably the major categories of content words in English, have traditionally been associated with things, actions, and properties respectively. As is well known, cases of mismatch between semantic type and part of speech abound: for example, it is not difficult to find verbs that do not describe an action (e.g., *to be*, *to understand*) and, conversely, nouns that do refer to an action (e.g., *slap*, *conquest*; see Koptjevskaja-Tamm 2015). In the light of examples such as these, the semantic approach to parts of speech has been reassessed, especially within cognitive linguistics, in terms of prototypicality. From this perspective, parts of speech are seen as prototype-based categories, in which "[t]ypical nouns describe INDIVIDUAL PHYSICAL OBJECTS, typical verbs describe PHYSICAL ACTIONS, and typical adjectives designate PROPERTIES" (Murphy 2010: 144). Under this account, both nouns and verbs prototypically express physical percepts. Given that accessibility to the senses is considered as the main feature of conceptual concreteness (see Section 1), this broad semantic characterization of parts of speech already allows us to hypothesize that nouns and verbs might overall be more concrete compared to adjectives, because physical objects and actions are usually experienced through the senses, while not all properties are. That is, properties may be either perceivable through the senses (e.g., the properties expressed by the adjectives *yellow*, *bitter*) or not perceivable through the senses (e.g., *intelligent*, *free*).

To formulate a more precise hypothesis, and one which also includes other parts of speech, we need to introduce a further distinction: that between punctual and relational concepts. As summarized by Prandi, "[p]unctual concepts classify *individuals* and give access to instances of masses, while relational concepts give access to *properties* of individuals and instances of masses, as well as to *processes* involving individuals and instances of masses" (Prandi 2004: 122; our emphasis). This ontological distinction, which dates back at least to Aristotle (Prandi 2004: 123), is connected to the distinction between the prototypical representatives of the major word classes, with punctual concepts (e.g., TREE) typically being the meanings of nouns, and relational concepts (e.g., BLUE, GIVE) typically being the meanings of verbs and adjectives (Prandi 2004: 124). Within cognitive linguistics, a comparable distinction was introduced in Langacker (1987a), where it is proposed that linguistic predications are either nominal or relational, with nominal predications

corresponding to nouns and relational predications corresponding to verbs, adjectives and adverbs.

The privileged association of nouns with punctual concepts (and nominal predications) on the one hand, and that of adjectives and verbs with relational concepts (and relational predications) on the other, allows us to make more specific predictions concerning the correspondence between parts of speech and degrees of concreteness. Sapir (1921: 102) already observed that there is a close connection between relational concepts and abstractness; more recently, Asmuth and Gentner (2017: 2016) state that "overall it is very likely that a relational word will also be abstract", and Borghi and Binkofski (2014: 2) posit relationality as one of the defining features of abstract concepts, since the latter "evoke properties and relations more than objects and events". The distinction between punctual and relational concepts therefore allows us to predict not only that adjectives and verbs will overall be rather abstract, but also that verbs will be more abstract than nouns, due to their relational nature. The idea that between nouns and verbs, nouns are relatively more concrete than verbs was also suggested by Givón (1979), based on the observation that nouns typically refer to entities that exist in time and space, while verbs would usually not be anchored in space. To sum up so far, going from more to less concrete we will have: nouns > verbs > adjectives.

With respect to the other parts of speech, adverbs are notoriously difficult to define as a category (see discussion in Pittner et al. 2015; Rauh 2015), especially because of their close connection with adjectives, which has even led some scholars to suggest that English adverbs and adjectives form a single class (Giegerich 2012). What is worth noting here is that not only are adverbs often morphologically related to adjectives (such as via the deadjectival suffix -ly) but, like adjectives, they also typically express relational concepts. As mentioned above, Langacker (1987a), for instance, characterizes both adverbs and adjectives as relational categories that differ from verbs with respect to being atemporal. We can therefore predict that adverbs, like adjectives, will overall be rated as being rather abstract.

Finally, function words, such as prepositions and conjunctions, are likely to lie on the most abstract end of the scale, because their prime role is coding relations. Some function words are moreover the result of grammaticalization, which is commonly held to involve a process of abstraction from more concrete meanings (Heine and Kuteva 2002; Hopper and Traugott 2003; Lehmann 2015; Traugott 1982; Žirmunskij 1966). Overall, our hypothesis is therefore that, starting from the most concrete part of speech, we will have: nouns > verbs > adjectives and adverbs > function words.

It is worth stressing that our predictions concerning parts of speech, like those that we will make in the following sections for the other linguistic variables, are probabilistic. That is, we are concerned here with the relative degree to which

particular values of a linguistic variable tend to associate with particular degrees of concreteness. Of course, there are, for instance, adjectives that may be considered rather concrete, such as *blond* and *wet* (which express properties perceived through our senses), and abstract nouns, such as *optimism* and *ugliness*. Interestingly, de-adjectival quality nouns like *ugliness* were referred to in the medieval philosophical tradition as *abstracta* in opposition to the adjectives they derive from, called *concreta* (Rainer 2015: 1269).

In the analyses below, we use corpus-derived parts-of-speech tags from the SUBTLEX corpus of movie subtitles (Brysbaert et al. 2012). The SUBTLEX-derived part-of-speech tags come with information about how frequent each word form is used as a particular part of speech. For example, the word *furl* is indicated to be a verb 67% of the time in the SUBTLEX corpus, and 33% it is a noun. In the following analysis, we used the dominant parts of speech for each word form. This is an important detail of our analysis since the participants, when they rated the corresponding word forms in the concreteness rating study, presumably thought of the most dominant part of speech when they saw the word in isolation. Importantly, the corpus-derived part-of-speech tags do not take semantics into account, they are exclusively based on a formal characterization of lexical categories in terms of their functioning in a sentence. By correlating these formal distinctions with the concreteness ratings, we assess the extent to which semantic differences correspond to the formal criteria. This not only serves to show that lexical category is an important factor to consider in psycholinguistic studies using concreteness ratings, it also empirically tests the cognitive-linguistic idea that lexical categories differ in their semantics.

Using a corpus of movie subtitles may seem like a strange choice from a methodological standpoint, but the SUBTLEX-US corpus has been argued to emulate spoken language really well and it has been demonstrated that it closely corresponds to behavioral data above and beyond other corpora (Brysbaert and New 2009). In addition, the part-of-speech tags generated from this corpus are one of the most extensive sets of part-of-speech tags that are also readily available for statistical analysis, and are widely used in psycholinguistic research (Brysbaert et al. 2012). The results we obtain here replicate with other part-of-speech tags (e.g., from the English Lexicon Project, Balota et al. 2007), which shows that the choice of SUBTLEX-US as a corpus does not matter for our analysis of part-of-speech.

2.2 Morphological structure

Research on abstract nouns indicates that a useful criterion for identifying them, among all nouns, is based on their morphological features. In English, as observed by Zamparelli (2020: 203) "one could regard as abstract all the nouns derived from

the suffixes *-ness*, *-ity*, *-tion* or *-hood*, *-itude*, *-cy*, *-ment*, *-ship* […], or more generally, all the nouns derived from gradable adjectives". <u>Additional abstract-noun-forming suffixes listed by Plag (1999: 67) are the deverbals *-age* (*steerage*), *-al* (*betrayal*), *-ance* (*annoyance*), *-y* (*enquiry*), and the denominals *-age* (*orphanage*), and *-ism* (*despotism*).</u> Reilly and Kean (2007) demonstrate that nouns with suffixes are rated as more abstract than nouns without suffixes. Here, we investigate whether suffixation may be useful as a signal of abstraction not only for nouns (which, incidentally, should be overall rather concrete, see Section 2.1) but also for words of other parts of speech. To test this hypothesis, we examined the most frequent suffixes of English: if the most frequent suffixes can be classed as abstraction-triggers, then there are chances that suffixed words are overall more abstract than non-suffixed ones.

