Form to meaning mapping and the impact of explicit morpheme combination in novel word processing

Rolando Bonandrini [*], Simona Amenta, Simone Sulpizio, Marco Tettamanti, Alessia Mazzucchelli, Marco Marelli

Department of Psychology, University of Milano-Bicocca, Milan, Italy

ARTICLE INFO

ABSTRACT

In the present study, we leveraged computational methods to explore the extent to which, relative to direct access to semantics from orthographic cues, the additional appreciation of morphological cues is advantageous while inducing the meaning of affixed pseudo-words. We re-analyzed data from a study on a lexical decision task for affixed pseudo-words. We considered a parsimonious model only including semantic variables (namely, semantic neighborhood density, entropy, magnitude, stem proximity) derived through a word-form-to-meaning approach (*ngram*-based). We then explored the extent to which the addition of equivalent semantic variables derived by combining semantic information from morphemes (*combination*-based) improved the fit of the statistical model explaining human data. Results suggest that semantic information can be extracted from arbitrary clusters of letters, yet a computational model of semantic access also including a *combination*-based strategy based on explicit morphological information better captures the cognitive mechanisms underlying human performance. This is particularly evident when participants recognize affixed pseudo-words as meaningful stimuli.

1. Introduction

Extracting meaning from written letter strings is fundamental to interact with the external world in a rapid and efficient way. As soon as we encounter a word we know, we immediately and automatically activate a set of mental representations. Such representations provide cues on what that given word indicates, as well as on the most adaptive reaction we should have towards it.

Remarkably, if this ability was limited only to familiar word-meaning associations, every time we come across an unfamiliar yet plausible word (e.g., a neologism), we would have no clue on what it relates to and on how to react to it. On the contrary, although we may have never seen a specific letter string before, we might have a vague idea on what its meaning could be. In this direction, it was shown that semantic features associated with a pseudo-word determine their perceived meaningfulness (Marelli & Baroni, 2015) and the ease with which they are classified as non-words in a lexical decision task (Hendrix & Sun, 2021, Günther & Marelli, 2020; Sulpizio, Pennucci, & Job, 2021). This evidence suggests that the mechanisms that allow the extraction of meaning from letter strings can be generalized to novel, unfamiliar stimuli such as neologisms and pseudo-words (Gatti, Marelli & Rinaldi, 2022).

A particular case of pseudo-words is represented by novel affixed words (or affixed pseudo-words), i.e., unattested complex words composed by an existing stem and an affix. An example of affixed pseudo-word is *tablist*: its stem is an existing word (e.g., *table*), its affix

---

is existing as well and commonly encountered in other derived words (e.g., *-ist*, as in psychologist, pianist, guitarist), but the whole affixed word is not attested. Although not all these forms are equally interpretable from the semantic point of view (e.g., *windowless* vs. *windowist*, Marelli & Baroni, 2015), it is fairly easy to acknowledge that even a difficult-to-interpret form such as *windowist*, probably refers to someone related to/that works with windows.

The relevance of this type of pseudo-words is not trivial, as the ability to correctly interpret them can determine the ease with which we can adapt to a novel context. For instance, the term *tablature* (a type of musical notation) is formed by an unquestionably known stem (i.e., *table*) and a commonly encountered affix (e.g. *-ture*, as in signature), although the complete form may be unknown. Still, being able to use the familiar "chunks" of the word to get a sense of its meaning might come useful to figure out that such alternative notation to the classic one on the pentagram exists, and that it might be useful when learning to play a new instrument. Yet, determining the mechanisms necessary for extracting meaning from affixed pseudo-words is not straightforward.

1.1. A role for morphology?

In trying to deal with an affixed pseudo-word, one might parse it in its morphological constituents, namely its stem and affix. This intuition is not trivial: experimental evidence that recognition of a stem (e.g., *fail*) is speeded by prior presentation of a morphologically related word (e.g., *failure*; Drews & Zwitserlood, 1995; Rastle et al., 2000), together with the observation of stem frequency effects in lexical decision tasks (New et al., 2004), have suggested that morphological information is important for word recognition (morphological framework).

For what concerns pseudo-words, ever since the pioneering work by Taft and Forster (1975, 1976) it has been known that rejecting a non-word in a lexical decision task is more difficult when it contains embedded morphemes (e.g., "*dejuvenate*") than when it does not (e.g., "*depertoire*"). In a similar vein, Burani et al. (1997) reported that suffixed pseudo-words (e.g., "*buyist*") are more difficult to reject than pseudo-words composed by an existing stem and a non-existing suffix (e.g., "*buyost*"), although this turned out to occur only for frequent word-endings. More recently, Crepaldi et al. (2010) provided evidence that suffixed pseudo-words (e.g., *gasful*) take longer to be rejected than orthographic controls with non-morphological endings (e.g., *gasfil*). A similar result has been observed with compound pseudo-words, with rejection times being related to the extent of the contribution of each constituent to the estimated meaning of any given compound form (Günther & Marelli, 2020).

However, morphological analysis might represent neither the only, nor the simplest gateway towards meaning for an affixed pseudo-word. Accordingly, psycholinguistic research has provided evidence that lexical processing is sensitive to the correspondence between sub-lexical units and meaning, which might be sufficient to inform about word meaning (Taft & Kougious, 2004). In other words, the identification of morphemes may not be necessary: the same type of information could be extracted from sublexical strings without formal recognition of these letter clusters as morphological units (form-to-meaning framework). Support for this view comes from the computational literature. The connectionist framework has described morphological effects as emerging from connections - mediated by an intermediate hidden layer- between word form and meaning (Harm & Seidenberg, 2004; for a review, see Stevens & Plaut, 2022). In a similar vein, the more recent Naïve Discriminative Learning (NDL) framework (Baayen et al., 2011) has described morphological effects as a mere epiphenomenon of the association - in this case not mediated by a hidden layer - between clusters of letters (e.g., bigrams) and a symbolic representation of the meaning of words. This perspective suggests that morphology might simply represent a special case of a more general-level language systematicity (Amenta, Günther & Marelli, 2020; Crepaldi, Amenta, Marelli, 2019).

In principle, the form-to-meaning framework is simpler than the morphological framework, as the former does not require explicit formal knowledge on morphemes to describe access to meaning of affixed pseudo-words. Moreover, it is a more general-level explanation, since it works also for words that are morphologically simple. Given its inherent simplicity and its ability to explain human performance, under a principle of parsimony the form-to-meaning framework should be considered the preferred (i.e., the *baseline*) explanation of meaning extraction from affixed pseudo-words, unless a more complex account provides a significant improvement in the fit to behavioural data. In our case, the "more complex" account is constituted by the morphological framework, as this latter constitutes a specific condition for semantic access (i.e., one that would be involved in the case of morphologically complex stimuli) that requires some understanding of how morphemes contribute to meaning. In such perspective, it is critical to test whether, in inducing the meaning of an affixed pseudo-word, the combination of semantic information conveyed by morphemes (i.e., stems and affixes) provides any advantage with respect to the extraction of meaning from non-morphologically-defined sub-lexical units.

The present study attempts to shed light on this issue by building on existing approaches developed in the domain of computationally-implemented theories of human semantic representations. Indeed, at a verbal-conceptual level both the form-to-meaning explanations and the morpheme-based explanations are consistent with behavioral data; the predictions of the two schools of thought concerning empirical effects are largely overlapping, and both are principally able to account for phenomena in the morphological processing literature. In such a scenario, simple experimental approaches offer limited possibilities for model adjudication. In this perspective, from a quantitative point of view, computational implementations of these explanations offer the unique opportunity to test whether - in describing how humans extract meaning from an affixed pseudo-word - we should prefer a description grounded on the more parsimonious form-to-meaning framework alone, or rather the overt extraction of morphological elements is necessary and/or advantageous. In the next two paragraphs we will review how semantic representations can be profitably conceptualized from a computational point of view, and how existing computational frameworks conceive operations on semantic representations to induce the meaning of affixed pseudo-words.

1.2. Concepts as vectors

The so-called *distributional hypothesis* (Harris, 1954) is the assumption that words with a similar meaning occur in similar contexts. From this assumption it follows that there is a link between a word distribution over contexts and its meaning. This principle constitutes the foundation of contemporary computational implementations of (distributional) semantic models, in which the meaning of a word is represented by a vector containing numbers quantifying the degree of association between that word and a set of contexts. For example, let's assume we want to quantify the meanings of the words *cherry* and *car*. A convenient way to do so could be to look for these words in the books we have on our bookshelf, and to count how many times each word appears in each book. For illustration purposes, let's assume that four books are on our shelf: "the book of food", "the book of small objects", "the book of heavy objects", and "the book of means of transport". The word *cherry* appears 200 times in the first book, 180 in the second, 1 time in the third and 0 in the fourth. Conversely, the word *car* appears 0 times in the "book of food", 3 times in the "book of small objects", 160 times in the "book of heavy objects", and 190 times in the "book of means of transport". What we just obtained are semantic vectors of four dimensions each (corresponding to the contexts in which word occurrences were counted), respectively, for *cherry* [200 180 1 0] and *car* [0 3 160 190]. In analogy with this example, the first step in the formation of a semantic space in classical models of computational semantics is the extraction - within large linguistic corpora - of word occurrences in different contexts, such as documents (Griffiths et al., 2007; Landauer & Dumais, 1997), or a window of surrounding words (Lund & Burgess, 1996; Mikolov, Chen, et al., 2013; Mikolov, Sutskever, et al., 2013; Pennington et al., 2014). However, in computationally implemented semantic models the relationship between words and semantic dimensions is often less intuitive than in the example presented above. On the one hand, weights in semantic vectors are rarely raw co-occurrences: they either represent the result of weighting schemes applied to co-occurrences (e.g., positional weighting: Lund & Burgess, 1996; log-entropy weighting: Martin & Berry, 2007; point- wise mutual information: Church & Hanks, 1990), or they might represent weights in neural networks (Mikolov, Chen, et al., 2013; Mikolov, Sutskever, et al., 2013), in which words represent the input layer from which their context (i.e., surrounding words) is predicted in the output layer (or vice-versa). On the other hand, semantic dimensions in these models are typically *sub-symbolic*, i.e., semantic dimensions are not necessarily translatable into contexts with a directly intelligible meaning. This is particularly evident when *dimensionality reduction* is applied (see Günther, Rinaldi & Marelli, 2019 for a review) to reduce the number of semantic dimensions with the aim to produce a more compact, yet equally informative model.

