Journal of Memory and Language 148 (2026) 104746

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Morphemes in the wild: Modelling affix learning from the noisy landscape of natural text

Maria Korochkina [a],*, Marco Marelli [b], Kathleen Rastle [a]

[a] Department of Psychology, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, United Kingdom
[b] Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, Milan, 20126, Italy

ARTICLE INFO

Dataset link: https://osf.io/sf2bh/

Keywords:
Morphology
Learning
Reading
Popular books
Lexical statistics
Computational modelling
Distributional semantics

ABSTRACT

Morphological knowledge serves as a powerful heuristic for vocabulary growth and contributes significantly to the speed and efficiency of reading. While research has long sought to explain how the knowledge of derivational morphology is acquired, previous approaches have struggled to capture the nuanced and complex ways in which derivational morphemes are used in written language, particularly that these morphemes contribute to meaning in a graded manner and that noise introduced by misleading forms (e.g., *deliver*) can impede learning. Our approach builds on earlier insights but moves beyond them by combining a large-scale analysis of vocabulary used in 1,200 popular books with computational modelling to explore how learning of derivational affixes may occur from text containing naturally occurring noise. We use a compositional distributional semantic model to investigate what can be learned about the meanings of individual English prefixes and suffixes through reading and evaluate the model's performance against data from 120 adults in a lexical processing task. Our findings demonstrate that, despite the presence of noise, natural text contains sufficient structure to support the extraction of core affix semantics, and that readers are attuned to the complex patterns that shape affix use in the wild. This work contributes a new dimension to a more principled and psychologically grounded account of morpheme learning, and we discuss both this contribution and the broader insights it offers for language research.

Introduction

Human language is a productive system in which a finite set of elements are combined to express an infinite range of thought. At the word level, this productivity arises from a language's morphology — the system that governs how words are formed. Most approaches to morphology analyse the structure of words in terms of *morphemes*, which are the smallest units in a language that carry independent meaning (e.g., *un-* in *unhappy* means 'not'). Morphemes are typically divided into two types: *inflectional* morphemes, which generate different grammatical forms of the same word (e.g., *create* + *-(e)d* → *created*), and *derivational* morphemes, which enable the formation of entirely new words (e.g., *create* + *-ion* → *creation*). Research shows that most unfamiliar words we encounter are new combinations of morphemes, suggesting that knowledge of morpheme meanings and functions is an important heuristic for vocabulary growth (e.g., Baayen et al., 1993; Brysbaert et al., 2016; Korochkina et al., 2024; Nagy & Anderson, 1984; Rastle & Taylor, 2018). In this article, we focus on derivational morphemes in English, which include prefixes (e.g., *un-* in *unhappy*) and suffixes (e.g., *-ness* in *happiness*).

Numerous studies have established that skilled readers capitalise on their knowledge of derivational morphemes to facilitate comprehension of both familiar and unfamiliar words as they read (see, e.g., Amenta & Crepaldi, 2012; Rastle & Davis, 2008; Stevens & Plaut, 2022, for reviews), and there is consensus that general knowledge of derivational morphology is also crucial for the production of spoken and written words (e.g., Treiman & Cassar, 1996; Zwitserlood et al., 2000). What remains less clear, however, is how readers come to appreciate the derivational regularities of their language. Research demonstrates that, in English, knowledge of inflectional morphology develops earlier than that of derivational morphology. In her seminal work, Berko (1958) demonstrated that children as young as four can apply some inflectional knowledge to create new words, but even at age seven, they still show very limited knowledge of common derivational affixes like *-er* and *-y*. This protracted development of derivational knowledge has prompted researchers to examine more closely the sources and nature of exposure to derivational morphology. This research has shown that derivational regularities in English are both more salient (e.g., Berg & Aronoff, 2017; Ulicheva et al., 2020) and more abundant (e.g., Dawson

* Corresponding author.
E-mail address: maria.korochkina@rhul.ac.uk (M. Korochkina).

https://doi.org/10.1016/j.jml.2026.104746
Received 1 August 2025; Received in revised form 11 December 2025; Accepted 13 January 2026
Available online 17 January 2026

et al., 2023; Korochkina & Rastle, 2025) in written language than in spoken language. These observations have led theoretical accounts to propose that reading experience likely plays a particularly important role in the acquisition of derivational knowledge (e.g., Rastle, 2019), particularly given that formal instruction in morphology in English is often patchy and inconsistent (e.g., Nunes & Bryant, 2006). In this article, we elucidate how exposure to text contributes to morpheme learning and what readers can infer about the meanings of individual affixes in English through recreational reading.

Morpheme learning as a statistical learning problem

Affix morphemes typically do not occur in isolation, so their meanings and functions must be inferred through experience with words that contain them (e.g., *unhappy*, *unclean*, *unwise*, *unafraid*). An influential idea that has guided research over the past two decades is that skilled readers' morpheme knowledge reflects an accumulation of experience with statistical regularities in the language environment (e.g., Gonnerman et al., 2007; Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999; Seidenberg & Gonnerman, 2000; Treiman et al., 2020). This idea stems from the recognition that morphological patterns in language are quasiregular, meaning that they are generally systematic but not entirely consistent. For instance, the same morpheme does not always contribute predictably to the meaning of a word (e.g., Aronoff, 1976). To illustrate, the suffix *-er* clearly denotes agency in words like *baker*, *teacher*, and *banker*, but its contribution in *dresser* is less straightforward. Similarly, the word *grocer* appears to follow the same derivational pattern as *teacher*; however, the element *groc(e)* has no identifiable meaning in contemporary English, making it difficult recognise *grocer* as a morphologically complex word and to determine the exact contribution of *-er* to its meaning. By contrast, *corner* also resembles *teacher* in structure, yet in this case the letter sequence *er* does not contribute meaningfully to the word at all, as there is no semantic relationship between *corn* and *corner*, and their orthographic similarity is purely coincidental.