The MorphoLex database (Sánchez-Gutiérrez et al. 2018) contains information on a set of derivational morphological variables for 68,624 complex words from the English Lexicon Project (Balota et al. 2007). Such variables also include measures of suffix frequency. For our purposes, the relevant measure is morphological family size (Schreuder and Baayen 1997), that is, morphological type frequency defined as the number of word types in which a given morpheme, in our case a given suffix, is a constituent: "For instance, in the example {*attendance*, *pleasance*, *pleasure*, *appearance*}, the suffix *-ance* has a morphological family size of 3 {*attendance*, *pleasance*, *appearance*}, while the root *-pleas-* has a morphological family size of 2 {*pleasure*, *pleasance*}" (Sánchez-Gutiérrez et al. 2018: 1572). Based on this measure, we selected the 20 most frequent suffixes, listed here in Table 1.

The suffixes in Table 1 comprise about 84% of the morphologically complex words in the MorphoLex data. We can notice that the majority of such suffixes form adjectives (*-ly*, *-ic*, *-able*, *-est*, *-ious*, *-ive*, *-less*; e.g., *lovely*, *terrific*, *measurable*, *smartest*, *cautious*, *active*, *hopeless*) and/or adverbs (*-ly*, *-est*; e.g., *honestly*, *soonest*); words that contain such suffixes should therefore be rather abstract, based to the hypothesis outlined in Section 2.1, according to which adjectives and adverbs should overall be more abstract compared to most other parts of speech. The verbalizer *-ize* (e.g., *humanize*) should also form relatively abstract words, again based on the hypothesis in Section 2.1 (that is, verbs should overall have "intermediate" concreteness values). As for nominal suffixes, most are among the typical abstraction-triggering ones listed in works on abstract nouns mentioned above (*-al*, *-ness*, *-ity*, *-ion*, *-ance*); the remaining two form nouns referring to persons, which, on the contrary, are arguably concrete (*-er*, *-ist*, e.g., *plumber*, *artist*).

Since the majority of the most frequent derivational suffixes of English form words referring to rather abstract concepts, we expect morphologically complex words (with suffixes) to be overall more abstract than morphologically simplex words (without suffixes). For nouns only, this result has already been established by

Table 1: The most frequent English derivational[1] suffixes according to morphological family size (from https://github.com/hugomailhot/MorphoLex-en). The third column shows the lexical category (or categories) of words formed by each suffix.

Suffix	Family size	Resultant lexical category
-ly	2,898	A, Adv
-y	2,486	A, N
-er	2,274	N, V
-ion	1,599	N
-al	1,431	A, N
-ness	1,243	N
-ic	1,014	A, N
-ate	937	N, A, V
-able	872	A
-est	714	A, Adv
-ious	617	A
-ity	580	N
-ive	580	A, N
-ant	464	A, N
-ist	462	N
-ize	430	V
-less	368	A, Adv
-ory	356	N, A
-ful	343	A, N
-ance	323	N

Reilly and Kean (2007). Here, we extend their result to more parts of speech (not only considering nouns), and we look at differences between types of suffixes, focusing on the suffixes shown in Table 1. Moreover, we are interested in investigating whether the prediction made above (suffixed words being more abstract) extends to words with more than just one suffix, and whether abstractness grows as a function of the precise number of morphemes, i.e., does abstractness increase monotonically when words become progressively more morphologically complex? It has been shown that word length correlates with abstractness (see Lewis and Frank 2016; and the experimental studies in Reilly et al. 2012, 2017), and hence we want to further assess the extent to which morphological structure and length are independently associated with abstractness.

---

1 Although inflectional suffixes were manually removed by the authors of MorphoLex (Sánchez-Gutiérrez et al. 2018: 1571), the list also includes the superlative suffix *-est*, traditionally considered as inflectional: its status as inflectional or derivational is however discussed in the literature (e.g., Blevins 2006; Fábregas 2014).

2.3 Countability

For the class of nouns, a much researched distinction is that between count and mass nouns (e.g., Fieder et al. 2014; Moltmann 2020). Linguists generally base the count/mass distinction on morphosyntactic criteria (Jespersen 1924). English count nouns, for instance, allow plural marking (*cats*, *hats*) and determination by cardinal numerals (*two cats*, *one hat*), while mass nouns generally don't (**sands*, **one sand*). The count/mass distinction, however, does not characterize nouns, but senses of nouns, that is, a given noun (e.g., *matter*) may be classified as count in one of its senses (e.g., 'a vaguely specified concern', as in *several matters to attend to*) and as mass in another (e.g., 'that which has mass and occupies space', as in *physicists study both the nature of matter and the forces which govern it*) (examples from Kiss et al. 2016: 2810–2811). In addition, a noun in a single sense (e.g., *cake* as a 'baked good') may also be count or mass depending on the context (e.g., *a cake* vs. *some cake*), to the point that, as Langacker observed (1987b: 67), "[g]iven proper circumstances, almost any count noun can be construed as designating a homogeneous, unbounded mass; it may thereby come to function as a mass noun grammatically". This is reflected in Langacker's well-known example *After I ran over the cat with our car, there was cat all over the driveway* (1987b: 67), where the count noun *cat* displays mass noun semantics. This has led some scholars to argue that the count/mass distinction concerns in fact neither lexemes nor senses of lexemes, but ways of using them (see discussion and references in Franzon and Zanini 2019). According to Chierchia (2010: 111), the fact that "some nouns are ambiguous, [and] most can be coerced" appears to be a property that is universally associated with the count/mass distinction, and he labels it "elasticity" (for an interpretation of contextual countability shifts in terms of coercion, see, among others, Michaelis 2005; Pustejovsky 1995).

In addition to the differences in morphosyntactic properties and behaviour, it is also often assumed that the distinction between count and mass nouns corresponds to a semantic and conceptual distinction, the precise nature of which is still much debated (see discussion in Doetjes 2017; Fieder et al. 2014; Kiss et al. 2021; Langacker 1987a; Middleton et al. 2004; Rips and Hespos 2019; Talmy 2000; Wierzbicka 1985). In this regard, several researchers have remarked about the count/mass distinction being also characterized in terms of concreteness. Mass nouns can be more abstract or more concrete (*hope* vs. *sand*), and the same holds for count nouns (*idea* vs. *cat*), so that "[c]ountability has often been considered 'orthogonal' to concreteness in the description of nouns" (Franzon and Zanini 2019: 169). However, the literature suggests that there may be a privileged relation between mass nouns and abstraction.

First, grammars of English and theoretical studies occasionally mention in passing that most mass nouns are abstract (e.g., Gillon 2017: 296–297; Quirk et al. 1972: Ch. 4), a hypothesis for which there also is some corpus-based evidence. For example Katz and Zamparelli (2012) show that among the nouns that are most frequently used in mass contexts, almost all are also abstract. Second, the case of "elastic" nouns is particularly revealing in this respect. Since elasticity concerns both concrete and abstract nouns, concrete nouns (e.g., *cake*) can be used as count (*a cake*) or mass (*some cake*), and abstract nouns (e.g., *hope*) can be used as count (*She had two hopes for her future*) or mass (*She sees some hope for the future*) (on countability within abstract nouns see Husić 2020; Zamparelli 2020). Interestingly for our concerns, it has been argued that concrete nouns are relatively more abstract in their mass uses (*some cake*) than in their count uses (*a cake*), because "the former entail the suppression of the reference to shape, which is a salient property in the representation of entities" (Franzon and Zanini 2019: 167). Finally, if we consider concepts expressed by concrete nouns only, experimental research has shown that count ones (e.g., *apple*) afford direct manual grasp, while mass ones referring to substances and aggregates (e.g., *water*, *sand*) tend to require the intermediation of an instrument, such as a cup or a spoon (De Felice 2015). Since graspability correlates with concreteness (Pexman et al. 2019), this may indicate that concrete mass nouns are relatively more abstract than concrete count nouns. For all these reasons, we expect mass nouns to be, on average, more abstract than count nouns.