In the context of the present work, we will build on computational implementations of semantic knowledge to explore how humans exploit available information to induce the meaning of affixed pseudo-words. In what follows, we will specify the mechanisms by which previously existing approaches can be used for this purpose.

1.3. Vector representations for affixed pseudo-words: Ngram and combination-based approaches

Previous literature suggests that there are at least two different ways through which the meaning of affixed pseudo-words can be determined, based on existing information in a semantic space (for a similar discussion on existing words, see Amenta, Günther & Marelli, 2020). The first approach assumes that, in addition to word-level information, also sub-word information (i.e., chunks of *n* letters, or *ngrams*) can be used as input to induce semantic vectors. This represents an implementation of the form-to meaning framework. Therefore, the meaning of a given letter string (regardless of whether it is a familiar word or not) can be determined simply by summing the semantic vectors of its complete form (if it exists) and those of its constituent *ngrams*. Thus, the meaning of the word *pretentious* could be obtained by summing the semantic vector of the full word with those of its *ngrams*: *pretentious + prete + reten + etent + tenti + entio + ntiou + tious*. This possibility is implemented in the Fasttext model (Bojanowski et al., 2017). Remarkably, in a recent work, Hendrix and Sun (2021) showed that the semantic measures derived from this model predict human response times in a lexical decision task. Moreover, Hendrix and Sun were also able to observe that performance for a given item was influenced by the average amount of relatedness between the item and its five most semantically similar words (i.e. semantic neighborhood density, see paragraph 3.3 for a detailed explanation of the meaning of this variable). Of note, this was true both when the target stimulus of the lexical decision task was a real word (a denser semantic neighborhood facilitated performance) and when it was a pseudo-word (a denser semantic neighborhood hindered performance). Gatti et al. (2022) combined this framework with a priming paradigm and highlighted that the same associative mechanisms governing word meaning also subserve the processing of pseudo-words. This was done for both English and Italian, thus providing cross-linguistic face validity in favor of this method and the psycholinguistic viability of Fasttext also for Italian. Crucially, the main feature of the *ngram-based* approach is that it does not require the assumption that morphemes play a critical role in determining stimulus meaning, as morpheme weights are as prominent as those of linguistically undefined (and, in principle, meaningless) *ngrams*. From a broader point of view, this framework embodies at a computational level the assumption that information from linguistically undefined orthographic chunks is enough to induce the meaning of a morphologically complex letter string (Taft & Kougious, 2004).

An alternative line is provided by *combinatorial* approaches, which represent implementations of the morphological framework. These approaches build on the assumption that morphology allows to generate novel meanings by means of the combination of known meaning-bearing elements (Amenta, Günther & Marelli, 2020). The computational foundations of this framework were laid by the pioneering work of Mitchell and Lapata (2010), in which semantic representations for short phrases were extracted by applying mathematical operations (such as addition or element-wise multiplication) to the vectors of the constituents. More recently, it has been proposed that, given the semantic vector *u* of a word (constituent) and the semantic vector *v* of another constituent, the semantic vector of their combined form can be computed as $c = M * u + H * v$. The terms M and H represent weight matrices estimated from training data indicating semantic features that a word activates when it assumes the roles of, respectively, Modifier or Head (Guevara, 2010). This

framework has been formally implemented in the Compounding as Abstract Operation in Semantic Space model (CAOSS, Marelli, Gagné & Spalding, 2017), that proved to accurately predict relational priming data in humans (Marelli et al., 2017), as well as to accurately model the effects of integrating the constituent meanings into a single new compound representation (Günther & Marelli, 2020; Günther & Marelli, 2019; Günther, Marelli, & Bölte, 2020). It is worth noting that both the *ngram* and the *combination-based* approaches are efficient at the computational level and proved to have face validity in psychological terms (by providing accurate predictions of human performance). Importantly, both methods are potentially able to capture morphological phenomena. Yet, the *ngram*-based approach builds aspecifically on sublexical elements: some of these elements happen to be morphemes, but the system does not a-priori distinguish morphemes and non-morphological sublexical chunks, and has no mechanism explicitly dedicated to morphological information; the model has no morphological awareness, and any morphological effect it displays will be a by-product of more general processes. On the other hand, the *combination-based* approach is informed of the morphological structure of both words (during model training) and pseudo-words (when testing the model predictions against human data), and explicitly combines stem and affix to induce the meaning of the whole string, while ignoring other sublexical chunks; the mechanisms implemented in this model are specifically geared towards morphological information.

1.4. Aims

In this work, we will test whether - relative to direct access to semantics from orthographic cues (*ngram*-based model) - the additional explicit consideration of morphology (simulated through a *combination*-based model) provides a sizeable contribution while deriving the meaning of an affixed pseudo-word (Fig. 1). This will be assessed in terms of the models' ability to predict human performance patterns while dealing with affixed pseudo-words in a lexical-semantic task: if the combination-based approach is able to explain human performance over and above the ngram-based one, this will indicate that morphological processing cannot be completely reduced to a general form-meaning procedure. To this end, we will re-analyse data by Amenta et al. (under review), which are fully available at: https://osf.io/mpfdy/.

After having tested this hypothesis, we will explore to what extent the diverse estimates generated by the models can be informative of the cognitive underpinnings of novel word processing.

2. Code availability

The code used in this work is available at: https://osf.io/qp2ad/.

3. Methods

3.1. Behavioral data

In this study, we employed as a test case the data provided in Amenta et al. (under review). We refer to the original paper for details about aims, methods and results of their lexical decision experiment. In the current study, we focus on the novel derivation data. We induce vector representations for the novel derivation stimuli (see below) and test the impact of the semantic variables extracted from them against the behavioral data. In what follows, a brief description of the sample, experimental stimuli and experimental procedure is provided.

From Amenta et al. (under review), we selected only participants who were native speakers of Italian (N = 51 (40 females), mean age = 23, SD = 4, all recruited at the University of Milano-Bicocca, Italy). Participants performed a non-speeded lexical decision task, during which they were asked to decide if the stimulus presented on a computer screen was an Italian word or not ("YES"= "it is an Italian word"; "NO"= "it is not an Italian word"). Novel Derivation strings consisted of a novel legal (i.e., syntactically well-formed) combination of an existing stem plus an existing suffix (e.g., *stelloso* interpretable as "*star-like*" - literally "*starrous*"). Thirteen derivational suffixes were used (-abile, -aggine, -eria, -zione, -ezza, -ismo, -iere, -ista, -itudine, -mento, -enza, -oso, -tore) to generate the set of all complex pseudo-words. Novel derivations were, on average, 9.80 (SD = 1.56) characters long. In the current work, we focused on novel derivation strings only (affixed pseudo-words). The non-lexicality of novel derivations was checked at the time of behavioral data collection (between 2017 and 2018) by means of dictionaries (Treccani and Repubblica), Google hits, and SUBTLEX-IT (corpus size: 128.339.065, Crepaldi et al, 2015)[1]. Given their status of non-words, frequency for these stimuli was effectively 0 in the reference corpora.

---

[1] The original work by Amenta et al. (under review), included a total of 250 items, of which 50 morphologically complex words, 50 morphologically simple words, 50 novel derivations (analyzed here), 50 affixed non-words (composed by nonsense pseudo-stems and an existing suffix; e.g., *sterable*), and finally, 50 non-affixed non-words (including an existing stem with a non-morphological ending; e.g., *sockel*). The same 13 suffixes were used for all affixed items (words, non-words and novel derivations), moreover all non-affixed items ended in the same 13 non-morphological endings. The use of the same endings (morphemic or not) aimed at limiting the variability in the relative frequency of endings (see Amenta et al., under review, for further details). Across all stimuli categories, the average length of the stimuli used in the original experiment is 8.712 characters (SD= 2.148). The average length of the novel derivation stimuli analyzed in this work is 9.800 characters, SD= 1.564.
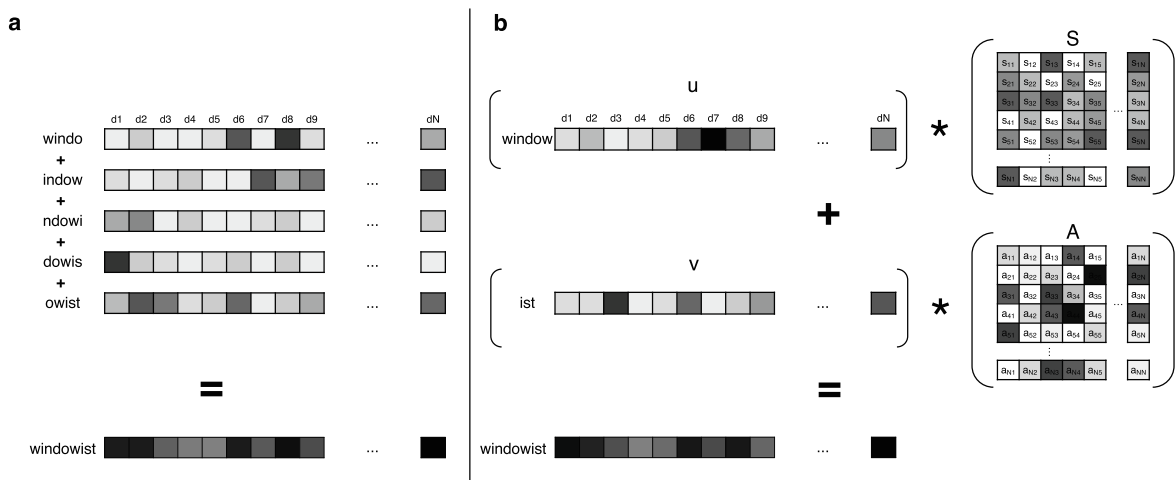
Fig. 1. Graphic representation of the operations applied to the semantic space by the (**a**) *ngram* and (**b**) *combination*-based models to induce the meaning of an affixed pseudo-word. The terms *d1-d9* indicate semantic dimensions, *u* is the semantic vector of the stem, *S* is the weight matrix for the stem vectors, *v* is the semantic vector of the affix, *A* is the weight matrix for the affix vectors.