The quasiregular nature of morphology implies that readers' processing of multimorphemic words should vary depending on the transparency of mappings between meaning, sound, and spelling (e.g., Seidenberg & Gonnerman, 2000), and data simulations using connectionist models provide some support for this view (e.g., Plaut & Gonnerman, 2000). However, these simulations lack grounding in actual linguistic data, as they are based on artificial languages with limited vocabularies and use tasks and materials that only loosely approximate real morphological systems. For example, Plaut and Gonnerman (2000) modelled differences in English and Hebrew morphology by representing stems and affixes as syllables, each of which was assigned a binary pattern. The meanings of these stems and affixes were defined by first assigning each syllable a vector of semantic features (10 randomly selected features for stems and 5 for affixes) and then randomly deactivating a specified proportion of its active features while activating the same proportion of features elsewhere. The stems varied in the frequency with which they combined with affixes, and there was also variation in the extent to which the meanings of the stem-affix combinations preserved the canonical meanings of the constituent stem and affix. The differences between English and Hebrew were simulated by creating 'transparent words', in which all features of the stem and affix meanings were retained (representing Hebrew), and 'opaque words', in which none of the features were retained (representing English). While this approach captured the broad characteristics of morphological structure in the two languages, the instantiation of morphology in the simulation was highly abstract and did not approximate the actual morphology of either language in detail. Consequently, while these simulations have advanced understanding of how input statistics relate to lexical processing, they do not inform how these statistics relate to real text distributions or account for important features of these distributions (e.g., type frequencies of affixes).

Recent research involving computational analyses of lexical databases has begun to articulate these relationships. Ulicheva et al. (2020) extracted 154 English suffixes from the CELEX database (Baayen et al., 1993) and analysed how reliably each suffix was associated with the lexical category of the words in which it appeared. They then examined whether gradations in this reliability were reflected in skilled readers' morpheme knowledge. To illustrate, both *-ness* and *-y* are used to form nouns in English; however, *-ness* is more diagnostic of noun status because *-y* also appears in adjectives (e.g., *cloudy*, *itchy*; see also Berg and Aronoff, 2017, for a similar discussion). Ulicheva et al. (2020) demonstrated that skilled readers' lexical processing mirrored these distributional statistics: the more diagnostic a suffix was of a particular lexical category, the more likely participants were to classify a nonword containing that suffix as belonging to that category (and see, Treiman et al. (2020), for a related finding). In a follow-up study, Ulicheva et al. (2021) found that participants' sensitivity to these spelling-meaning regularities was associated with their reading experience, providing further support for the idea that morpheme knowledge reflects the graded relationship between form and meaning in written language, and that text experience is critical for assimilating this relationship. These conclusions align with findings from studies on the acquisition of morphologically complex novel words, which show that knowledge of how derivational affixes contribute to word meaning facilitates learning, even when new words contain unfamiliar stems (e.g., Dawson et al., 2021a; Merkx et al., 2011; Nathaniel et al., 2024; Tucker et al., 2016).

In parallel to research examining how the distributional properties of individual derivational morphemes in language corpora relate to skilled readers' morpheme knowledge, a complementary line of work has investigated how these same properties influence morpheme learning. In Tamminen et al. (2015), adult readers were presented with definitions of novel words created by combining a familiar stem with a novel suffix (e.g., *mowlomb*, *lynchlomb*). Although the function of the suffix was not explicitly explained, the definitions were designed to allow participants to infer its potential meaning. Tamminen et al. (2015) found that participants' ability to extract and generalise the affix's meaning to new combinations (e.g., *fetchlomb*) depended on the number of different stems it appeared with (i.e., its *type* frequency) and the consistency with which it modified the stem's meaning across the novel words (see Behzadnia et al. (2023), for a similar discussion, and Mirkovic et al. (2019), and Tamminen et al. (2012), for research on memory mechanisms supporting this generalisation process in the learning of inflectional and derivational regularities, respectively). This finding is significant in that it established the prerequisites for derivational learning in the absence of explicit instruction and raised the question of whether these prerequisites are met in real-world reading.

Recent research using large language corpora has made important advances in addressing this question. Korochkina and Rastle (2025) analysed 1,200 books popular among British children and adolescents to examine how this text experience may support derivational learning. They found that most affixes occur in relatively few distinct words (i.e., have low type frequency) and often require specialised linguistic knowledge to identify. For example, an emerging reader can easily recognise the suffix *-er* in *teacher*, as removing it leaves a meaningful word (*teach*) that is clearly related in meaning. In contrast, identifying *-er* in *grocer* requires knowing that *groc(e)* comes from the Latin *grossus* ('gross') and that *grocer* originally referred to someone who sold goods in bulk. Most children (and adults) are unlikely to possess such knowledge and so will need to rely on spelling to parse words into morphemes. Using a computational algorithm to simulate this process, Korochkina and Rastle (2025) showed that derivational affixes were clearly identifiable from spelling alone (i.e., without etymological knowledge) in only half of prefixed words and one third of suffixed words. The algorithm also revealed that about 5% of monomorphemic words were likely to be mistakenly parsed as affixed based on their spelling, with some affixes more prone to these 'false alarms' than

others. These are words like *corner*, where spelling-based segmentation (*corn* + *-er*) is intuitive but misleading, as the apparent suffix does not contribute to the word's meaning. Korochkina and Rastle (2025) proposed that only words that appear morphologically complex in spelling – that is, genuine affixations like *teacher* and false alarms like *corner*, but not unparsable forms like *grocer* – will contribute to readers' morpheme experience, but in different ways. While experience with genuine affixations will benefit derivational learning, exposure to false alarms will hinder it by increasing noise and creating uncertainty about the affix's true function. This account thus links the theoretical notion of morpheme knowledge as a graded construct to specific features of reading experience that might give rise to such gradations, offering a concrete formulation of how derivational learning from print may unfold. By specifying the proportion of complex words likely to contribute to learning affix meanings, as well as the 'noise' that may pose barriers, this account also helps explain why general sensitivity to derivational regularities in online processing tasks does not emerge until mid-to-late adolescence (Beyersmann et al., 2012; Dawson et al., 2018, 2021b).