In our analyses, we use data from the Bochum English Countability Lexicon (Kiss et al. 2016), a manually annotated database that addresses the issues discussed above, namely, that the count/mass distinction does not apply to nouns but to senses of nouns, and that a noun in a given sense can often be both count and mass. As for the first issue, the database provides countability information for noun-sense pairs rather than for nouns, e.g., $matter_1$ ('a vaguely specified concern') is annotated as count, while $matter_2$ ('that which has mass and occupies space') is annotated as mass. The second issue is addressed by the fact that each noun-sense pair is assigned to one of four distinct classes rather than to count or mass only, i.e., count, mass, both count and mass, neither count nor mass. The database contains ≈11,800 noun-sense pairs that have been assigned to one of the four classes based on the answers that annotators gave to several syntactic and semantic questions about the usage of the noun-sense pair. For example, a noun-sense pair is annotated as count only if the answer to "Can the noun-sense pair in its plural form appear together with *more*?" is "yes" (Kiss et al. 2016: 2811).

2.4 Etymology

Whether conceptual concreteness plays a role in diachronic change or stability of languages' vocabulary is a matter of debate (Monaghan and Roberts 2019: 149). In the specific case of English, however, social and cultural correlates of well-known historical facts allow us to formulate predictions about the degree of concreteness of words originating from two languages in particular: Latin and French. These two languages are major sources of lexical borrowing in English, with Latin and French words forming a large portion of the vocabulary of Modern English (for an overview of external influences on English, see Durkin 2014; Miller 2012).

The borrowing of Latin lexemes reached its quantitative peak during the Renaissance (see Oxford English Dictionary timelines),[2] when Latin, as the international language of science, provided lexical material for new technic, scientific, and philosophical concepts: it is therefore likely that most Latin words that entered the lexicon in this period, and thus a considerable number of Latin loanwords in general, are rather abstract. French, whose words started being borrowed after the Norman Conquest (1066), continued to enjoy great political and cultural prestige for centuries. For example, a disproportionate amount of English legal terms, which are highly abstract, derive from French (*lawyer*, *attorney*, *mortgage*, *defendant*, *culprit*, *jury*, *larceny*, *parole*, *plaintiff*, etc.) because the Normans were in power and ruled the court. Orr (1944: 3) speaks of "the far more abstract or intellectualized vocabulary of French". Textbooks and popular publications, moreover, often mention triplets like *rise*, *mount* and *ascend*, or *go*, *depart* and *exit*, where "[t]he Anglo-Saxon word is typically a neutral one; the French word connotes sophistication; and the Latin or Greek word, learnt from a written text rather than from human contact, is comparatively abstract and conveys a more scientific notion" (Hitchings 2008: 21).

Although the idea that the English vocabulary is stratified in such a clear-cut way is over-simplistic, the contact history with Latin and French leads us to expect that words of Latin and French origin will, on average, be more abstract than the remaining part of the vocabulary, composed of words of Germanic and other origin. Reilly and Kean (2007) already provide some evidence for this, because they showed that nouns of Latinate origin are overall more abstract compared to those of other origin. We will see whether our data confirm their result for more than just nouns. In contrast to the other variables (part of speech, morphological structure, and countability) this prediction is based on extra-linguistic considerations, i.e., contingent facts about the history of English. Taking etymology into account is, however, also important for theoretical reasons: because English mass nouns are disproportionately French- or Latin-derived, our hypotheses about the

---

2 https://www.oed.com/timelines.

abstractness of mass nouns need to control for the fact that we expect these nouns to be more abstract on purely historical grounds. Similarly, the idea that morphological complexity is correlated with abstractness needs to be assessed while controlling for etymology, because Latin words have a tendency to be slightly more morphologically complex, as reported below. Thus, certain hypotheses motivated by linguistic theory need to be assessed while holding etymology constant to make sure that these two factors are not confounded.

To establish the origin of words, in the analyses below we retrieve etymological information from the Oxford English Dictionary.

2.5 Limitations: language-specificity and omitted variables

There were some variables that we excluded from our consideration even though they would seem relevant on the grounds of cognitive linguistic theory. We did so partially to limit the scope of our investigation, but also because predictions for these other variables were not as clear-cut. One of the variables we do not consider here is polysemy. Intuitively, it could be hypothesized that more polysemous words should be more abstract because i) metaphor is a prominent mechanism of semantic extension, and ii) metaphoric extended meanings of polysemous words are known to often be more abstract than the meanings from which they stem (Lakoff and Johnson 1980). For instance, the polysemous word *support* has a concrete meaning that refers to a structure that holds a weight, and a relatively more abstract metaphorical meaning that refers to the help or approval given to a friend, an idea, or an organization.

However, predictions for the relation between polysemy and concreteness ratings are not as clear-cut. First, polysemy also includes relations between senses that are not based on metaphor, and hence that do not clearly follow a concrete > abstract direction, such as metonymy, which often involves relations between concrete senses. For example, the word *dish* is metonymically polysemous between a food sense (*This is a great dish*) and a container sense (*Can you hand me the dish?*), both of which are rather concrete. Second, polysemy does of course not only involve metaphor and metonymy, but also hypernymy (e.g., *cow* 'female' to the more generalized sense *cow* 'of either sex') and hyponymy (e.g., *to drink* 'anything' to the more specialized sense *to drink* 'only alcohol'), both of which can be seen as making opposing predictions with respect to concreteness. Third, even for metaphorical polysemy, predictions are not as clear-cut. Metaphor is generally thought to primarily involve mappings from concrete to abstract meanings, which means that 1) we can expect metaphorically polysemous words to be more abstract (because they have acquired abstract senses), but also 2) since metaphor is thought

to preferentially draw from concrete source domains, metaphorically polysemous words could also be, on average, more concrete.

Additionally, there are methodological reasons for us not to consider polysemy here. In concreteness rating studies such as Brysbaert et al. (2014), ratings are provided for isolated word forms, not for word-meaning pairs or words in context. We therefore do not know which senses of polysemous words have been rated. As observed by Löhr (2021), the same word could plausibly have multiple different concreteness ratings depending on which sense is implied, something that has been empirically addressed by Reijnierse et al. (2019), who show that when the meanings of a word are disambiguated, the elicited concreteness ratings are actually different. Another methodological issue is the fact that it is not clear what the best way of quantitatively operationalizing polysemy at large scale is, especially vis-à-vis the fact that some word senses are more dominant than others (cf. Werkmann Horvat et al. 2021). We therefore leave it for future studies to explore the relation between polysemy and concreteness ratings.

Finally, it is worth highlighting why we focus on one language only. First, the choice of the relevant variables to be investigated is at least in part language-dependent (e.g., one would not consider morphological structure when working on a strongly isolating language). Second, concreteness scores are language-dependent too, because it cannot be assumed that there is a perfect match between two lexemes from two different languages – however semantically similar such lexemes may be – and a given concept. The English noun *air*, for instance, has a concreteness score of 4.11/5 in Brysbaert et al. (2014), while the French noun *air* 'air' has a much more abstract score of 1.93/5 in Bonin et al. (2018). Such discrepancies across languages may depend on a variety of motivations, ranging from actual differences in conceptualization to semantic anisomorphism between the two lexemes or, more trivially, to slight differences in the instructions provided to speakers. This led us to focus in this paper on a single language, leaving a much-needed exploration of cross-linguistic variation in the relation between linguistic features and conceptual concreteness for future research.

3 Methods

3.1 Datasets

3.1.1 The main dataset: concreteness ratings

The main dataset of concreteness ratings comes from Brysbaert et al. (2014) large-scale rating study, where 40,000 English word forms were rated for concreteness

on a scale from 1 (= maximally abstract) to 5 (= maximally concrete). Over four thousand U.S.-resident native speakers of English participated in the study, with a wide range of ages and educational levels being represented. The instructions given to participants defined concrete words as those that "refer to things or actions in reality, which you can experience directly through one of the five senses", and abstract words as those that "refer to meanings that cannot be experienced directly but which we know because the meanings can be defined by other words" (Brysbaert et al. 2014: 906). In using these ratings, we adhere to the idea that the concreteness/abstractness distinction is best characterized in terms of whether concepts are accessible to the senses (Paivio et al. 1968). This is the most common way of operationalizing concreteness, but there are other operationalisations possible (Dunn 2015). The concreteness ratings will be used as the main response variable (dependent variable) in all subsequent analyses, with different linguistic predictors as independent variables.