3.2. Computational modelling

Ngram-based model. The *ngram*-based semantic vectors for affixed pseudo-words (Fig. 1a) were derived from the trained semantic space for Italian available on the Fasttext website (https://fasttext.cc/docs/en/crawl-vectors.html), trained on Common Crawl and Wikipedia using a Continuous Bag Of Words (CBOW) method with position weights and *ngrams* of length 5 (default setting in the Fasttext model. For further technical details regarding the model, please refer to Bojanowski et al., 2017). For each stimulus (affixed pseudo-word), semantic vectors were obtained. It is important to consider that -in the context of this work- the *ngram* approach represents the baseline on top of which the additional contribution of explicitly defined and combined morphological information was evaluated. In this regard, larger *ngrams* provide more defined semantic information (i.e., closer to that of actual words) than shorter *ngrams*, as testified by better performance of models with larger *ngrams* (Bojanowski et al., 2017). Previous research also described advantages in learning efficiency with no significant drop in performance for Fasttext models only using 5-grams compared to models including *ngrams* from 3 to 5 (Grave et al., 2018). Hence, the 5-gram option we adopted ensures we are considering a strong baseline. Secondly, an empirical psycholinguistic validation of the model for Italian already exists in the literature, and it is based on a Fasttext model built using *ngrams* of length 5 (Gatti et al., 2022). Since the present work builds on the validation by Gatti and colleagues, the same parameters (including *ngrams* of length 5) were used for the sake of comparability.

Composition-based model. We used an adaptation to derived words of the CAOSS (Marelli et al., 2017) model, which was originally developed to characterize the semantic operations underlying compounding. CAOSS is not the only computational model entailing operations in the semantic space that could be applied to simulate affixation processes. Indeed, the Functional Representations of Affixes in Compositional Semantic Space (FRACSS) approach (Marelli & Baroni, 2015) conceives each affix as a linear function that can be applied to the semantic vector of a stem to modify its meaning. The choice of adopting CAOSS instead of FRACSS (Marelli & Baroni, 2015) is motivated by two reasons. First, the former requires a more limited training set than the latter; second, the possibility offered by CAOSS of having one single common computational mechanism underlying affixation – rather than a different function for each affix (as in the FRACSS) – is more parsimonious from a cognitive standpoint.

In the context of the CAOSS framework, affixation could be regarded as a specific type of compounding applied to morphemic units (a stem and an affix, as opposed to two free-standing words), deriving a semantic vector for an affixed pseudo-word as $c = S \ast u + A \ast v$ (Fig. 1b). In this context, *u* and *v* are the semantic vectors of the stem and affix, respectively, and *S* and *A* are weight matrices applied to the *u* and *v*, capturing the semantic change that a lexical element undergoes when it assumes the role of Stem or the role of Affix[2] (see Lazaridou et al. (2013) for a proof of concept for the feasibility of a combinatorial approach in the context of deriving meaning from affixed pseudo-words).

Our CAOSS-inspired model involves a training phase and a testing phase. The goal of the training phase is to leverage familiar representations of stems, affixes, and affixed words to estimate the *S* and *A* weight matrices. During the training stage, the values of S and A are optimized so that if the semantic vector of an existing stem (e.g., *explore*) multiplied by S is summed to the semantic vector of an existing affix (e.g., *-ation*) multiplied by A, the result is as similar as possible to the semantic vector of the existing affixed form (e.g., *exploration*). Once the model is trained on existing data, the resulting S and A matrices can be used to generate meaning for novel combinations of stems and affixes (affixed pseudo-words).

---

[2] This notation (S and A) is proposed here to refer to weight matrices of Stems and Affixes, respectively. The reader should be mindful that, from a computational point of view, they don't differ from the originally proposed M and H weight matrices (Marelli et al., 2017).

The training list, including word-stem-affix pairs, was obtained from the *DerIvaTario* (Talamo et al., 2016), a corpus of Italian affixed forms annotated for morphological, morpho-syntactic, and morpho-semantic features. From the initial set of 11,147 items entries, three items in which affix information was missing were discarded. Entries with more than one affix (e.g. *antiproibizionista* [anti- + proibizion + -ista], antiprohibitionist; 5157 entries) and where stem information was missing (594 entries) were excluded. Entries with prefixes (491 entries) were also excluded. One entry (*polemico*, polemical), in which a conversion from noun (*polemica*, polemic) occurred, without any recognizable affixation process, was excluded as well. At this stage, a set of items involving a total of 41 different affixes was obtained. Given that the affixes *-aggine* and *-itudine* have been adopted to create experimental stimuli for the behavioral task, but were not part of this pool, words with these affixes were manually added to the list. To do so, among the top most frequent 200,000 entries in the Italian Fasttext semantic space, items ending in *-aggine* and *-itudine* were identified, and entries with a recognizable stem + affix structure were included in the training set. Double entries (29 items) were excluded from the set, together with compound words (22 items). To avoid the inclusion of items composed by a stem pseudo-word, the correctness of infrequent stems was manually checked. Raw frequency was derived from the SUBTLEX-IT corpus (Crepaldi et al., 2015). Among entries with raw frequency ≤ 10, 65 wrongly identified stems were manually corrected. Correctness of affixation was manually checked and the affix for 2 entries were manually reassigned (*settarismo*, (sectarianism): original affix = ario, corrected affix = ismo; *intuitivamente* (intuitively): original affix = ivo, corrected affix = mente). Items with a non-intuitive relationship[3] between stem, affix, and affixed word (including words derived from Latin and Greek words) were excluded (59 stems). Four incorrect entries were further excluded. At this step we computed affix frequency by summing (for each affix) the frequencies (derived from the SUBTLEX-IT) of the entries in the set. An overview of stem frequency and of the possible affixes for each stem included in the behavioural study can be found in the Supplementary Material (Table S1). Out of the remaining items, only those present among the first 200,000 entries in the Italian Fasttext semantic space were included. This step finally narrowed down the training set to 3,570 words. A total of 41 affixes were present; we took into consideration those (36) involved in more than 5 word examples. The list of affixes with more than 5 items is reported in Table 1.

These affixes were included in the final training set, which involved a total of 3,551 word entries. In line with the method proposed by Westbury & Hollis (2019), for each affix, the associated semantic vector was then computed by averaging the semantic vectors of all the words sharing that particular affix from the first 200,000 entries of the Italian Fasttext semantic space. One entry (*scrivenza*), among the first 200,000 entries of the Italian Fasttext semantic space was deemed as non-existent, as testified by its absence in the SUBTLEX-IT corpus, and thus manually excluded from the semantic space.

A new semantic space was then obtained, by merging the 199,999 vectors of the Italian Fasttext semantic space with the 36 vectors of the affixes. The combinatorial system was trained on the 3,551-words training set by using the *full additive* option available in DISSECT (Dinu & Baroni, 2013). The default least square regression optimization approach was adopted. Combination was then applied to the affixed pseudo-words in order to induce their meaning representations. For the affix *-eria,* included in experimental stimuli, two different options were available according to the notation of the *DerIvaTario* corpus (*-1eria* and *-2eria*). In order to model the *-eria* affix used in the behavioral experiment, the -2eria affix was used because of greater productivity in the training set (44 items for *-2eria* vs. 8 items for *-1eria*).

It is worth noting that the *combination-based* approach does not introduce novel, external information (relative to the *ngram* approach) for affixes in semantic vectors. Rather, semantic information for affixes used in the *combination*-based approach is derived (through averaging, on the basis of Westbury and Hollis, 2019) from the *ngram*-based semantic vectors themselves. In other words, the combination-based approach simply arranges *ngram*-based semantic information according to explicitly defined linguistic units (i.e., morphemes). This means that any performance difference between models is not due to the compositional one having received additional information at the ngram level compared to the form-meaning-mapping approach.

3.3. Model-based measures

The main rationale behind data analyses was to determine whether, while inducing the meaning of affixed pseudo-words, the appreciation of the meaning information conveyed by morpheme combination (*combination-based* approach) provides any additional advantage relative to form-meaning regularities from sublexical units which are not necessarily morphologically defined (*ngram-based* approach).