In a subsequent study, Korochkina et al. (2026) proposed a mathematical formulation of their theoretical model. Unlike previous studies that typically focused on a small set of frequent affixes that generally attach to free stems (e.g., *-er*, *-ness*, *-ly*), they selected 12 derivational affixes from Korochkina et al. (2026) that varied in both type frequency and their participation in false alarms. For each affix, two metrics were computed: (1) orthography-based type frequency, calculated as the number of distinct words in which the affix is recognisable based on spelling alone, and (2) a false alarm penalty, which quantified the extent to which affix experience should be penalised due to the presence of false alarms. The logic behind this metric is that once we have an estimate of how many words a reader is likely to analyse morphologically based on spelling alone (as achieved by the Korochkina and Rastle (2025) algorithm), we can use linguistic knowledge to classify these into genuine affixations or false alarms, and compute a measure expressing the potential harm caused by false alarms. Korochkina et al. (2026) then examined how these two affix properties influenced lexical processing using a morpheme interference task. In this task, participants make word/nonword decisions on morphologically structured nonwords (e.g., *unwood*, *woodness*) and matched orthographic controls without morphological structure (e.g., *ubwood*, *woodnels*). The morpheme interference effect is the finding that readers make more errors and are slower when rejecting morphologically structured nonwords. This effect is interpreted as evidence that skilled readers have acquired sensitivity to morphological regularities of their language (e.g., Crepaldi et al., 2010; De Simone et al., 2024; Taft & Forster, 1975).

Consistent with previous research, skilled readers in Korochkina et al. (2026) exhibited robust morpheme interference effects in both accuracy and response times. However, the key finding from that study is that participants' performance on morphologically structured nonwords varied as a function of the two affix metrics Korochkina et al. (2026) defined: the higher the orthography-based type frequency and the lower the false alarm penalty of the affix, the more errors participants made and the slower their responses were. This result indicates that participants found it easier to reject nonwords containing affixes that appeared in fewer distinct words and were associated with more false alarms, possibly because their implicit knowledge of these affixes was weaker compared to more frequent affixes that were less prone to false alarms. Critically, the model that included orthography-based type frequency and the false alarm penalty explained readers' processing of nonwords with different affixes significantly better than models that either assumed that all historically complex words contributed equally to learning or that disregarded the false alarms.

The findings of Korochkina et al. (2026) align with those of Ulicheva et al. (2020), providing further evidence that derivational knowledge gained through text experience appears to depend on what readers can readily identify from spelling, even if this entails relying on cues that do not reliably reflect true morphological structure (i.e., false alarms) or overlooking potentially informative exemplars (e.g., words with bound stems). This body of work advances the field by proposing concrete metrics for summarising morpheme presentation and distribution in text. However, this approach relies on linguistic expertise to determine whether a word containing a given affix is 'useful' (e.g., *teacher*) or 'harmful' (e.g., *corner*) for learning. In real-life reading, learners do not have access to such information, meaning that even these metrics fall short of capturing how the average reader navigates the noisy input from which morpheme knowledge emerges. Moreover, this approach does not account for the fact that morphemes contribute to word meaning in a graded manner. For example, both *teacher* and *dresser* would be placed in the 'useful' category; however, the contribution of *-er* differs between them and is arguably less transparent in *dresser*. Similarly, *artist*, *typist*, and *racist* would all be classified as useful for learning, yet the suffix *-ist* functions differently in each case (e.g., a *typist* is someone who types, but a *racist* is not someone who races). Thus, a more ecologically valid test of Korochkina and Rastle (2025)'s account should not only simulate how readers without specialised knowledge learn to distinguish between useful and harmful cases, but also account for differences in semantic consistency across affixes. In this study, we address both goals using a compositional distributional semantic model.

Morpheme learning through the lens of distributional semantics

Compositional distributional semantic models build on distributional semantic models, which represent word meanings as vector coordinates in a high-dimensional vector space (e.g., Amenta et al., 2020; Turney & Pantel, 2010). These models are grounded in the distributional hypothesis, which suggests that a word's meaning can be inferred from the contexts in which it appears (Harris, 1954). This hypothesis is based on the observation that words used in similar contexts tend to have similar meanings, while words used in distinct contexts have more divergent meanings. For example, *boat* and *ship* denote semantically similar concepts, and both commonly co-occur with words like *water*, *sea*, and *passenger*. Distributional semantic models extract such co-occurrence patterns from large language datasets and compute word vectors based on these patterns using machine learning techniques such as matrix factorisation (e.g., Latent Semantic Analysis; Landauer & Dumais, 1997), word embedding algorithms (e.g., word2vec; Mikolov et al., 2013), or transformer architecture (e.g., BERT; Devlin et al., 2019). These models align well with patterns observed in human language processing and are therefore considered to offer a psychologically plausible account of semantic memory (e.g., Günther et al., 2019; Jones et al., 2006, 2015; Landauer & Dumais, 1997; Mitchell et al., 2008). *Compositional* distributional semantic models extend this distributional approach to semantics by incorporating *compositionality* — the process by which meaningful linguistic units (e.g., words or morphemes) combine to form new meanings (e.g., *snow* + *man* → *snowman*; *weight* + *-less* → *weightless*) (e.g., Baroni et al., 2014; Marelli & Baroni, 2015; Mitchell & Lapata, 2010).