It is important to state that the data quality of the concreteness measurements is not uniform. First, the concreteness ratings may be more volatile for rare words that are not known by a sufficient number of people. Because of this, we used word prevalence norms (Brysbaert et al. 2016; Keuleers et al. 2015) to restrict the dataset to only those words that are known by at least 95% of the population. This led to an exclusion of 10,266 data points (28.5% of the full dataset). A second aspect of data quality relates to the variability of concreteness ratings across different participants. It is known that words at the extreme points of the concreteness continuum have lower standard deviations across raters (Pollock 2018), i.e., raters agree more with each other for words that are clearly abstract or clearly concrete. There are multiple ways of entering standard deviations into the analysis. We decided to use weighted regression, where words with higher standard deviations (i.e., more variability across participants) contribute less to the overall result. This way, we do not have to exclude words based on an arbitrary cut-off value (e.g., the 50, 60, 70% lowest SD words). However, of course it is methodologically interesting to note whether the incorporation of standard deviation matters for the outcome of this analysis. To assess that this is the case, we compared the fit of standard regression models to those that incorporate standard deviations as regression weights.

3.1.2 Predictor datasets

We consider the following variables presented in Table 2, with predictions based on the discussion in Section 2.

We use corpus-derived part-of-speech tags from the SUBTLEX movie subtitle corpus (Brysbaert et al. 2012), morpheme parses from MorphoLex (Sánchez-Gutiérrez et al. 2018), morpheme counts from the English lexicon project (Balota et al. 2007),

Table 2: Predictions about concreteness for the linguistic variables considered.

Linguistic variable	Prediction for concreteness
Part of speech	Nouns > verbs > adjectives and adverbs > function words
Morphological structure	Morphologically simplex words > suffixed words
Countability	Count nouns > mass nouns
Etymology	Other > French, Latin

countability data from the Bochum English Countability Lexicon (Kiss et al. 2016), and etymological data from the Oxford English Dictionary.

3.1.3 Lemmatization

Our hypotheses are specified with respect to the structure of the lexicon and make no clear predictions with respect to whether inflected forms differ in concreteness. The Brysbaert concreteness rating study included some inflected forms, which show only small differences in concreteness (e.g., *dog* 4.85, *dogs* 5; *eye* 4.9, *eyes* 4.85). Here, we lemmatized all forms using the texstem package version 0.1.4 (Rinker 2018), thus getting rid of inflectional morphology (averaging over the different concreteness values). This is not only motivated based on the small differences in concreteness ratings, but also because it facilitates merging and comparison across datasets (e.g., the countability lexicon considers lemmas), and because it allows us to avoid violating the independence assumption of standard regression (i.e., including inflections would mean that the same lemma has multiple datapoints associated with it, thus artificially increasing our sample size).

3.2 Statistical analysis

All analyses were performed with R version 4.1.1 (R Core Team 2019) and the tidyverse package version 1.3.1 (Wickham et al. 2019). The car package version 3.0.11 (Fox and Weisberg 2018) was used to compute variance inflation factors to assess collinearity. All datasets and code are made publicly available in the following Open Science Framework repository: https://osf.io/rej6b/.

We performed a series of regression analyses, each time with concreteness as the main response variable. We first consider each linguistic predictor in isolation before we perform a simultaneous regression analysis with all predictors. One reason for performing separate regression analyses first is the presence of missing values for different words in the different data sets, as there are only few words

having data for all variables. Running individual regressions allows us to assess each variable's influence on concreteness ratings for the biggest dataset possible. Another reason for performing separate regressions is that some variables only have values for certain subgroups of the data. In particular, the count/mass distinction only exist for nouns. After having performed the individual regressions, we performed a simultaneous analysis with all predictors together.

4 Results

4.1 Lexical category (part of speech)

For our first analysis, we consider a word's "dominant part of speech", which corresponds to the part of speech that a word occurred most frequently in within the SUBTLEX corpus. For example, the word form *leak* was used 81% as a noun and 19% as a verb. This word form would thus be treated as a noun in the following analysis. Figure 1 shows the concreteness ratings as a function of part of speech.

Except for adverbs, the results were largely as predicted: nouns were the most concrete ($M = 3.59$, $SD = 1.03$), followed by verbs ($M = 2.95$, $SD = 0.81$), adjectives ($M = 2.49$, $SD = 0.73$), function words ($M = 2.38$, $SD = 0.77$), and adverbs ($M = 2.08$, $SD = 0.55$). An omnibus test reveals that there is an overall effect of part of speech ($F(4, 19123) = 1722.0$, $p < 0.0001$) that described a considerable 26% of the variation in concreteness ratings across words ($R^2 = 0.26$). When the same analysis was repeated with regression weights penalizing high-SD words, the part of speech predictor described even more variance, $R^2 = 0.34$.
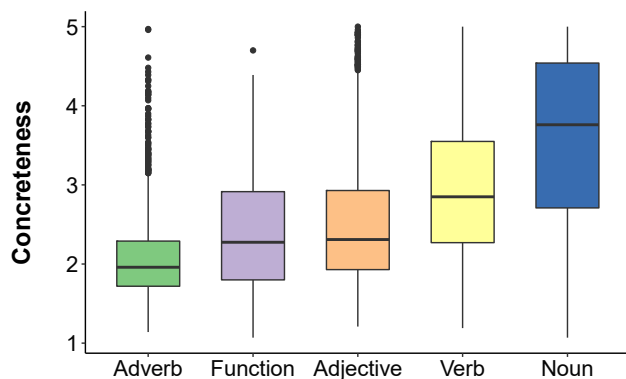

Figure 1: Concreteness ratings as a function of part of speech.

4.2 Morphological structure

In their analysis of morphological structure, Reilly and Kean (2007) excluded compound words. Here, we follow this analysis approach given that our hypotheses for morphological structure (laid out in Section 2.2) relate to derivation and not to compounding as a word formation process. We compiled a list of English compound words by collating the lists of Juhasz et al. (2015), Gagné et al. (2019), and Kim et al. (2019), which was used to exclude these words.

First, we look at the presence or absence of suffixes. This analysis compares morphologically simplex words (1 word = 1 morpheme) against morphologically complex words with one of the suffixes listed in Table 1 (1 word = 2 morphemes). Words without suffix were considerably more concrete ($M = 3.44$, $SD = 1.03$) than words with suffix ($M = 2.77$, $SD = 0.89$). The corresponding linear model ($t = 32.80$, $df = 11{,}164$, $p < 0.0001$) described 9% of the variance in concreteness ratings. Again, the weighted regression performed even better, describing 13% of the variance.

Figure 2 shows the result broken up by the most frequent suffixes (corresponding to the suffixes listed in Table 1). This shows that Reilly and Kean's (2007) claim that words with suffixes are on average more abstract clearly needs to be qualified. While we have replicated this finding in our previous analysis for a larger dataset than they consider (including not only just nouns), Figure 2 shows that there are vast differences in the average concreteness of particular suffixes. In fact, words with the suffix *-er* (as in *worker, dreamer, swindler, pensioner*) had similar average concreteness ($M = 3.91$, $SD = 0.66$) to monomorphemic words without suffixes ($M = 3.44$, $SD = 1.03$). This clearly shows that it matters *which* suffix one considers. Words with the suffix *-ly* were the most abstract ($M = 2.11$, $SD = 0.42$), in line with the previous analysis of part-of-speech tags, given that this suffix forms adverbs in English, which we have found to be the most abstract. An omnibus test for suffixed words showed that there was a reliable effect of type of suffix ($F(19, 3284) = 114.5$, $p < 0.0001$) that described 40.0% of variation in concreteness ratings for these words. When regression weights were incorporated to penalize high-SD words, this described variance rose to 47%.
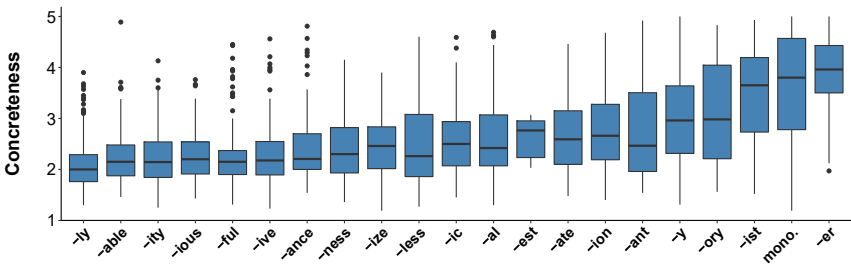

Figure 2: Concreteness ratings as a function of specific suffixes.