For this reason, for each dependent variable, comparisons were performed between a regression model including only semantic estimates obtained by means of the *ngram* approach (*baseline* model) and a regression model including the estimates of the baseline model (derived from the *ngram* approach) together with the same estimates derived from the *combination* approach (*test* model). The choice of using the *ngram* approach as a *baseline* was guided by the principle of parsimony: in both approaches, chunks of orthographic information are used as gateways towards semantic information. However, the combinatorial approach also involves appreciation of linguistic patterns (stems and affixes) that could in principle already be implicitly captured by the *ngram* approach: it is hence important to test whether the combinatorial model predictors explain variance in behavioral data over and above the *ngram*-based predictors.

To test the effects of the two models on behavioral data, the following semantic variables were computed: *stem proximity, semantic neighborhood density, entropy,* and *vector magnitude.* In line with Marelli and Baroni (2015), *stem proximity* was computed as the cosine

---

[3] e.g., in "bellico" = "military", the stem "bell-" comes from the latin "bellum" = "war" but it's not a word in Italian.

Table 1
List of affixes included in the training set and number of word examples for each affix. Affixes preceded by a number (1 or 2) indicate that there are different possible applications of the same affix, according to the notation included in the *DerIvaTario* (Talamo et al., 2016).

Affix	Examples	Affix	Examples	Affix	Examples
-1aio	19	-eggiare	36	-ità	169
-1eria	8	-etto	120	-itudine	10
-1ino	139	-evole	34	-izia	10
-2aio	13	-ezza	87	-izzare	48
-2eria	44	-ico	270	-mente	256
-2ino	31	-iera	42	-mento	146
-accio	12	-iere	58	-nza	26
-aggine	6	-iero	15	-oso	201
-aggio	19	-ificare	29	-tore	247
-ale	356	-ismo	115	-torio	22
-ario	103	-ista	166	-tura	64
-bile	122	-istico	28	-zione	480

similarity between the vector of an affixed pseudo-word and the vector of its stem. This variable was included in analyses to quantify the extent to which readers leverage semantic information derived from the stem to infer the meaning of the novel word. More specifically, both a linear and a quadratic term for *stem proximity* were included. Indeed, Marelli and Baroni (2015) observed that in a task in which ratings of semantic acceptability were asked (i.e., how easy it is to assign a meaning) for affixed pseudo-words, the highest semantic acceptability was detected at intermediate scores of stem proximity. According to Marelli and Baroni, such quadratic effect is due to the fact that attributing meaning to an affixed word is particularly difficult both when stem proximity is particularly low (i.e., when the relationship between stem and affixed word is obscure: e.g., in *windowist*), and particularly high (i.e., relative to the meaning of the stem, semantic information provided by the affix is just redundant; e.g., in *attorneyist*).

The variable *semantic neighborhood density* provides a measure of how densely populated the region of semantic space in which the target item is located is. The underlying assumption is that a meaningful new concept should be located in a portion of semantic space populated by a rich set of related concepts that already have an existing lexical counterpart, while a meaningless new concept should be far from any concept meaningful enough to have been lexicalized. This variable is operationalized as the average semantic (cosine) similarity between the semantic vector of an item and those of its closest semantic neighbors, with the number of closest neighbors constituting a free parameter. Also in line with Marelli and Baroni (2015), we computed *semantic neighborhood density* as the average cosine similarity between a given semantic vector and its 10 closest semantic neighbors. Cosine similarity and semantic neighbors were obtained through the R package *LSAfun* (Günther et al., 2015).

The variable *entropy* provides a measure of the distribution across different (subsymbolic) semantic dimensions. From a mathematical point of view, the entropy of a vector is lower when it has a skewed distribution with just few dimensions having large values, higher when the distribution tends to be uniform (Cover & Thomas, 2006). In other words, *entropy* measures the degree of uniformity of the distribution of weights across semantic dimensions for a given item. If vector loading is concentrated in a limited number of semantic dimensions, then that stimulus is expected to have a well-defined meaning (low entropy). On the contrary, if vector loading is uniformly distributed across a wide number of semantic dimensions (high entropy), then that stimulus does not convey any specific meaning (i.e., the meaning is uncertain). Previous evidence suggests that entropy is negatively related to the degree of meaningfulness of affixed pseudo-words (Marelli & Baroni, 2015).

Given that semantic vectors contained negative values, the formula for entropy was not directly applicable. For this reason, exclusively for the computation of entropy, vectors were transformed as follows: first, each element of the vector was summed to the absolute value of the minimum entry contained in the vector and to a constant representing the inverse of the number of vector dimensions. Secondly, for the obtained vector, each element was divided by the sum of all vector entries. This transformation yielded a vector of proportions summing to 1, analogous to a probability distribution. Entropy was then calculated on transformed vectors (being d_i the ith element of a transformed semantic vector $(t(\vec{v}))$ and N being vector dimensions) as:

$$entropy(t(\vec{v})) = -\sum_{i=1}^{N} d_i log(d_i)$$

It is worth highlighting that entropy is typically computed with reference to vectors of frequencies (or proportions), while semantic vectors indicate coordinates in a semantic space. Yet, each semantic vector can be conceived as proportional to a vector that indicates how likely it is to find semantic information for that stimulus in any of the semantic dimensions (which can sometimes be conceptualized as different semantic domains or "topics") taken into account. As it happens, this latter vector is exactly the result of the preliminary transformation we applied to semantic vectors prior to computation of entropy and is consistent with Marelli and Baroni's (2015) interpretation of entropy as a measure of topic-specificity of a concept. From an information theory standpoint, a highly entropic semantic vector indicates high uncertainty on the semantic domains carrying the meaning for that stimulus.

The variable *vector magnitude* was also included. It has been previously used in the domain of semantic combination at the phrase level and it proved to be able to account for human performance in semantic judgements for adjective-noun combinations (Vecchi et al., 2017). It represents semantic loading across dimensions in a vector and, from a mathematical point of view, it is computed as the Euclidean norm of the vector x (with n being the number of semantic dimensions):

$$\|x\| = \sqrt{x_1{}^2 + x_2{}^2 + \cdots + x_n{}^2}$$

From a geometric point of view, if we consider a semantic vector as corresponding to the coordinates of a point in the semantic space, vector magnitude represents the length of the arrow whose tail is in the origin of the space (i.e., the point in space in which all dimensions are equal to zero) and whose tip is in the point in space corresponding to the coordinate of interest. For example, in a toy semantic space composed of three dimensions, a word with a semantic vector of [0.2 1 2.5] will have a magnitude equal to $\sqrt{0.2^2 + 1^2 + 2.5^2} = \sqrt{0.04 + 1 + 6.25} = \sqrt{7.29} = 2.7$.

Finally, in all models, we introduced (log-transformed) stem frequency (Burani et al., 1984; Burani & Thornton, 2011), (log-transformed) affix frequency (Burani & Thornton, 2003), and word length (Marelli & Baroni, 2015; Yap et al., 2015). Item order was also included as a predictor to rule out the possibility that any result of potential interest could be due to within-task learning dynamics. Table 2 reports the correlations between the semantic predictors used in analyses.

3.4. Analyses

Trials in which reaction times were longer than 2,500 ms (123 trials, 4.24%) were discarded. In none of the trials the RT was – according to the neurophysiology of reading processes (Cohen et al., 2000; Dehaene, 1995; Kuriki et al., 1998) – implausibly short (the shortest recorded reaction time was 417 ms). For each participant, separately for each trial, response choice in dichotomous form (1 = word, 0 = non-word) and log-transformed RTs were employed as dependent variables in a series of mixed-effects models. Logarithmic transformation (Baayen, 2008) was adopted to obtain a better approximation to a Normal distribution than that of raw RTs. Subjects and stimuli were modeled as random intercepts. Response choice data were analyzed by means of logistic regressions.

As far as RTs are concerned, two different sets of models were fit: one for affixed pseudo-words classified as words ("YES" responses) and one for affixed pseudo-words classified as "non-words" ("NO" responses). Therefore, for each of the two models (namely, the *baseline* [ngram] and the *test* [ngram + combination] approach), three analyses were performed: a model on response choice, a model on (log-transformed) RTs for affixed pseudo-words classified as words ("YES" responses), and a model on (log-transformed) RTs for affixed pseudo-words classified as non-words ("NO" responses). All analyses were performed by means of the statistical software R (4.0.3) and the lme4 package (Bates et al., 2015). For each dependent variable, analyses proceeded as follows: we first fitted the *baseline* model containing item order, stimuli length, affix frequency and stem frequency, plus the semantic variables density, entropy, magnitude, and stem proximity both as linear and quadratic term, with these variables obtained via an *ngram*-based approach. We subsequently fitted the *test* model, involving the same *ngram*-based variables of the *baseline* model plus the *combination*-based variables. Baseline and test models were compared by means of a likelihood ratio test, complemented by exploration of the Akaike Information Criterion (AIC). Whenever the *test* model provided a better fit than the *baseline* model, we ran ancillary analyses in which each semantic variable derived from the *combination*-based approach was separately entered as a single predictor in the *baseline* model. Each resulting model was then compared to the *baseline* model, with the aim of evaluating the degree of impact of each specific semantic feature emerging from the combinatorial procedure. In each model, we used the R package *performance* to compute the conditional R^2 (Nakagawa et al., 2017) and to check for multicollinearity.

The stimuli corresponding (for each ngram– and combination-based variable) to the maximum and the minimum value can be found in the Supplementary Material (Table S2). Values of ngram- and combination-based semantic variables can be found in the Supplementary Table S3.