One of the most prominent compositional distributional semantic models is the *Compounding as Abstract Operation in Semantic Space* (CAOSS) model (Marelli et al., 2017), which was developed to capture the semantic processing of compound words. Compounds are formed by combining two constituent words (e.g., *snow* + *man* → *snowman*), whose meanings may interact in various ways to create a new meaning. For example, while a *snowman* is a human-like figure made of snow, a *mailman* is not a man made of mail. Likewise, an *airbed* is a type of bed filled with air, but an *aircraft* is not a vehicle made of air but rather one capable of flying and thus traversing through air. The CAOSS model addresses this complexity by assuming that individual constituents undergo a transformation when combined to form a new lexical item, and it estimates this transformation by representing it as a numerical matrix. Once these matrices (one for each constituent slot) are learned, they can be used to generate vector representations for novel compounds (e.g., *aircup*). Research indicates that these model-generated

representations align well with speakers' intuitions about the meanings of novel compounds, suggesting that the model provides a plausible approximation of compositionality as a cognitive process (e.g., Günther & Marelli, 2020). Importantly, although the CAOSS model was originally developed to capture the meanings of compound words, it is not limited to this type of morphological composition and has also been successfully applied to other forms of derivational morphology, including suffixation (Bonandrini et al., 2023, and see Lazaridou et al., 2013, for a proof of concept of the combinatorial approach to affixation).

The present study

In the present study, we trained the CAOSS model on data from a large corpus of books popular among children and young people in the United Kingdom (Korochkina et al., 2024), with the goal of exploring what the model can learn about the meanings of individual derivational affixes used in these texts. While we build on previous work in our implementation of the model, we introduce several important innovations. The most critical of these concerns the data used to train the models in our study. Previous studies considered only genuine instances of affixation (e.g., *teacher*) for model training, meaning that the input the models received simulated a pristine learning environment devoid of noise. By contrast, we constructed our training sets to reflect the real-life affix experience of a typical reader, rather than relying on linguistically informed judgments about what aspects of morpheme experience should be relevant for learning. Thus, our training data included all words that give the impression of compositionality based on spelling, regardless of their etymological status. This meant incorporating both genuine affixations and false alarms such as *corner*, making our study the first to apply the CAOSS model to input in which signal co-occurs with noise. Our second innovation was to apply the CAOSS model to both suffixes and prefixes. As previous work focused exclusively on suffixation, this study marks the model's first implementation in the domain of prefix semantics.

Our overarching aim was to uncover (1) what the CAOSS models can learn about the meanings of individual prefixes and suffixes, and (2) to what extent the model's knowledge of affix meanings can explain skilled readers' lexical processing. To achieve this, we constructed a set of metrics capturing different aspects of the models' knowledge of the individual affixes from Korochkina et al. (2026) and assessed how well these metrics accounted for the behavioural patterns observed in the morpheme interference task reported in that study. We also compared these metrics with the false alarm penalty metric from Korochkina et al. (2026) to evaluate whether CAOSS models can capture aspects of meaning on par with their binary, linguistically informed classification, and potentially even go beyond it. Finally, in a complementary analysis, we explored whether these behavioural patterns in nonword processing can be explained more effectively by considering not only the properties of the affixes themselves but also the combination of stem and affix. We begin by describing the model architecture and training procedure in detail, before outlining how we computed the CAOSS-based metrics and evaluated them against the behavioural data.

Modelling affix and nonword representations

Model architecture

Within the CAOSS model, derivational affixation is understood as a form of compounding, applied to a stem and an affix morpheme (e.g., *weight* + *-less* → *weightless*) rather than to stems, as seen in compounds (e.g., *snow* + *man* → *snowman*). Mathematically, this is expressed as follows:

$$v_c = M_a \times v_a + M_b \times v_b \quad (1)$$

Here, v_a, v_b, and v_c are vectors in a multi-dimensional semantic space that encode the meanings of a stem (e.g., *weight*), an affix (e.g., *-less*), and an affixed word combining the two (e.g., *weightless*), respectively. The vector representations of stems and affixed words are referred to as *lexicalised* representations because they are typically derived from a semantic space built on natural language usage (i.e., from corpora). M_a and M_b are transformation matrices applied to these vectors so that their sum – referred to as the *compositional* representation of the affixed word – approximates the lexicalised representation v_c as closely as possible. At the outset, the CAOSS model must be provided with the vector representations of the stems, affixes, and affixed words (v_a, v_b, v_c). Its task is to estimate the transformation matrices M_a and M_b in a *training phase*.

This training objective implies that the CAOSS model is not designed to store explicit semantic representations of individual affixes, but rather to learn transformation operations that enable stems and affixes to be composed into complex words. Because the model aims to learn affixation as a compositional process, the transformation matrices are not item-specific but consist of a single matrix for the affix slot and a single matrix for the stem slot. Once training is complete, the learned transformation matrices can be applied to the original affix vectors to derive *transformed* affix vector representations. While these derived representations were not learned with the explicit goal of capturing affix meanings, they nevertheless reflect how the model has internalised the semantic contribution of each affix within compositional word formation. These transformed representations can then be used both to (a) infer the model's implicit representation of affix semantics and (b) generate compositional vector representations for novel stem-affix combinations (e.g., *floorless*, *cakeless*, *penless*). In this article, we utilise both of these functionalities of the CAOSS model, and we outline their implementation in the next section. In the remainder of this section, we describe the model-building process, which involved assembling the training sets and constructing the vector representations of stems, affixes, and affixed words that the model required to learn the transformation matrices. The code for the entire modelling process is available in this project's repository on the Open Science Framework (https://osf.io/sf2bh/).