In a separate analysis, we regressed concreteness ratings onto morpheme counts from the English Lexicon Project. This analysis still excludes compound words but focuses on the full range of morphologically complex words, including words with multiple derivational affixes (e.g., *unbelievable*, *surpassingly*, *harmoniousness*). As shown in Figure 3, words with more morphemes were more abstract. We incorporated morpheme count as a continuous variable into the regression model, with an estimated decrease of 0.53 concreteness rating points for each additional morpheme (t = 70.61, df = 15,933, p < 0.0001). The morpheme count variable described 24% of the variance in concreteness ratings, and 31% in the weighted regression. Our analyses present a further extension of the results by Reilly and Kean (2007), which is that it is not just a binary distinction between morphologically complex and simplex words that matters, but instead, morphological complexity is monotonically related to the concreteness/abstractness distinction.

As mentioned above, Lewis and Frank (2016) found that word length was related with abstractness or, more specifically, with the related concept of conceptual complexity (see also Kelly et al. 1990). Because of this, it makes sense to estimate whether the morpheme count result above is independent of word length. In a simultaneous regression, we incorporated phoneme counts from the English Lexicon Project (Balota et al. 2007) as an additional predictor, on top of morpheme counts. For this analysis, there was an effect of phoneme count on abstractness (b = −0.08, SE = 0.004, t = 32.33, df = 15,932, p < 0.0001), as well as an effect of morpheme count (b = −0.36, SE = 0.01, t = 19.94, df = 15,932, p < 0.0001). The fact
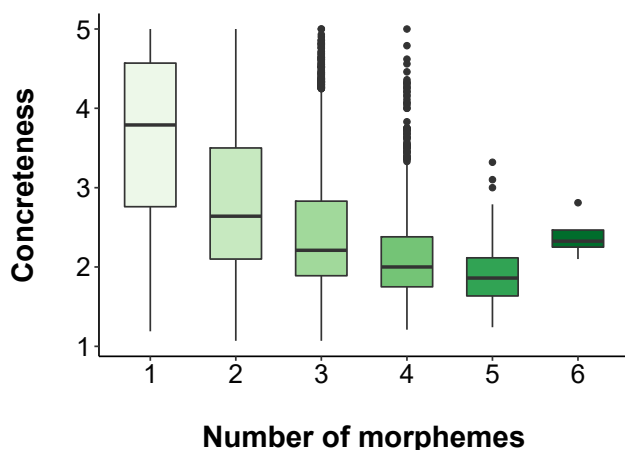

Figure 3: Concreteness ratings as a function of morpheme counts.

that the morpheme count coefficient shrunk (from −0.53 concreteness rating points per morpheme above to −0.36 concreteness rating points in this analysis) suggests that in the previous analysis, the effect of word length was in fact confounded with morpheme length. However, both have independent effects on abstractness. Comparison of coefficients as well as of R^2 shows that incorporating length only leads to slightly better model fit (25% as opposed to 24% variance described), which suggests that morphology matters more than just length.

4.3 Countability

For nouns only, we investigated the distinction between mass and count nouns, using data from the Bochum English Countability Lexicon (Kiss et al. 2016). The BELC database is organized around senses (see Section 2.3), rather than word forms. To match this data to the concreteness ratings, which do not separate senses, we decided to analyse only the subset of those nouns for which *all* senses were either count or mass without exception. This dataset included a total of 3,599 count nouns and 801 mass nouns. As seen in Figure 4, count nouns were on average more concrete ($M = 3.78$, $SD = 0.91$), and mass nouns relatively more abstract ($M = 2.76$, $SD = 0.95$) ($t = 28.74$, $df = 4,398$, $p < 0.0001$), and described 16% of the variance in concreteness ratings. As before, the weighted regression penalizing high-SD words described even more variance, 19%.

4.4 Etymology

We processed etymologies for a total of 287,341 unique word senses from the Oxford English Dictionary (OED). For sequences of borrowing (e.g., French < Latin),
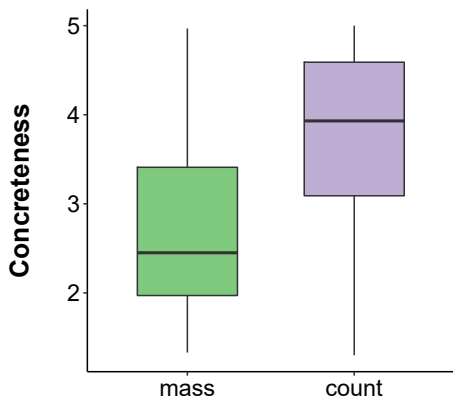

Figure 4: Concreteness ratings as a function of the count/mass noun distinction.

we used the proximate source (in this case, French) rather than the ultimate source (in this case, Latin), as it is the language that is most directly tied to the contact history of English. Of course, distinct word senses can have distinct etymological histories: typically, from the perspective of historical linguistics and OED, this is homonymy, with the same word form resulting from different origins. To match the OED data to the data from the concreteness rating study (which does not distinguish between word senses, as discussed above), we collapsed etymologies across distinct entries using a majority vote criterion (i.e., if 9 etymologies for a word form were indicated to be French-derived and 3 were indicated to be Latin-derived, we assigned "French" to the word form). In most cases, this majority vote criterion did not have to be exercised because different word senses were indicated to have the same etymologies. For the classification of an etymology as "French", we included "law French" but excluded "Canadian French" and "French Creole". Words that were not Latin or French were assigned the "other" category, which included words of Germanic origin as well as borrowings from other languages (e.g., Dutch, Chinese), words derived from person names or place names, imitative forms, and uncertain etymologies.

On average, Latin-derived words were the most abstract/least concrete (M = 2.86, SD = 0.95), closely followed by French-derived words (M = 2.95, SD = 0.99). Words in the 'other' category were much higher in concreteness (M = 3.21, SD = 1.10), as shown in Figure 5. An omnibus test reveals a reliable effect of the factor 'etymological origin' ($F(2, 25686) = 150.24$, $p < 0.0001$). However, this factor only described a total of 1% of the variance. The described variance rose to 1.5% in the weighted regression that penalizes high-SD words.
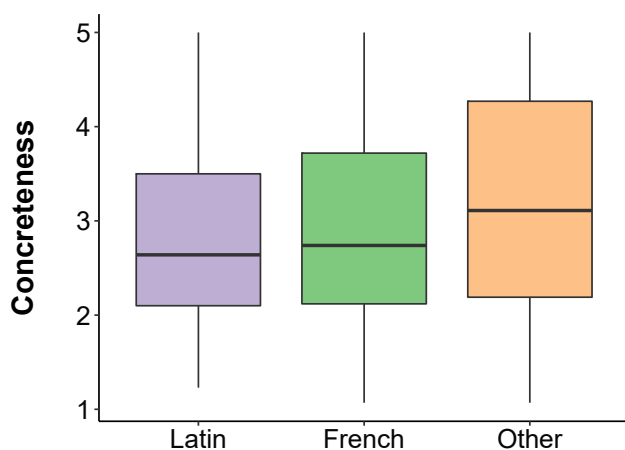

Figure 5: Concreteness as a function of etymological origin.

As discussed above, it is important to keep in mind that etymology interacts with the other variables considered so far, specifically morphological complexity and the count/mass noun distinction. A simple Chi-square test tabulating etymology against the countability data from Section 4.3 shows that the count/mass distinction differed reliably across etymologies ($\chi^2 = 37.77$, $df = 2$, $p < 0.0001$), with French words being much more likely to be mass nouns (Pearson adjusted standardized residual = +5.2). Similarly, Latin words were more likely to be mass nouns (+2.4). Finally, "other" words were much more likely to be count nouns (residual: −6.1). With respect to morphological complexity, Latin-derived words also had higher morpheme counts ($M = 2.10$) than French-derived words ($M = 1.82$), with "other" words being in between the two ($M = 2.07$). A Poisson regression model (used because the dependent measure morpheme count is a count variable) shows that etymological origin affects morpheme counts (likelihood ratio test against null model: $\chi^2 = 75.0$, $df = 2$, $p < 0.0001$). The next section will see whether etymological origin has a unique effect, or whether its contribution to predicting concreteness ratings can be reduced to the other variables.