Average cosine similarity between semantic vectors obtained by means of the *ngram-based* and *combination-based* approaches was 0.506 (SD = 0.085, Fig. 2). This suggests that, although the models have a certain degree of overlap in their predictions, they might capture rather different processes, and hence cannot be reduced to one another. Indeed, Spearman's rank correlation between the *ngram-based* and *combination-based* approaches was only significant ($\rho = 0.619$ p < 0.001) for the variable magnitude (density $\rho = 0.169$, p = 0.239; entropy $\rho = -0.128$, p = 0.374; stem proximity $\rho = 0.195$, p = 0.172). A pairwise scatterplot of the *ngram*-based and *combination*-based semantic variables used in the statistical models can be found in the Supplementary Material (Fig. S1).

4. Results

Across participants (i.e., after averaging by participant), the average probability of a "YES" response was 0.262 (SD = 0.146). In a complementary fashion, the average probability of a "NO" response was 0.734 (SD = 0.146). Across stimuli (i.e., after averaging by item), the average probability of a "YES" response was 0.264 (SD = 0.232). The average probability of a "NO" response was 0.736 (SD = 0.232). As shown in Table 3, the test models (including combination-based measures) significantly improved the model fit

Table 2
Spearman's rank correlations between semantic predictors.

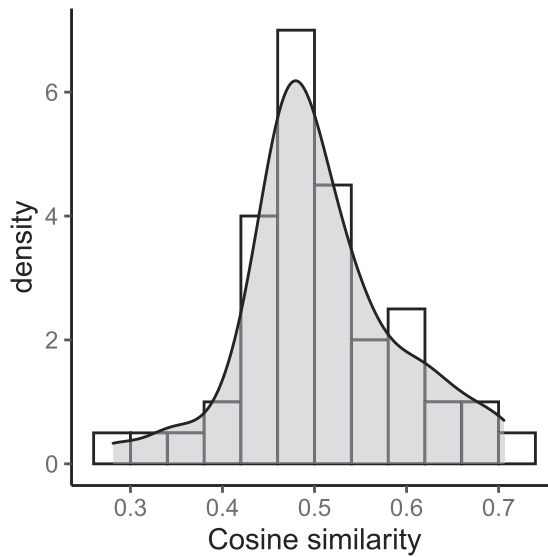|	Ngram – based predictors					combination – based predictors			
	Density	Stem proximity	Magnitude	Entropy		Density	Stem proximity	Magnitude	Entropy
Density	–				**Density**	–			
Stem Proximity	0.316	–			**Stem Proximity**	−0.030	–		
Magnitude	0.018	−0.069	–		**Magnitude**	0.152	0.486	–	
Entropy	−0.278	−0.137	−0.333	–	**Entropy**	−0.069	−0.101	0.043	–

Fig. 2. Distribution of the cosine similarity between semantic vectors obtained by means of the ngram-based and combination-based approaches.

Table 3
Results of the three model comparison procedures. The label "YES" responses refers to affixed pseudo-words classified as words; the label "NO" responses refers to affixed pseudo-words classified as non-words.

Dependent variable	Model	Parameters	AIC	logLikelihood	Deviance	Conditional R^2	Chisq	Df	Pr (>Chisq)
Response choice	baseline (ngram only)	12	2108.2	−1042.1	2084.2	0.521			
	test (ngram + combination)	17	**2102.2**	**−1034.1**	2068.2	0.533	15.916	5	**0.007**
log*(RT) "YES"*	baseline (ngram only)	13	395.18	−184.59	369.18	0.450			
response	**test (ngram + combination)**	18	**386.15**	**−175.07**	350.15	0.449	19.029	5	**0.002**
log*(RT) "NO"*	**baseline (ngram only)**	13	**157.34**	−65.672	131.34	0.535			
response	test (ngram + combination)	18	158.47	**−61.233**	122.47	0.537	8.878	5	0.114

(according to both the AIC and the likelihood ratio test criteria) when added to the baseline in the response choice analysis and in the analysis on RTs for "YES" responses. This finding suggests the significance of the added value of combination towards semantic access in this set of stimuli recognized as meaningful words. Conversely, the additional explicit information about morpheme combinations did not significantly improve model fit when considering the "NO" responses.

We thus conducted ancillary analyses. We compared – for response choice and RTs for "YES" responses – the *baseline* model with models containing the same variables of the *baseline* model plus one semantic variable derived from the combinatorial approach.

For what concerns response choice (Table 4), we observed that the *baseline* model plus the magnitude variable derived from the combinatorial approach was better than the *baseline* model alone according to the AIC and likelihood ratio test. None of the other variables significantly improved the model fit.

The best model was further explored, revealing a significant effect of (log) stem frequency (Chisq(1) = 9.887, p = 0.002), indicating that the more frequent the stem, the less affixed pseudo-words were classified as "words" (Fig. 3a). The linear effect of *ngram*-based entropy was also significant (Chisq(1) = 5.494, p = 0.019), indicating less frequent "word" responses for higher entropy values (Fig. 3b). The polynomial effect of *ngram*-based stem proximity was also significant (Chisq(2) = 10.102, p = 0.006). In particular, the quadratic effect was significant (Z = -2.849, p = 0.004, Fig. 3c), whereas the linear one was not (Z = 0.605, p = 0.545). Most crucially, these analyses revealed a significant negative effect of *combination*-based magnitude (Fig. 3d) on the probability of "word" responses (Chisq(1) = 14.121, p < 0.001).

For what concerns RTs for "YES" responses (affixed pseudo-words classified as words; Table 5), the *baseline* model plus the linear and quadratic effect of stem proximity from the combinatorial approach turned out to be better than the *baseline* model, according to the AIC and likelihood ratio test.

The best model was further explored. The effect of stem frequency was significant (Chisq(1) = 4.227, p = 0.040, Fig. 4a), indicating shorter RTs for more frequent stems. Similarly, the effect of affix frequency was significant (Chisq(1) = 7.529, p = 0.006, Fig. 4b), indicating shorter RTs for more frequent affixes. The polynomial (1st and 2nd order) component of *combination*-based stem proximity was also significant (Chisq(2) = 11.917, p = 0.003). In particular, the quadratic effect was significant (t(31.608) = 3.354, p = 0.002,

Table 4
Results of the ancillary model comparison procedure for Response Choice.

| Dependent variable	Model	Parameters	AIC	logLikelihood	Deviance	Conditional R^2	Chisq	Df	Pr (>Chisq)
Response choice	baseline (*ngram* only)	12	2108.2	−1042.1	2084.2	0.521			
	baseline + density (*combination*)	13	2110	−1042	2084	0.521	0.152	1	0.697
	baseline + entropy (*combination*)	13	2108.1	−1041	2082.1	0.520	2.061	1	0.151
	baseline + magnitude (*combination*)	**13**	**2096.3**	**−1035.2**	**2070.3**	**0.531**	**13.847**	**1**	**< 0.001**
	baseline + stem proximity (combination)	13	2107.4	−1040.7	2081.4	0.522	2.702	1	0.100
	baseline + stem proximity (combination) + stem proximityˆ2 (combination)	14	2108.1	−1040	2080.1	0.522	4.093	2	0.129
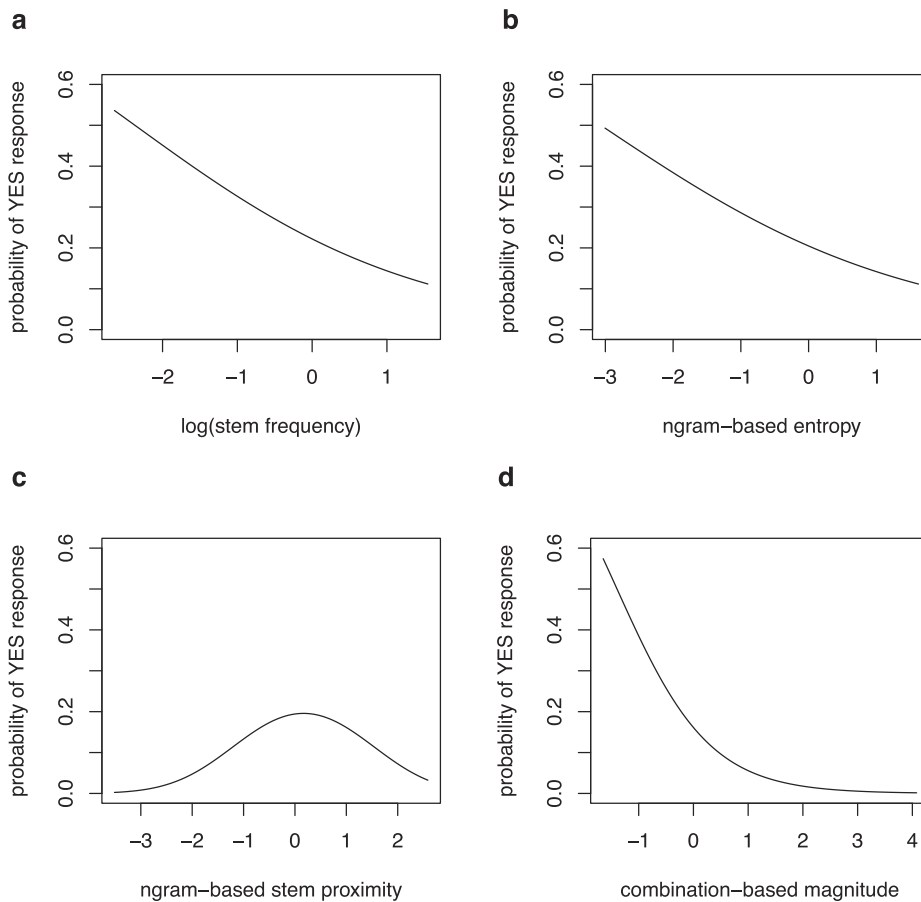
Fig. 3. Effects on the probability of a "YES" response (affixed pseudo-words classified as words) of the variables (**a**) log(stem frequency), (**b**) *ngram*-based entropy, (**c**) *ngram*-based stem proximity, (**d**) and *combination*-based magnitude. All predictors were scaled.