Training sets

We built two CAOSS models: one for generating vector representations of the most common *prefixes* in popular books and the other for the most common *suffixes*. We implemented separate models for prefixed and suffixed words because prefixation and suffixation represent different compositional processes. The functions of suffixes are typically distinct from those of prefixes; for example, suffixes often change the grammatical class of a word, whereas prefixes typically modify the meaning of the stem in more subtle ways (e.g., Marchand, 1960). As explained in the previous section, the CAOSS model aims to learn affixation as a word-formation process; therefore, during training, it acquires the transformations required to combine stems and affixes into complex words. Reflecting this objective, the model learns a *single* matrix for the affix slot and a *single* matrix for the stem slot rather than separate matrices for each individual stem or affix that could occupy those slots (i.e., the matrices are role-general rather than item-specific). Given this approach, combining prefixed and suffixed words into a single model could have forced the affix matrix to conflate fundamentally different contributions, potentially misrepresenting the functions of both types of affixes. We created and trained these models using the DIStributional SEmantics Composition Toolkit (DISSECT; Dinu et al., 2013), a publicly available Python 2 module hosted at https://github.com/composes-toolkit/dissect.

The list of common prefixes and suffixes included 23 prefixes and 25 suffixes identified by Korochkina and Rastle (2025) as the most frequently used in the CYP-LEX corpus (Korochkina et al., 2024). This corpus consists of books that UK residents aged 7+ *choose* to read. The selection of books was based on popularity, and, consequently, many titles in the 13+ age band are also frequently read by adults (e.g., *The*

Hunger Games, *The Maze Runner*, *A Game of Thrones*, *Atonement*, *1984*). We therefore considered that the texts in this corpus constitute a valid approximation of morphemic exposure experienced by typical young adults with average reading habits.

To create the training sets, we first identified all distinct words in the 1,200 books of the CYP-LEX corpus that appear to contain the 48 common affixes using the RegEx algorithm developed by Korochkina and Rastle (2025). This algorithm identifies words that appear to contain derivational affixes based on their orthographic form by scanning for letter strings that pair existing stems with the target affixes. It activates only when the orthographic patterns associated with each affix (e.g., *ness*, *er*, *de*) are positioned appropriately (prefixes at the beginning and suffixes at the end of the word). When a word contains the pattern in the correct position, the 'affix' is detached, and the algorithm checks whether the remaining letter string (the 'stem') exists in CYP-LEX. If it does, the word is tagged as affixed; if not, the algorithm moves on. Korochkina and Rastle (2025) incorporated rules into the algorithm to address common orthographic alterations arising from derivational affixation, such as silent *-e* deletion (e.g., *adore* + *-able* → *adorable*, *create* + *-or* → *creator*) and consonant doubling (e.g., *sun* + *-y* → *sunny*, *run* + *-er* → *runner*, *admit* + *-ance* → *admittance*). These rules align with those taught as part of the English National Curriculum. The full list of rules, the algorithm code, and additional details are provided in Korochkina and Rastle (2025).

The supervision required by the RegEx algorithm is minimal: it only needs to locate letter strings corresponding to affixes based on key orthographic rules. Beyond this, no guidance regarding the etymological status of words is provided. By identifying affixed words solely from orthographic information (the presence of letter sequences), the algorithm captures both genuine cases of affixation and false alarms — words that contain the affix pattern orthographically but are not actually derived through affixation. For example, *er* appears at the end of *teacher*, *grocer*, and *corner*. *Teacher* and *grocer* are genuinely affixed words; however, while *teacher* can be segmented into *teach* + *er*, *grocer* cannot (there is no word *groc(e)* in modern English) and hence is not identified as a relevant exemplar. In contrast, removing *er* from *corner* leaves *corn*, which is a word attested in CYP-LEX, so the algorithm tags *corner* as containing the suffix *-er*. Likewise, because the algorithm accounts for the silent *-e* rule, it identifies both *writer* (*write* + *er*) and *badger* (*badge* + *er*) as relevant exemplars, even though *badger* has no derivational relationship to *badge*.

The RegEx algorithm was applied to all distinct words in the CYP-LEX database, resulting in identification of 7,642 words containing the 23 most common prefixes and 13,472 words containing the 25 most common suffixes. These words were used to construct the training sets: the CAOSS prefix model training set included 7,642 triplets (stem, prefix, prefixed word), while the CAOSS suffix model training set included 13,472 triplets (stem, suffix, suffixed word). A detailed breakdown of triplet counts per affix is provided in the Supplementary Material, available on OSF (https://osf.io/sf2bh/). The training sets themselves (pref_train_set.txt and suf_train_set.txt) are also available in the OSF repository.

Vector representations for model training

As explained in "Model architecture", the CAOSS model requires vector representations for the stem (v_a) and the affixed word (v_c) as inputs. In our case, this meant obtaining vector representations for all prefixed and suffixed words, as well as their stems, from the prefix and suffix training sets we prepared. Our initial plan was to derive these vector representations by training a distributional semantic model on the CYP-LEX corpus (Korochkina et al., 2024), as word representations based on popular books are likely to better approximate morphemic experience gained through reading than those derived from other sources. Distributional semantic models learn word meanings from the contexts in which words occur, and therefore training them requires that each word appears frequently in the corpus. For example, training a *word2vec* model (Mikolov et al., 2013) – one of the most commonly used and best-performing distributional semantic models – requires each word to appear at least 10 times (e.g., Baroni et al., 2014). However, 51% of the distinct words in the CYP-LEX corpus occur less than 10 times across 1,200 books (see Korochkina et al., 2024, for a detailed discussion of this finding). This feature would have prevented the *word2vec* model from generating adequate representations for half the words in the corpus. Consequently, we opted to use the *subs2vec* tool (van Paridon & Thompson, 2021) to obtain representations for the stems and affixed words instead, as its vectors were trained on very large corpora and are therefore expected to be of higher quality than those we could generate from the CYP-LEX corpus.