4.5 Simultaneous regression analysis

So far, we have only considered each linguistic property by itself. However, as the previous section has discussed, some of the predictors are confounded with each other. To assess whether the different variables have independent effects when controlling for the others, we entered them all into the same simultaneous regression model. To account for the count/mass noun distinction in the full model, the "noun" category of the part-of-speech factor was split into two levels: "mass noun" and "count noun". Thus, the full model contained a total of four predictors: part of speech (including the count/mass noun distinction), the number of morphemes, the number of phonemes, and etymology. We used the rsq package version 2.2 (Zhang 2021) to compute partial R^2 as an estimate of the unique contribution of each predictor to the total variance of concreteness ratings.

The results are shown in Table 3. With all four predictors together, the overall model described 46% of the variance in concreteness ratings, and 56% for the weighted regression model. Partial values suggest that part of speech (including the count/mass distinction) describes the biggest share of variance, followed by morpheme counts, phoneme counts, and etymology, in that order.

5 Discussion and conclusion

Our results show that linguistic factors are indeed statistically associated with concreteness ratings, and in a manner that can largely be predicted based on

Table 3: Partial R^2 for the simultaneous regression model, and the simultaneous weighted regression model.

	Partial R^2	Partial R^2 (weighted regression)
Part of speech	0.30	0.38
Morpheme count	0.08	0.12
Phoneme count	0.03	0.03
Etymology	0.02	0.02

(cognitive) linguistic theory. As ratings are never for concepts directly, but they are always collected via tasks that necessitate mediation through words (Löhr 2021), considering linguistic factors is important. Here, we first show how our analyses provide empirical evidence for observations found in the linguistic literature. Next, we highlight methodological implications and discuss the relevance of our findings for empirical research in cognitive science.

Adverbs, function words, and adjectives are rated to be most abstract, followed by verbs, while nouns are rated to be most concrete. This fits the idea that parts of speech that are inherently relational, depending on other words in the same sentence, are also considered to be more abstract. A few words of comment are in order for the most abstract end of the scale (adverbs, function words, and adjectives). We did not make specific predictions about differences between adjectives and adverbs, but the fact that adverbs turned out to be more abstract is not surprising considering that prototypical adverbs are modifiers of verbs, and secondarily of adjectives, that is, of relational lexemes. As for function words, their mean rating is slightly less abstract than expected, probably because this heterogenous and small class contains many pronouns referring to people (*she*, *everyone*, etc.) and prepositions referring to positions in space (*behind*, *under*, etc.), which are judged as rather concrete by the participants of Brysbaert et al. (2014).

The part of speech result also has important ramifications for how parts of speech are defined within cognitive linguistic theory. While non-cognitive theories prefer formal criteria to delineate parts of speech, cognitive linguistics also emphasizes semantic criteria, highlighting differences in the conceptual structures that tend to go together with such distinctions as nouns versus verbs (Givón 1979; Langacker 1987a). However, that semantic differences actually reliably go together with part of speech contrasts across large sets of lexical items has rarely even been empirically demonstrated in a quantitative fashion (for an exception, see Strik Lievers and Winter 2018). From this perspective, our analyses show that lexical category differences do in fact go together with semantic differences, as measured by a large-scale concreteness rating study. This can be seen as a quantitative

confirmation of a core hypothesis of cognitive linguistics, which is that parts of speech do in fact tend to differ in their semantics.

We have furthermore shown that there are meaningful differences within the noun category, with mass nouns rated as more abstract than count nouns. This again provides empirical evidence for cognitive linguistic approaches that emphasize the notional basis of the distinction between mass and count nouns (Langacker 1987a: 203). More specifically, our analyses confirm the privileged relationship between mass (uses of) nouns and abstractness observed in the theoretical literature (Gillon 2017; Katz and Zamparelli 2012) and fit the idea that words with less individuated meanings are more abstract. And, by controlling for etymology, we show that this result is independent of the fact that mass nouns are also more likely to be French or Latin, two sources of English words that are associated with abstract meanings.

We have also shown that morphological structure is statistically associated with concreteness. This had already been established for nouns by Reilly and Kean (2007), however, only for the coarse measure of whether there was or was not suffixation. Here, we add to this result in four ways. First, we show that morphological structure is correlated with concreteness not just for nouns, but across all parts of speech, and controlling for part of speech. Second, we show that there are important differences between different suffixes, in part depending on the lexical category of the words formed by each suffix (e.g., suffixes that form adjectives and adverbs, like *-ly* and *-able*, tend to lie on the most abstract end of the scale). Third, we show that concreteness relates to the exact number of morphemes in a monotonic fashion, with decreasing concreteness for increasingly more morphemes. Fourth and finally, we show that phonological word length is correlated with abstractness independently of word length that is attributable to morphological complexity (Kelly et al. 1990; Lewis and Frank 2016).

Reilly and Kean (2007) furthermore found, for nouns only, that etymology mattered with respect to the concreteness dimension. This is in and of itself an important result, as it suggests that language history bears an imprint onto the semantic structure of the lexicon: the distribution of concrete/abstract words across semantic pockets in the lexicon is influenced by contact history. We extended this result to more words and different parts of speech, finding again that indeed, French and Latin words are more abstract than words of "other" origin.

Our results also make an important methodological point for any investigation that considers rating data, which are increasingly becoming used in cognitive linguistics (for an overview, see Winter, to appear). Pollock (2018) showed that standard deviations across participants are not uniformly distributed across the concreteness rating scale, with more extreme words (very abstract or very concrete) having lower standard deviations. Here, we took standard deviation into

account by adding it in the form of regression weights. These allowed low-SD words to contribute more towards the overall estimate than high-SD words. Across *all* analyses, we found that adding these regression weights increased model fit. This clearly shows that linguistic variables are more strongly associated with concreteness ratings for those ratings on which participants actually agree with each other. Results are weaker for more variable words. While there is no guarantee that this pattern carries over to other domains, it suggests that future psycholinguistic research using the norms should consider words with low standard deviations.

We also hope that our findings will speak to the cognitive science community more broadly. In line with recent literature stressing that the distinction between abstract and concrete concepts is the result of the interaction between multiple factors, going beyond perceptibility (Borghi et al. 2018; Harpaintner et al. 2018; Kiefer and Harpaintner 2020; Villani et al. 2019), we provided evidence that linguistic features also play a relevant role. To put it differently, our findings show that abstract and concrete words differ not only in the degree to which they are accessible to the senses, but also in their linguistic properties. This can in turn be seen as further evidence of the close association between conceptual and linguistic distinctions that is advocated by cognitive linguists. On a practical level, the results of our study suggest that, first, further effort should be put into looking for ways of capturing the concrete/abstract distinction that are not word-based (see Langland-Hassan et al. 2021). Second, researchers using word-based concreteness datasets should balance for linguistic factors when selecting stimuli for empirical studies of conceptual concreteness. For example, experiments that use nouns as stimuli should take into consideration the distinction between count and mass nouns, given that they tend to be characterized by different degrees of concreteness. Moreover, although much empirical research currently focuses on nouns, it would be important to verify whether and how its results generalize to other lexical categories.

More generally, we would like our study to raise awareness of the fact that concreteness ratings are associated with words, rather than with concepts directly. It could be the case that speakers who provided the concreteness ratings, faced with a long list of words to judge during the data collection, developed a strategy to perform the task less effortfully. For instance, they may have relied on linguistic shortcuts, rather than thinking deeply about the meaning and deciding how concrete it is, e.g., subconsciously knowing that verbs are more abstract than nouns, and relying on this information to quickly perform the rating task. Alternatively, it could also be the case that the ratings are indeed ratings about conceptual content (not words), but that linguistic factors have influenced such ratings. Participants may have mentally imagined a concept like SHRINKING and formed an idea about its concreteness, but then, realizing that this was manifested

in a verbal form, may have decreased their first judgment about this concept's concreteness (which is 3.2 out of 5 in Brysbaert et al., 2014). Our study, while not allowing us to know what speakers are really tapping into when performing concreteness rating tasks, contributes to highlighting the complexity of the relation between linguistic and conceptual structures. To conclude: although language is probably the best way through which to collect concreteness ratings, it is important that scholars acknowledge that the scores may measure linguistic phenomena together with the conceptual concreteness of the underlying concepts.