Fig. 4c), while the linear one was not (t(31,408) = -1.536, p = 0.134).

Since in the case of RTs for "NO" responses the *baseline* model (i.e., the one containing *ngram*-based semantic variables only) turned out to be better than the one that also included semantic variable derived from the combination-based approach, the significance of the factors included in the *baseline* model was further explored.

Such exploration revealed the significant effect of item order (Chisq (1) = 62.375, p < 0.001, Fig. 5a), the significant effect of stimulus length (Chisq(1) = 11.733, p < 0.001, Fig. 5b), and the significant effect of *ngram*-based entropy (Chisq(1) = 4.412, p = 0.036, Fig. 5c).

In none of the estimated models multicollinearity turned out to be an issue (Variance Inflation Factor < 10 for all effects in all analyses).

Table 5

Results of the ancillary model selection procedure for log(RTs) of "YES" responses (i.e., affixed pseudo-words classified as "words").

Dependent variable	Model	Parameters	AIC	logLikelihood	Deviance	Conditional R^2	Chisq	Df	Pr (>Chisq)
log(RT) "word" response	baseline (*ngram* only)	13	395.18	−184.59	369.18	0.450			
	baseline + density (*combination*)	14	393.72	−182.86	365.72	0.453	3.464	1	0.063
	baseline + entropy (*combination*)	14	396.79	−184.39	368.79	0.451	0.392	1	0.531
	baseline + magnitude (*combination*)	14	397.02	−184.51	369.02	0.453	0.158	1	0.691
	baseline + stem proximity (combination)	14	396.58	−184.29	368.58	0.451	0.601	1	0.438
	baseline + stem proximity (combination) + stem proximity^2 (combination)	**15**	**385.80**	**−177.90**	355.80	0.442	13.381	2	**0.001**
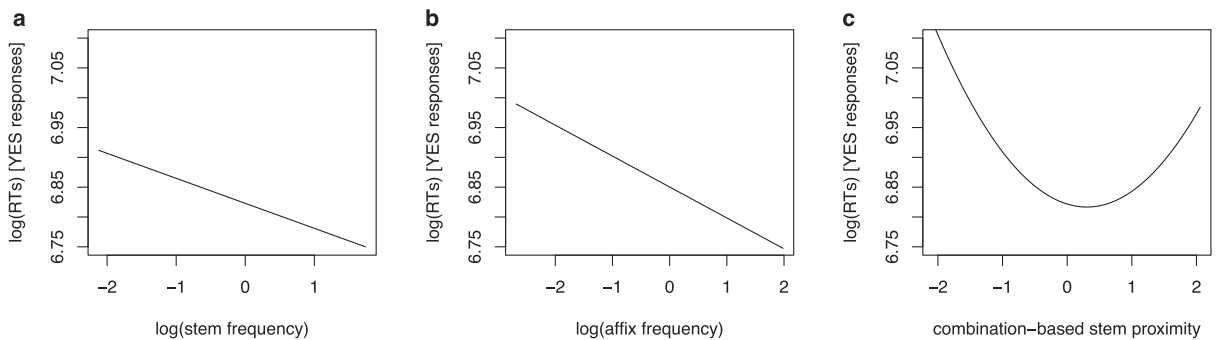
Fig. 4. Effects on log(RTs) for "YES" responses (i.e., affixed pseudo-words classified as "words") of the variables (**a**) log(stem frequency), (**b**) log (affix frequency), and (**c**) *combination*-based stem proximity. All predictors were scaled.
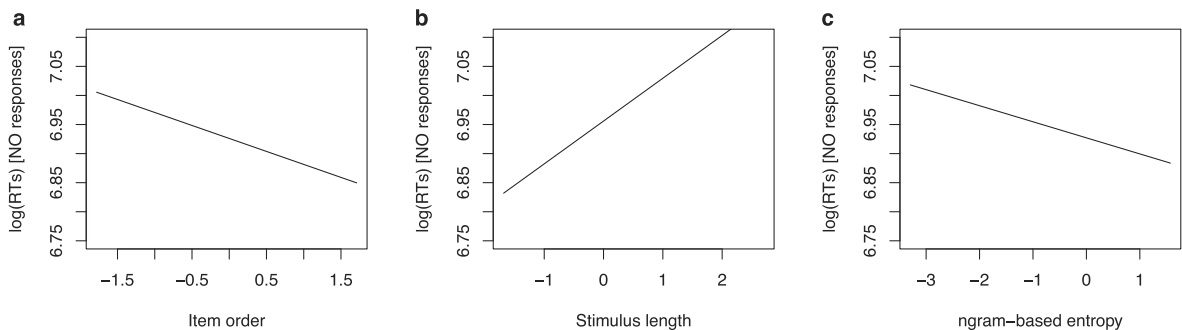
Fig. 5. Effects on log(RTs) for "NO" responses (i.e., affixed pseudo-words classified as "non-words") of the variables (**a**) item order, (**b**) stimulus length, and (**c**) *ngram*-based entropy. All predictors were scaled.

5. Discussion

When we come across a novel word constituted by an existing stem and an existing affix, such as in an affixed pseudo-word, a viable strategy to derive its meaning might be considering its constituents. However, although psycholinguistic research has suggested that morphology plays a role in this regard (Burani et al., 1997; Crepaldi et al., 2010; Taft & Forster, 1975, 1976), there is also evidence suggesting that orthographic cues can be sufficient to access semantic representations and that morphological mediation is not necessary (e.g., Harm & Seidenberg, 2004; Baayen et al., 2011). In this latter perspective, morphological effects would be a pure epiphenomenon of general-level form-meaning mapping processes.

In the current study, we built on computational models to explore the extent to which, relative to direct access to semantics from orthographic cues, morpheme combination can provide a sizable contribution while deriving the meaning of an affixed pseudo-word. In particular, we considered two models: a *ngram*-based approach – derived from Fasttext (Bojanowski et al., 2017) – in which meaning is conveyed by orthographic chunks (*ngrams*), and a *combination*-based approach (Marelli et al., 2017), according to which the meaning of an affixed pseudo-word is derived by combining semantic information from the stem with semantic information from the affix. Due

to its more parsimonious nature, we took the former model as a baseline, and evaluated to what extent the additional consideration of morphological combination improved the fit of the statistical models aiming to explain human performance in a non-speeded lexical decision task. To do so, we included in the *baseline* model the effects of item order, stimuli length, affix frequency and stem frequency, and a set of semantic variables (neighborhood density, entropy, magnitude, stem proximity, and a first- and second-order stem proximity effect) derived by means of the *ngram* approach. The *test* model included the same variables included in the *baseline* model plus neighborhood density, entropy, magnitude, stem proximity (first and second order effect) computed via the *combination*-based approach.

Among the non-semantic variables included in the models, stem frequency, affix frequency, item order and stimulus length turned out to have a significant effect (although these effects were not consistently significant across models). In particular, stem frequency was negatively associated with the proportion of "YES" responses, although it turned out to have a facilitating effect on RTs in case of "YES" responses. This could indicate that familiar stems induce a fast, intuitive "YES" response. When this early response is successfully inhibited, stem frequency then contributes to unravel the implausibility of novel stem + affix combinations. Affix frequency turned out to be associated with faster RTs in the case of "YES" responses as well. In the case of RTs for "NO" responses, item order proved to have a significant impact: affixed pseudoword stimuli occurring later in the task were associated with faster "NO" responses. Conversely, stimuli length had a positive effect on RTs for "NO" responses: longer words were associated with longer RTs. These results on non-semantic variables highlight the existence of different processes underlying acceptance and rejection of affixed pseudowords as word stimuli.

Semantic variables derived from both the *ngram*-based and *combination*-based approaches variables proved to have explanatory value towards human performance, as demonstrated by their significance across different models. In particular, the *ngram*-based semantic variables that provided significant contributions were entropy and stem proximity. The *ngram*-based entropy (which measures how semantic information is uniformly distributed across semantic dimensions) was associated with higher probability and faster RTs for "NO" responses (i.e., affixed pseudo-words classified as "non-words"). This suggests that when semantic information extracted from letter chunks is more uncertain, participants tend to consider an incoming orthographic stimulus as a non-word. In other words, if a letter string could mean anything – according to the perceived meaning of the chunks of letters that compose it -, then it is probably not a word. This result appears in line with previous evidence indicating that letter strings whose semantic vector is more entropic (i.e., more uniformly distributed) tend to be associated with lower ratings of perceived meaningfulness (Marelli & Baroni, 2015). Yet, the present finding extends this proposal, by suggesting this effect to be particularly relevant for non-morphemic form-to-meaning associations. Indeed, in none of the models the addition of *combination*-based entropy resulted in a better fit than the baseline model containing *ngram*-based entropy.

The significant quadratic effect of *ngram*-based stem proximity (i.e., the semantic resemblance between the full affixed non-word and its stem) indicates lower probability of a "YES" response (i.e., an affixed pseudo-word classified as a "word") if the stem and the full form are either too similar or too distant from a semantic point of view. In line with the results from Marelli and Baroni (2015), our finding indicates that when the semantic relation between the affixed pseudo-word and its stem is either obscure (low stem proximity), or redundant (high stem proximity), the stimulus is less likely to be considered as a meaningful word.