The *subs2vec* tool provides 300-dimensional vector representations for a large number of words across 55 languages, generated using the *fastText* implementation of the *skipgram* algorithm (Bojanowski et al., 2017). For each language, *subs2vec* offers three types of representations: those trained on the OpenSubtitles corpus (which includes all subtitles available in the OpenSubtitles archive for that language), those trained on the Wikipedia corpus (containing all Wikipedia entries in that language), and those trained on a combination of both. For this study, we selected the English semantic space trained on the combined OpenSubtitles and Wikipedia datasets, as it has been shown to outperform models trained on either corpus alone (van Paridon & Thompson, 2021). This semantic space includes vector representations for 870,220 unique entries and is freely available at https://github.com/jvparidon/subs2vec/. It contains vector representations for all prefixed and suffixed words, as well as their stems, included in our training sets; therefore, we extracted *all* word representations from this space.

In addition to the stem and affixed word representations, the CAOSS model also requires affix vector representations (v_b) to generate the transformation matrices M_a and M_b. These initial vector representations for the 48 affixes used in our training sets were created using the method proposed by Westbury and Hollis (2019), which derives an affix's vector representation by averaging the vector representations of all words containing that affix. Westbury and Hollis (2019) demonstrated that this approach effectively captures category-defining information, such as part of speech and morphological family, and aligns with human word category judgments even when morphological false alarms like *corner* are included in the averaging process. Therefore, in the present study, we computed each affix's initial vector representation by averaging the vector representations of *all* distinct words identified by the RegEx algorithm as containing that affix, including both genuinely affixed words like *teacher* and morphological false alarms like *corner*.

Construction of CAOSS-based metrics

As described in "Model architecture", the transformation matrices acquired by the CAOSS model during training indicate how the model has internalised the semantic contributions of stems and affixes during compositional word formation. During training, we provided the model with affix vector representations, computed using the Westbury and Hollis (2019) method, to serve as proxies for affix meaning. By multiplying each affix vector by the appropriate affix transformation matrix (one for prefixes and one for suffixes) after training, we obtain a *transformed* representation of the affix's meaning. These transformed vectors encode the model's acquired knowledge of each affix's semantic function *within* complex words; however, their high-dimensional format (300-element numerical vectors) necessitates an additional processing step to translate the encoded information into variables suitable for analytical purposes (e.g., statistical modelling). This additional step applies mathematical procedures to the transformed representations to yield *metrics* that represent different facets of affix meaning.

As stated in the Introduction, we evaluated the performance of the CAOSS models using morpheme interference data from Korochkina et al. (2026). Accordingly, all metrics were computed for the 12

affixes and the morphologically structured nonwords containing these affixes that were used in that study. Details of the materials (including which affixes were selected and how the nonwords were constructed), participants, and data analysis are provided in "Analytical approach to behavioural data". We selected metrics that have been shown to capture meaningful internal structure in distributional semantic models, with prior research demonstrating that they can explain variance in human lexical processing (see Sections "Nonword-based metrics" and "Affix-based metrics" below). This involved deriving two classes of metrics from the CAOSS models for each affix and morphologically structured nonword: *affix-based metrics*, which quantify the model's acquired knowledge of affix semantics, and *nonword-based metrics*, which reflect the model's intuitions about the meaningfulness of specific stem-affix combinations. We selected three metrics for each class to obtain a richer and more comprehensive picture of what the model has learned. The metrics within a class all draw on information encoded in the transformation matrices and are therefore correlated; however, they are not redundant, and each captures a distinct aspect of the model's knowledge (the highest absolute Spearman correlation among the affix-based metrics is .51, and .11 among the nonword-based metrics). We emphasise that these metrics are neither the mechanism nor the model of morpheme learning. Rather, they function as analytical tools for probing the CAOSS model and, in this sense, serve as proxies for what the model has learned.

The first step in computing all metrics was to derive the transformed vector representations for each of the 12 affixes from Korochkina et al. (2026). The *affix-based metrics* were computed directly from these transformed affix representations. To compute the *nonword-based metrics*, we followed several steps. First, we generated compositional representations for all morphologically structured nonwords containing the 12 affixes from Korochkina et al. (2026). This step used Eq. (1) from model training; however, since the transformation matrices (M_a and M_b) had already been learned and the transformed affix representations were available, we now solved this equation for v_c, which in this context represented the nonwords. These nonword representations are *compositional* because they are constructed by the CAOSS model using the learned transformations (unlike real words, nonwords cannot have pre-existing lexicalised representations). These nonword representations formed the basis for computing the nonword-based metrics.

Once the metrics were computed, we examined whether they could account for differences in how skilled readers from Korochkina et al. (2026) processed morphologically structured nonwords, and whether the performance of these metrics was comparable to that of the false alarm penalty metric reported in the same study. In what follows, we describe the construction of all metrics. We begin with the nonword-based metrics, as their computation closely follows prior work, and then explain how we extended the same foundational logic to derive the affix-based metrics. The code used to compute the metrics is available in this project's OSF repository (https://osf.io/sf2bh/).