Data availability statement

The datasets and analysis scripts used during the current study are available in the following OSF repository: https://osf.io/rej6b/.

References

Asmuth, Jennifer & Dedre Gentner. 2017. Relational categories are more mutable than entity categories. *Quarterly Journal of Experimental Psychology* 70(10). 2007–2025.
Baker, Mark. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge: Cambridge University Press.
Baker, Mark & William Croft. 2017. Lexical categories: Legacy, lacuna, and opportunity for functionalists and formalists. *Annual Review of Linguistics* 3(1). 179–197.
Balota, David A., Melvin J. Yap, Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson & Rebecca Treiman. 2007. The English lexicon project. *Behavior Research Methods* 39(3). 445–459.
Binder, Jeffrey R., Chris Westbury, Keri A. McKiernan, Edward T. Possing & David A. Medler. 2005. Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience* 17(6). 905–917.
Blevins, James P. 2006. English inflection and derivation. In Bas Aarts & April McMahon (eds.), *The handbook of English linguistics*, 507–536. Malden, MA, USA: Blackwell Publishing.
Bolognesi, Marianna & Gerard J. Steen (eds.). 2019 *Perspectives on abstract concepts: Cognition, language and communication*. Amsterdam: John Benjamins.
Bonin, Patrick, Alain Méot & Aurélia Bugaiska. 2018. Concreteness norms for 1,659 French words: Relationships with other psycholinguistic variables and word recognition times. *Behavior Research Methods* 50(6). 2366–2387.

Borghi, Anna M., Laura Barca, Ferdinand Binkofski & Luca Tummolini (eds.). 2018. Varieties of abstract concepts: Development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373. 20170121.

Borghi, Anna M. & Ferdinand Binkofski. 2014. *Words as social tools: An embodied view on abstract concepts*. New York: Springer.

Brysbaert, Marc & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4). 977–990.

Brysbaert, Marc, Amy Beth Warriner & Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46(3). 904–911.

Brysbaert, Marc, Michaël Stevens, Paweł Mandera & Emmanuel Keuleers. 2016. How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology* 7. 1116.

Brysbaert, Marc, New Boris & Emmanuel Keuleers. 2012. Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods* 44(4). 991–997.

Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174(1). 99–149.

Conca, Francesca, Virginia Maria Borsa, Stefano Cappa & Eleonora Catricalà. 2021. The multidimensionality of abstract concepts: A systematic review. *Neuroscience & Biobehavioral Reviews* 127. 474–491.

Ćoso, Bojana, Marc Guasch, Pilar Ferré & José Antonio Hinojosa. 2019. Affective and concreteness norms for 3,022 Croatian words. *Quarterly Journal of Experimental Psychology* 72(9). 2302–2312.

Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.

De Felice, Irene. 2015. *Language and affordances*. University of Pisa Unpublished PhD thesis.

Doetjes, Jenny. 2017. The count/mass distinction in grammar and cognition. *Annual Review of Linguistics* 3(1). 199–217.

Dunn, Jonathan. 2015. Modeling abstractness and metaphoricity. *Metaphor and Symbol* 30(4). 259–289.

Durkin, Philip. 2014. *Borrowed words: A history of loanwords in English*. Oxford: Oxford University Press.

Fábregas, Antonio. 2014. Adjectival and adverbial derivation. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*, 276–296. Oxford: Oxford University Press.

Fieder, Nora, Lyndsey Nickels & Britta Biedermann. 2014. Representation and processing of mass and count nouns: A review. *Frontiers in Psychology* 5. 589.

Fox, John & Sanford Weisberg. 2018. *An R companion to applied regression*. Thousand Oaks, CA: Sage Publications.

Franzon, Francesca & Chiara Zanini. 2019. Different degrees of abstraction from visual cues in processing concrete nouns. In Marianna Bolognesi & Gerard J. Steen (eds.), *Perspectives on abstract concepts: Cognition, language and communication*, 167–184. Amsterdam: John Benjamins.

Gagné, Christina L., Thomas L. Spalding & Daniel Schmidtke. 2019. LADEC: The large database of English compounds. *Behavior Research Methods* 51(5). 2152–2179.

Gao, Chuanji, Laura B. Baucom, Jongwan Kim, Jing Wang, Douglas H. Wedell & Svetlana V. Shinkareva. 2019. Distinguishing abstract from concrete concepts in supramodal brain regions. *Neuropsychologia* 131. 102–110.

Giegerich, Heinz J. 2012. The morphology of *-ly* and the categorial status of 'adverbs' in English. *English Language and Linguistics* 16(3). 341–359.

Gillon, Brendan S. 2017. Semantic categorization. In Henri Cohen & Claire Lefebvre (eds.), *Handbook of categorization in cognitive science*, 2nd edn., 291–311. Amsterdam: Elsevier.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Guasch, Marc, Pilar Ferré & Isabel Fraga. 2016. Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods* 48(4). 1358–1369.

Harpaintner, Marcel, Natalie M. Trumpp & Markus Kiefer. 2018. The semantic content of abstract concepts: A property listing study of 296 abstract words. *Frontiers in Psychology* 9. 1748.

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. New York: Cambridge University Press.

Hitchings, Henry. 2008. *The secret life of words: How English became English*, 1st American edn. New York: Farrar, Straus and Giroux.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*, 2nd edn. Cambridge: Cambridge University Press.

Husić, Halima. 2020. *On abstract nouns and countability. An empirical investigation into the countability of eventuality denoting nominals*. Ruhr-Universität Bochum Unpublished PhD thesis.

Jespersen, Otto. 1924. *The philosophy of grammar*. London: George Allen and Unwin.

Juhasz, Barbara J., Yun-Hsuan Lai & Michelle L. Woodcock. 2015. A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods* 47(4). 1004–1019.

Katz, Graham & Roberto Zamparelli. 2012. Quantifying count/mass elasticity. In Jaehoon Choi, Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz & Alex Trueman (eds.), *Proceedings of the 29th west coast conference on formal linguistics*, 371–379. Somerville, MA: Cascadilla Proceedings Project.

Kelly, Michael H., Ken Springer & Frank C. Keil. 1990. The relation between syllable number and visual complexity in the acquisition of word meanings. *Memory & Cognition* 18(5). 528–536.

Keuleers, Emmanuel, Michaël Stevens, Paweł Mandera & Marc Brysbaert. 2015. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology* 68(8). 1665–1692.

Kiefer, Markus & Marcel Harpaintner. 2020. Varieties of abstract concepts and their grounding in perception or action. *Open Psychology* 2(1). 119–137.

Kim, Say Young, Melvin J. Yap & Winston D. Goh. 2019. The role of semantic transparency in visual word recognition of compound words: A megastudy approach. *Behavior Research Methods* 51(6). 2722–2732.

Kiss, Tibor, Francis Jeffry Pelletier & Halima Husić (eds.). 2021. *Things and stuff. The semantics of the count-mass distinction*. Cambridge: Cambridge University Press.

Kiss, Tibor, Francis Jeffry Pelletier, Halima Husic, Roman Nino Simunic & Johanna Marie Poppek. 2016. A sense-based lexicon of count and mass expressions: The Bochum English countability lexicon. *LREC 10 proceedings*, 2810–2814. European Language Resources Association. Available at: https://aclanthology.org/L16-1447.

Koptjevskaja-Tamm, Maria. 2015. Action nouns. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation*, 1195–2009. Berlin, Boston: Mouton de Gruyter.

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, Ronald W. 1987a. *Foundations of cognitive grammar*, vol. 1. Stanford: Stanford University Press.

Langacker, Ronald W. 1987b. Nouns and verbs. *Language* 63(1). 53.