For what concerns *combination*-based semantic variables, a quadratic effect of stem proximity was also detected for RTs, yet limited to affixed pseudo-words classified as "words": longer RTs were observed for extreme values of *combination*-based stem proximity. This suggests that the appreciation of the semantic relationship between an affixed pseudo-word and its stem might be a feature shared by the form-to-meaning (i.e., *ngram*) level of analysis and morphological parsing.

Within the context of *combination*-based semantic variables, a significant effect of magnitude was also detected. Indeed, *combination*-based magnitude (loading across semantic dimensions) was associated with higher probability of "NO" responses (i.e., an affixed pseudo-word classified as a "non-word"). The interpretation of this effect is made difficult by the fact that normative values of semantic vector magnitude for existing concepts are hard to determine. However, it is possible that, while considering an affixed pseudo-word as composed of semantic information derived from the stem and semantic information derived from the affix, widespread semantic loading across dimensions might lead to a "non-word" response. What is important to acknowledge is that this effect seems to be specific for the morphological analysis of stimuli. In this regard, it may be possible to conceive *combination*-based magnitude as a proxy for the semantic complementarity (i.e. non-redundancy) of the stem and affix of a given stimulus. Indeed, let's consider a "toy semantic space" that is only composed of two dimensions, with vector loadings for the stem being equal to (for instance) 4 and 1, respectively. If the vector of the affix has identical loadings on the same semantic variables (i.e. 4 and 1, respectively), once the semantic vectors of the stem and affix are summed,[4] the magnitude is equal to $\sqrt{(4+4)^2 + (1+1)^2} = 8.246$. Instead, if the affix has identical loadings but their distribution between factors is different from that of the stem (i.e. 1 and 4), then the magnitude is more limited, and precisely $\sqrt{(4+1)^2 + (1+4)^2} = 7.071$. If the same reasoning is extended to multidimensional semantic vectors, it appears evident that if semantic information from stem and affix is redundant (producing a greater vector magnitude), then it is reasonable to expect the resulting affixed pseudo-word to be more likely recognized as a non-word. This is in line with evidence showing that novel derived words whose meaning is very close to their stems are perceived as less meaningful by speakers of a language (Marelli & Baroni, 2015): a novel word is in fact expected to denote an at least partially novel meaning – redundant information in such elements is hence perceived as implausible.

---

[4] *combination*-based semantic vectors were obtained via (weighted) summation of the semantic vectors of the stem and of the affix in this work.

This result is interesting to interpret in light of the effect of *ngram*-based entropy. Indeed, while entropy – that can be considered as a measure of how scattered (uniformly distributed) semantic information is across different dimensions – plays a role when smaller orthographic-semantic units are considered (*ngrams*), vector magnitude is determinant when a higher-level combination of semantic information between stem and affixes is taken into account. However, both measures provide a summary of semantic activation across different dimensions. This suggests that, when we encounter an affixed pseudo-word during a non-speeded lexical decision task, the decision process is guided by mechanisms that determine overall loading/dispersion across the different semantic dimensions, although the weight of these processes changes depending on the level of analysis (*ngrams* or *morphemes*) being used, with a progressive shift from a more fine-grained (i.e., entropy) to a coarser (i.e. magnitude) index of overall semantic loading.

More broadly, relative to the baseline (*ngram*) model, our results show that the further inclusion of semantic variables from the *combination*-based approach improved the fit to behavioral data on response choice and on RTs for "YES" responses. This finding can be read in light of the debate in morphological processing (for recent reviews, see Marelli, Traficante, & Burani, 2020; Marelli, 2023) on the relationship between morphological and semantic information. In this regard, previous evidence that morphological priming occurs for complex words even in the absence of semantic relationship between prime and target (e.g., *corner* and *corn*) suggested that morphological decomposition might be a "semantically blind" process (Rastle & Davis, 2008). The present findings complement this view, by pointing out that even if morphemes are recognized based on their pure orthographic regularity, still they (stems and affixes alike) guide the extraction of meaning (Amenta, Marelli & Crepaldi, 2015), especially when the stimulus is treated as a real, meaningful word. Crucially, this would be accomplished by combining meanings conveyed by the activated morphemes.

However, the nature of this process seems to be different when the incoming stimulus is perceived as a "non-word". In that case, responses from the participants are not affected by combination-based variables. This suggests that the underlying process, in such a scenario, might rely on an automatic analysis driven by the surface level properties of the stimulus, and their semantic informativity as captured by the ngram approach. This is in line with evidence showing that embedded lexical elements can make a nonword more difficult to reject, even in the absence of morphological structure (e.g., Taft and Forster, 1976). Conversely, it is plausible that a deeper process, focused on the actual combination of morphemic units, would be engaged when the stimulus is perceived as an existing, and hence meaningful, element (i.e., in the case of "YES" responses).

So, how do we extract meaning from affixed pseudo-words? The present data suggest that, on the one hand, we look for semantic clues in letter chunks, regardless of their linguistic status and of the structure of the stimulus, and such clues can be sufficient to process that stimulus as meaningless. On the other hand, we do further explore stimulus structure and leverage semantic information contained in morphemes when it is available, and the contribution of this step is substantial to process the incoming stimulus as a word. Most importantly, by refraining from an aut-aut logic (that would have demanded a simple and direct comparison between the *ngram* and *combination*-based approaches to find the model best accounting for human data), our study was able to unravel how, rather than being mutually exclusive, form-meaning and combinatorial strategies seem to complement each other. This is testified here by the significance of semantic variables extracted with both *ngram* and *combination*-based approaches, and is in line with recent interpretations of advocating – in the transition between a naïve and an expert reading system – early recruitment of domain-general language-agnostic mechanisms (i.e., extraction of statistical regularities in word forms; Treiman et al., 2021; Ulicheva et al., 2021), and later reliance on linguistic representations of word chunks (De Rosa & Crepaldi, 2022). From this standpoint, form-meaning and combinatorial processes can be seen as both rooted in statistical learning, with the first one mirroring general-purpose learning, and the latter embodying the emergence of combinatorial operations specific to salient elements emerging from linguistic distributions (Stevens & Plaut, 2022). This view allows the reconciliation of results indicating the existence of domain-general semantic mapping processes (Hendrix & Sun, 2021), with those suggesting a specific morphological mediation via combinatorial strategies (Marelli & Baroni, 2015).

Still, the present data do not provide direct clues on how the temporal dynamics of these processes might unfold. A parallel interpretation would imagine these two processes to operate simultaneously throughout meaning extraction. Conversely, a sequential interpretation would entail that form-meaning strategies are used in earlier stages in order to have a sense of whether the target stimulus has a somewhat defined meaning or it could "mean anything" (effect of entropy). Then, the consideration of the relationship between parts of the strings and its full form (effect of stem proximity) might represent a bridge between word-form and morphology-based strategies, with this latter ultimately adopted to appreciate how non-redundant (effect of magnitude) the meaning obtained by combining semantic information form the stem and that of the affix is. Future integrations of the approach outlined in this work with experimental techniques best suited to explore the temporal dynamics of morphological processing (e.g., priming studies with different SOA, eye tracking, Electro- and Magneto-Encephalography) will be most welcomed to shed further light on the nature of dynamic interplay between form-meaning and combinatorial processes. Additionally, functional Magnetic Resonance Imaging studies, combined with multivariate representational similarity analysis, may be helpful in teasing apart the relative importance of form-meaning (ngram-based) versus morphology-meaning (combination-based) information coding at the neural level.

It is important to highlight the fact that -per se- a lexical decision task does not provide a pure measure of semantic access. Even so, the present data themselves provide quantitative empirical evidence (i.e., the significant effect of semantic variables on performance) that semantic access did indeed take place. In line with this, there is substantial evidence in the literature on semantic effects emerging in lexical decision tasks (see for instance Joordens & Becker, 1997; Perea & Rosa, 2002; Scaltritti et al., 2021; Sulpizio et al., 2021).

As a final remark, it is worth highlighting that the present results raise the need for computationally implemented models of human semantic knowledge to include a morphological parsing module. Since their earliest development, computational models of semantic memory have tried to develop more and more reliable representations, by leveraging (more or less direct) measures of word co-occurrence derived by large linguistics corpora (Lund & Burgess, 1996; Landauer & Dumais, 1997). In a narrow sense, these models might be unable to account for language productivity phenomena (Marelli, 2019). To put it more directly, they may simulate at

most a human unable to generalize learnt semantic regularities to novel, untrained linguistic elements. The *ngram*-based and the *combination*-based models taken into account in the present work address this problem by applying combinatorial operations in the semantic space either by leveraging semantic information provided by chunks of letters (*ngrams*), or the meaning of morphemes. The observation that combining the two approaches provides the best approximation to human performance (especially when affixed pseudowords are recognized as words) suggests that, aside *ngram*-based methods, future computational models of semantic knowledge should also include combination as a complementary tool to address language productivity.

6. Conclusions

The present findings suggest that, in the context of lexical decisions on affixed pseudo-words, word-form-to-meaning (*ngram*) mechanisms and processes of extraction (*combination*) of semantic information contained in morphemes are interwoven: behavioral performance was predicted by the degree of uncertainty (i.e. entropy) of semantic information at the *ngram* level and by the degree of overall semantic loading (i.e. magnitude) at the *combinatorial* level. The effect of the semantic similarity between stem and composed form turned out to be shared by the two mechanisms. In summary, our data suggest that, on top of the cognitively more parsimonious word-form-to-meaning (*ngram*) access mechanism, combinatorial processes mediated by morphological units play a prominent role in inducing the meaning of affixed pseudo-words, especially when these are recognized as words. This evidence draws the attention to the fact that morphological effects that can be observed in reading are (also) inherently semantic in nature, rather than solely lexical.