Nonword-based metrics

The nonword-based metrics were used to capture the model's knowledge of how meaningful novel stem-affix combinations are — that is, whether combining a given affix and stem produces a compositional representation that makes sense given the concepts already present in the world. These metrics were defined following Bonandrini et al. (2023). The *nonword diffuseness* metric (also known as *nonword entropy*), reflects the degree to which the nonword's meaning is well-defined or uncertain. Vector representations of linguistic units (e.g., words or morphemes) encode meaning across multiple dimensions, with each dimension thought to represent a different aspect of the unit's meaning. The diffuseness metric measures how the weights of these dimensions are distributed: when meaning is uncertain, the weights are spread across many dimensions, resulting in a uniform distribution and a high diffuseness score. Conversely, when the meaning is more specific and better defined, the weights are concentrated in a few dimensions, resulting in a low diffuseness score (Marelli & Baroni, 2015). We computed this metric using an adaptation of the Shannon entropy formula proposed by Bonandrini et al. (2023), which is necessary because the standard entropy formula cannot be applied to negative values that frequently occur in words' semantic vectors:

$$Diffuseness(t(v)) = -\sum_{i=1}^{N} d_i \log(d_i) \quad (2)$$

Here, $t(v)$ represents the nonword vector after a two-stage transformation. First, each element is adjusted by adding the absolute value of the minimum element in the vector and a constant representing the inverse of the number of vector dimensions. Second, each element is divided by the sum of all vector elements. This transformation ensures that the elements of the nonword vector sum to 1, making it analogous to a probability distribution. In the formula above, d_i represents the ith element in the transformed vector $t(v)$, and N denotes the number of dimensions in the vector (300).

The *nonword richness* metric reflects the semantic richness of the nonword's meaning. It is calculated by taking the Euclidean norm of the nonword vector, which is the square root of the sum of the squared values of the vector's elements:

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2} \quad (3)$$

Both diffuseness and richness are determined by the distribution of the vector dimensions, but they differ in the aspects of meaning they capture. Diffuseness reflects the degree of uncertainty in the nonword's meaning by measuring how the weights are spread across the dimensions. In contrast, richness depends on the specific values of these weights and thus captures the overall amount of information in the vector; it is therefore also known as *nonword magnitude* (e.g., Bonandrini et al., 2023). A higher nonword richness score indicates a broader range of meanings and greater semantic variation in its use, while a lower score suggests that the nonword's meaning is less rich and complex and that it provides limited semantic information. That said, the interpretation of this metric is not always straightforward in the context of morpheme combinations, and we refer readers to Bonandrini et al. (2023) for a more comprehensive discussion.

To compute the *nonword neighbourhood density* metric, we first identified the ten closest semantic neighbours of a nonword — that is, the ten words most similar in meaning to the nonword (Marelli & Baroni, 2015). Then, we calculated the cosine similarity between the nonword vector and each of its ten closest neighbours and averaged across the resulting values. The underlying assumption is that a vector encoding a meaningful new concept should occupy a region of semantic space densely populated by vectors of related concepts that have already been lexicalised (Marelli & Baroni, 2015). In contrast, a vector representing something nonsensical should lie far from any vectors corresponding to meaningful, lexicalised concepts. In other words, if a nonword is meaningful (e.g., *acneless*), it should be easy to think of related words (e.g., *pimple*, *teenager*, *lotion*), resulting in high neighbourhood density, whereas if a nonword is less meaningful (e.g., *hikeless*), few related concepts would come to mind, yielding low neighbourhood density. This metric does not reflect a nonword's lexical utility (i.e., how 'good' a nonword is): it neither assumes nor indicates whether a nonword is needed or would be useful in the language. Rather, it simply quantifies the alignment of the concept with the existing semantic space.

Consistent with morpheme interference theory, our premise for the nonword-based metrics was as follows: the more well-defined, rich, and well-aligned a nonword is with other words in the semantic space (i.e., lower diffuseness, higher richness, and higher neighbourhood density, respectively), the more word-like it should feel. Consequently, it should be harder to reject, leading to lower accuracy and longer response times.

Affix-based metrics

The affix-based metrics were intended to capture three distinct aspects of affix meaning and function. *Affix diffuseness* and *affix richness* were computed in the same way as nonword diffuseness and richness (see formulae in (2) and (3) above), but using the transformed affix vectors instead of the nonword vectors. These metrics reflect, respectively, the degree of diffusion and uncertainty in the affix's meaning, and the richness and complexity of that meaning. Finally, specifically for this study, we designed the *affix-word coherence* metric. This metric is derived by computing the cosine of the angle between the transformed vector representation of the affix and the vector representation of each word containing this affix (i.e., affixed words used in the training phase), and then averaging across the resulting cosine values. The metric thus reflects the *average* coherence of an affix's meaning relative to the meanings of the words in which it appears: higher values indicate greater alignment with the derived words containing that affix.

Following the same logic as for the nonword-based metrics, we expected that readers would have better knowledge of affixes whose meanings were richer, less diffuse, and more closely aligned with the meanings of the words that contain them. Consequently, nonwords containing these 'stronger' affixes should be harder to reject, resulting in lower accuracy and slower response times.

Analytical approach to behavioural data

As stated above, all metrics were computed for the affixes and morphologically structured nonwords used in Korochkina et al. (2026), and their performance was evaluated using the behavioural data from that study. In this section, we first provide details on the core aspects of the Korochkina et al. (2026) study, and then outline the analytical approach applied to the behavioural data.

The data

In Korochkina et al. (2026), 120 native English speakers based in the United Kingdom (mean age = 30 years, *SD* = 7 years, 63 female, 56 male, 1 non-binary) completed an online lexical decision task. Each participant saw 240 nonwords and 240 real words, included to balance the number of expected 'yes' and 'no' responses. Half of the nonwords (N = 120) were morphologically structured (e.g., *woodness*, *sheeper*, *motorate*), with each of 12 affixes paired with 10 different stems. The 12 affixes included six prefixes (*un-*, *mis-*, *dis-*, *pre-*, *re-*, *de-*) and six suffixes (*-ness*, *-ly*, *-able*, *-er*, *-ic*, *-ate*), and were selected because they had reasonably high type frequency in the CYP-LEX corpus (ensuring familiarity to participants) while offering variability in detectability and involvement in false alarms (Korochkina & Rastle, 2025). The remaining nonwords served as orthographic controls: each was created by modifying a single letter in the affix of a morphologically structured nonword (e.g., *sheepel*, *motorafe*, *woodnels*), ensuring that the only difference between these nonwords and the morphologically structured ones was the absence of morphological structure. In the present study, only the data for morphologically structured nonwords are used.