Langland-Hassan, Peter, Frank R. Faries, Maxwell Gatyas, Aimee Dietz & Michael J. Richardson. 2021. Assessing abstract thought and its relation to language with a new nonverbal paradigm: Evidence from aphasia. *Cognition* 211. 104622.

Lehmann, Christian. 2015. *Thoughts on grammaticalization*, 3rd edn. Berlin: Language Science Press.

Lewis, Molly L. & Michael C. Frank. 2016. The length of words reflects their conceptual complexity. *Cognition* 153. 182–195.

Löhr, Guido. 2021. What are abstract concepts? On lexical ambiguity and concreteness ratings. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-021-00542-9.

Michaelis, Laura A. 2005. Entity and event coercion in a symbolic theory of syntax. In Jan-Ola Östman & Mirjam Fried (eds.), *Constructional approaches to language*, 45–88. Amsterdam: John Benjamins.

Middleton, Erica L., Edward J. Wisniewski, Kelly A. Trindel & Mutsumi Imai. 2004. Separating the chaff from the oats: Evidence for a conceptual distinction between count noun and mass noun aggregates. *Journal of Memory and Language* 50(4). 371–394.

Miller, D. Gary. 2012. *External influences on English: From its beginnings to the renaissance*. Oxford: Oxford University Press.

Moltmann, Friederike (ed.). 2020. *Mass and count in linguistics, philosophy, and cognitive science*. Amsterdam: John Benjamins.

Monaghan, Padraic & Seán G. Roberts. 2019. Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. *Cognition* 186. 147–158.

Montefinese, Maria, Ettore Ambrosini, Beth Fairfield & Nicola Mammarella. 2014. The adaptation of the affective norms for English words (ANEW) for Italian. *Behavior Research Methods* 46(3). 887–903.

Murphy, M. Lynne. 2010. *Lexical meaning*. Cambridge: Cambridge University Press.

Orr, John. 1944. On some sound values in English. *British Journal of Psychology* 35(1). 1–8.

Paivio, Allan, John C. Yuille & Stephen A. Madigan. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology* 76(1, Pt.2). 1–25.

Paul, Hermann. 1891. *Principles of the history of language*. (Trans.) H. A. Strong. London: Longmans, Green, and Co.

Pexman, Penny M., Emiko Muraki, David M. Sidhu, Paul D. Siakaluk & Melvin J. Yap. 2019. Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior Research Methods* 51(2). 453–466.

Pexman, Penny M., Ian S. Hargreaves, Jodi D. Edwards, Luke C. Henry & Bradley G. Goodyear. 2007. Neural correlates of concreteness in semantic categorization. *Journal of Cognitive Neuroscience* 19(8). 1407–1419.

Pittner, Karin, Fabian Barteld & Daniela Elsner (eds.). 2015. *Adverbs: Functional and diachronic aspects*. Amsterdam: John Benjamins.

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

Pollock, Lewis. 2018. Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior Research Methods* 50(3). 1198–1216.

Prandi, Michele. 2004. *The building blocks of meaning: Ideas for a philosophical grammar*. Amsterdam: John Benjamins.

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, Mass: MIT Press.

Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech & Svartvik Jan. 1972. *A grammar of contemporary English*. London: Longman.

R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rainer, Franz. 2015. Quality nouns. In Peter O. Müller, Ingeborg Ohnheiser & Susan Olsen (eds.), *Word-formation*, 1268–1284. Berlin, Boston: Mouton de Gruyter.

Rauh, Gisa. 2010. *Syntactic categories: Their identification and description in linguistic theories*. Oxford: Oxford University Press.

Rauh, Gisa. 2015. Adverbs as a linguistic category (?). In Karin Pittner, Daniela Elsner & Fabian Barteld (eds.), *Adverbs: Functional and diachronic aspects*, 19–46. Amsterdam: John Benjamins.

Reilly, Jamie, Chris Westbury, Jacob Kean & Jonathan E. Peelle. 2012. Arbitrary symbolism in natural language revisited: When word forms carry meaning. *PLoS One* 7(8). e42286.

Reilly, Jamie & Jacob Kean. 2007. Formal distinctiveness of high- and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science* 31(1). 157–168.

Reijnierse, Gudrun, Christian Burgers, Marianna Bolognesi & Tina Krennmayr. 2019. How polysemy affects concreteness ratings: The case of metaphor. *Cognitive Science* 43(8). e12779.

Reilly, Jamie, Jinyi Hung & Chris Westbury. 2017. Non-arbitrariness in mapping word form to meaning: Cross-linguistic formal markers of word concreteness. *Cognitive Science* 41(4). 1071–1089.

Rinker, Tyler W. 2018. textstem: Tools for stemming and lemmatizing text version 0.1.4. *R package*. Available at: http://github.com/trinker/textstem.

Rips, Lance J. & Susan J. Hespos. 2019. Concepts of objects and substances in language. *Psychonomic Bulletin & Review* 26(4). 1238–1256.

Sánchez-Gutiérrez, Claudia H., Hugo Mailhot, S. Hélène Deacon & Maximiliano A. Wilson. 2018. MorphoLex: A derivational morphological database for 70,000 English words. *Behavior Research Methods* 50(4). 1568–1580.

Sapir, Edward. 1921. *Language. An introduction to the study of speech*. New York: Harcourt, Brace & World.

Schreuder, Robert & R. Harald Baayen. 1997. How complex simplex words can be. *Journal of Memory and Language* 37(1). 118–139.

Soares, Ana Paula, Ana Santos Costa, João Machado, Montserrat Comesaña & Helena Mendes Oliveira. 2017. The Minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods* 49(3). 1065–1081.

Spreen, Otfried & Rudolph W. Schulz. 1966. Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior* 5(5). 459–468.

Strik Lievers, Francesca & Bodo Winter. 2018. Sensory language across lexical categories. *Lingua* 204. 45–61.

Talmy, Leonard. 2000. *Toward a cognitive semantics*, vol. 1. Cambridge, Mass: MIT Press.

Traugott, Elizabeth Closs. 1982. From propositional to textual and expressive meanings; some semantic-pragmatic aspects of grammaticalization. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Perspectives on historical linguistics*, 245–271. Amsterdam: John Benjamins.

Villani, Caterina, Luisa Lugli & Marco Tullio Liuzza & Anna M. Borghi. 2019. Varieties of abstract concepts and their multiple dimensions. *Language and Cognition* 11(3). 403–430.

Vonk, Jet M. J., Loraine K. Obler & Roel Jonkers. 2019. Levels of abstractness in semantic noun and verb processing: The role of sensory-perceptual and sensory-motor information. *Journal of Psycholinguistic Research* 48(3). 601–615.

Wang, Jing, Julie A. Conder, David N. Blitzer & Svetlana V. Shinkareva. 2010. Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping* 31(10). 1459–1468.

Werkmann Horvat, Ana, Marianna Bolognesi & Aditi Lahiri. 2021. Processing of literal and metaphorical meanings in polysemous verbs: An experiment and its methodological implications. *Journal of Pragmatics* 171. 131–146.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry & Jim Hester. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). 1686.

Wierzbicka, Anna. 1985. "Oats" and "wheat": The fallacy of arbitrariness. In John Haiman (ed.), *Iconicity in syntax*, 311–342. Amsterdam: John Benjamins.

Winter, Bodo. to appear. Managing semantic norms for cognitive linguistics, corpus linguistics, and lexicon studies. In Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller & Lauren B. Collister (eds.), *The open handbook of linguistic data management*. Cambridge, MA: MIT Press.

Yao, Zhao, Jia Wu, Yanyan Zhang & Zhenhong Wang. 2017. Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods* 49(4). 1374–1385.

Zamparelli, Roberto. 2020. Countability shifts and abstract nouns. In Friederike Moltmann (ed.), *Mass and count in linguistics, philosophy, and cognitive science*, 191–224. Amsterdam: John Benjamins.

Zhang, Dabao. 2021. rsq: R-squared and related measures. *R package version 2.2*. Available at: https://CRAN.R-project.org/package=rsq.

Žirmunskij, Viktor. 1966. The word and its boundaries. *Linguistics* 4(27). 65–91.