CRediT authorship contribution statement

Rolando Bonandrini: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Simona Amenta:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **Simone Sulpizio:** Conceptualization, Writing – original draft, Writing – review & editing. **Marco Tettamanti:** Conceptualization, Writing – original draft, Writing – review & editing. **Alessia Mazzucchelli:** Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Marco Marelli:** Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code used in this work is available at: https://osf.io/qp2ad/.

Acknowledgements

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2023.101594.

References

Amenta, S., Foppolo, F., Badan, L. (under review). The role of morphological information in processing pseudo-words in Italian L2 learners: It's a matter of experience.

Amenta, S., Günther, F., & Marelli, M. (2020). A (distributional) semantic perspective on the processing of morphologically complex words. *The Mental Lexicon, 15*(1), 62–78. https://doi.org/10.1075/ml.00014.ame

Amenta, S., Marelli, M., & Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: Early morpho-orthographic and morpho-semantic effects in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(5), 1587. https://doi.org/10.1037/xlm0000104

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801686.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on I discriminative learning. *Psychological Review, 118*(3), 438. https://doi.org/10.1037/a0023851

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9 https://CRAN.R-project.org/package=lme4.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics, 5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Burani, C., & Thornton, A. M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. *Trends in Linguistics Studies and Monographs, 151*, 157–208. https://doi.org/10.1515/9783110910186

Burani, C., & Thornton, A. M. (2011). The interplay of root, suffix and whole-word frequency in processing derived words. In *Morphological structure in language processing* (pp. 157–208). De Gruyter Mouton.

Burani, C., Dovetto, F. M., Thornton, A. M., & Laudanna, A. (1997). Accessing and naming suffixed pseudo-words. In *Yearbook of morphology 1996* (pp. 55-72). Springer. https://doi.org/10.1007/978-94-017-3718-0_5.

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access. *Visible Language, 18*(4), 342–352.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics, 16*, 22–29.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*(Pt 2), 291–307. https://doi.org/10.1093/brain/123.2.291

Cover, T., & Thomas, J. (2006). Elements of information theory, USA: Jonh Wiley & Sons. In: Inc. https://doi.org/10.1002/047174882X.

Crepaldi, D., Amenta, S., & Marelli, M. (2019). For a probabilistic and multidisciplinary approach to the investigation of morphological processing. *Cortex, 116*, 1–3. https://doi.org/10.1016/j.cortex.2019.05.010

Crepaldi, D., Amenta, S., Pawel, M., Keuleers, E., & Brysbaert, M. (2015). SUBTLEX-IT. Subtitle-based word frequency estimates for Italian. In *Proceedings of the Annual Meeting of the Italian Association For Experimental Psychology* (pp. 10-12).

Crepaldi, D., Rastle, K., & Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition, 38*(3), 312–321. https://doi.org/10.3758/MC.38.3.312

De Rosa, M., & Crepaldi, D. (2022). Letter chunk frequency does not explain morphological masked priming. *Psychonomic bulletin & review, 29*(2), 589–599. https://doi.org/10.3758/s13423-021-02010-y

Dehaene, S. (1995). Electrophysiological evidence for category-specific word processing in the normal human brain. *Neuroreport, 6*(16), 2153–2157. https://doi.org/10.1097/00001756-199511000-00014

Dinu, G., & Baroni, M. (2013). Dissect-distributional semantics composition toolkit. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

Drews, E., & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 1098. https://doi.org/10.1037/0096-1523.21.5.1098

Gatti, D., Marelli, M., & Rinaldi, L. (2022). Out-of-vocabulary but not meaningless: Evidence for semantic-priming effects in pseudoword processing. *Journal of Experimental Psychology: General.*. https://doi.org/10.1037/xge0001304

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893*. https://doi.org/10.48550/arXiv.1802.06893.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*(2), 211. https://doi.org/10.1037/0033-295X.114.2.211

Guevara, E. (2010). A regression model of adjective-noun compositionality in distributional semantics. In *Proceedings of the 2010 Workshop on Geometrical Models of Natural Language Semantics*.

Günther, F., & Marelli, M. (2019). Enter sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(10), 1872. https://doi.org/10.1037/xlm0000677

Günther, F., & Marelli, M. (2020). Trying to make it work: Compositional effects in the processing of compound "nonwords". *Quarterly Journal of Experimental Psychology, 73*(7), 1082–1091. https://doi.org/10.1177/1747021820902019

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun-An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods, 47*(4), 930–944. https://doi.org/10.3758/s13428-014-0529-0

Günther, F., Marelli, M., & Bölte, J. (2020). Semantic transparency effects in German compounds: A large dataset and multiple-task investigation. *Behavior Research Methods, 52*(3), 1208–1224. https://doi.org/10.3758/s13428-019-01311-4

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science, 14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological review, 111*(3), 662. https://doi.org/10.1037/0033-295X.111.3.662

Harris, Z. S. (1954). Distributional structure. *Word, 10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(1), 157. https://doi.org/10.1037/xlm0000819

Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1083. https://doi.org/10.1037/0278-7393.23.5.1083

Kuriki, S., Takeuchi, F., & Hirata, Y. (1998). Neural processing of words in the human extrastriate visual cortex. *Brain Research Cognitive Brain Research, 6*(3), 193–203. https://doi.org/10.1016/s0926-6410(97)00030-x

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211. https://doi.org/10.1037/0033-295X.104.2.211

Lazaridou, A., Marelli, M., Zamparelli, R., & Baroni, M. (2013). Compositional-ly derived representations of morphologically complex words in distributional semantics. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*(2), 203–208. https://doi.org/10.3758/BF03204766

Marelli, M. (2019). Quantitative Methods in Morphology: Corpora and Other "Big Data" Approaches. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.602.

Marelli, M. (2023). The role of semantics in the processing of complex words. In D. Crepaldi (Ed.), *Linguistic Morphology in the Mind and Brain* (pp. 65–79). Routledge.

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review, 122*(3), 485. https://doi.org/10.1037/a0039267

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as Abstract Operation in Semantic Space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition, 166*, 207–224. https://doi.org/10.1016/j.cognition.2017.05.026

Marelli, M., Traficante, D., & Burani, C. (2020). Reading morphologically complex words: Experimental evidence and learning models. *Word Knowledge and Word Usage*, 553–592.

Martin, D. I., & Berry, M. W. (2007). Mathematical foundations behind latent semantic analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 35–56). Mahwah, NJ: Erlbaum.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546*.

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science, 34*, 1388–1439. https://doi.org/10.1111/j.1551-6709.2010.01106.x

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface, 14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

New, B., Brysbaert, M., Segui, J., Ferrand, L., & Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language, 51*(4), 568–585. https://doi.org/10.1016/j.jml.2004.06.010

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research Psychologische Forschung, 66*(3), 180–194. https://doi.org/10.1007/s00426-002-0086-5

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23*(7–8), 942–971. https://doi.org/10.1080/01690960802069730

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes, 15*(4–5), 507–537. https://doi.org/10.1080/01690960050119689

Scaltritti, M., Job, R., & Sulpizio, S. (2021). Selective suppression of taboo information in visual word recognition: Evidence for cognitive control on semantics. *Journal of Experimental Psychology: Human Perception and Performance, 47*(7), 934. https://doi.org/10.1037/xhp0000917

Stevens, P., & Plaut, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychonomic Bulletin & Review, 1–30*. https://doi.org/10.3758/s13423-022-02086-0

Sulpizio, S., Pennucci, E., & Job, R. (2021). The impact of emotional content on pseudoword recognition. *Psychological Research Psychologische Forschung, 85*, 2980–2996. https://doi.org/10.1007/s00426-020-01454-6

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior, 15*(6), 607–620. https://doi.org/10.1016/0022-5371(76)90054-2

Taft, M., & Kougious, P. (2004). The processing of morpheme-like units in monomorphemic words. *Brain and Language, 90*(1–3), 9–16. https://doi.org/10.1016/S0093-934X(03)00415-2

Talamo, L., Celata, C., & Bertinetto, P. M. (2016). DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure, 9*(1), 72–102. https://doi.org/10.3366/word.2016.0087

Treiman, R., Wolter, S., & Kessler, B. (2021). How sensitive are adults to the role of morphology in spelling? *Morphology, 31*(3), 261–271. https://doi.org/10.1007/s11525-020-09356-4

Ulicheva, A., Marelli, M., & Rastle, K. (2021). Sensitivity to meaningful regularities acquired through experience. *Morphology, 31*(3), 275–296. https://doi.org/10.1007/s11525-020-09363-5

Vecchi, E. M., Marelli, M., Zamparelli, R., & Baroni, M. (2017). Spicy Adjectives and Nominal Donkeys: Capturing Semantic Deviance Using Compositionality in Distributional Spaces. *Cognitive Science, 41*(1), 102–136. https://doi.org/10.1111/cogs.12330

Westbury, C., & Hollis, G. (2019). Conceptualizing syntactic categories as semantic categories: Unifying part-of-speech identification and semantics using co-occurrence vector averaging. *Behavior Research Methods, 51*(3), 1371–1398. https://doi.org/10.3758/s13428-018-1118-4

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 597. https://doi.org/10.1037/xlm0000064