Korochkina et al. (2026) demonstrated that participants were less accurate and slower at rejecting morphologically structured nonwords compared to their orthographic controls (i.e., a morpheme interference effect in both accuracy and response times). They then investigated whether participants' accuracy and speed in judging the lexical status of morphologically structured nonwords could be explained by the properties of the affixes used. For each nonword, they computed two metrics: orthography-based type frequency, defined as the number of distinct words in the CYP-LEX corpus in which the affix used in this nonword was identified from spelling alone, and the false alarm penalty, which captured the learning penalty associated with the affix's participation in false alarms in the CYP-LEX corpus. These metrics were then included as predictors in statistical models, with response accuracy and response times as the dependent variables. Korochkina et al. (2026) showed that lexical decision accuracy and reaction time varied as a function of orthography-based type frequency and the false alarm penalty of the affixes contained in the nonwords. These metrics accounted for differences in participants' behaviour more effectively than alternative variables (see the original paper for further details; the data and analysis code are available at https://osf.io/yq9h7/overview).

Analysis approach

As stated in the Introduction, our goal in this study was twofold. First, we aimed to determine whether the CAOSS models can learn key aspects of affix meaning and behaviour in a context that closely resembles how readers are likely to experience affixes in real-life reading, where input is often noisy and there is no explicit instruction on which aspects of this experience should be prioritised (i.e., genuine affixations) and which should be disregarded (i.e., false alarms). Second, we sought to use these models to better understand the extent to which skilled readers' lexical processing of nonwords is influenced by the meaningfulness of specific stem-affix combinations, beyond what can be explained by affix properties alone.

We addressed these goals in two analyses: *Analysis 1* examined how well the affix-based metrics accounted for variability in skilled readers' morpheme knowledge. We adopted an analytical approach similar to Korochkina et al. (2026), but replaced the false alarm penalty used in their study with three affix-based metrics derived from the transformed affix vector representations generated by the CAOSS models. A separate model was constructed for each of these metrics, and, like in Korochkina et al. (2026), each model included the orthography-based type frequency variable as a covariate. We then compared the fit of these three models to both the empirical data and to the false alarm penalty model from Korochkina et al. (2026). This comparison enabled us to examine whether the CAOSS model, which is not explicitly told which words to treat as useful or as false alarms, could account for patterns in readers' behaviour as effectively as the false alarm metric from Korochkina et al. (2026), where making this distinction required expert linguistic judgement.

The aim of *Analysis 2* was to evaluate whether the metrics capturing stem-affix compositionality (i.e., nonword-based metrics) can explain variability in skilled readers' lexical processing beyond what is accounted for by the affix-based metrics. To do this, we fitted two models – one for response accuracy and one for response times – for each affix-based metric, including the three nonword-based metrics as covariates, and evaluated whether these extended models provided a better fit to the data. This yielded a total of six models: three for accuracy and three for response times.

The data were analysed in R, version 4.2.1 (R. Core Team, 2022). For the response time data, we followed the preprocessing steps outlined in Korochkina et al. (2026), which included removing data points that clearly fell outside the data distribution and applying an inverse transformation as indicated by the Box–Cox test (Box & Cox, 1964). We used (Generalised) Linear Mixed-Effects Models in all analyses, and only response times for correct 'no' decisions were considered. The random effects structure of all models was identical to that in Korochkina et al. (2026), with varying intercepts for participants and items, determined using a method proposed by Bates et al. (2018). In all models, likelihood ratio tests were used to assess whether the null hypothesis of no effect could be rejected for a given predictor variable by comparing a model containing this predictor with one from which it had been removed. Because the aim of Analysis 1 was to compare the performance of the CAOSS-derived metrics with that of the false alarm penalty metric in explaining participants' behaviour, model selection was based on the Akaike Information Criterion (AIC; Akaike, 1973; 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978), with smaller values indicating a better fit to the data. In Analysis 2, we compared the models from Analysis 1 to those

Table 1
Results of Analysis 1 for models incorporating affix-based metrics derived from the CAOSS models and the false alarm penalty model from Korochkina et al. (2026). Models with affix-based metrics are ranked from best to worst based on AIC and BIC values (lower values indicate better model fit). The letter *β* denotes the *standardised* coefficient. The statistics for the false alarm penalty model are shown against a grey background to distinguish it from the models that are the focus of the present study.

Dependent variable	Metric	*β*	*SE*	*z*/*t*	*p*	AIC	BIC
Response accuracy	Affix richness	−0.51	0.04	−12.09	<.001	11,783	11,821
	Affix-word coherence	−0.44	0.04	−11.41	<.001	11,799	11,837
	Affix diffuseness	0.34	0.05	6.87	<.001	11,875	11,913
	False alarm penalty from Korochkina et al. (2026)	0.49	0.04	11.47	<.001	11,796	11,834
Response times	Affix richness	0.14	0.01	12.64	<.001	61,058	61,103
	Affix-word coherence	0.12	0.01	11.70	<.001	61,076	61,121
	Affix diffuseness	−0.09	0.01	−6.48	<.001	61,162	61,207
	False alarm penalty from Korochkina et al. (2026)	−0.13	0.01	−11.73	<.001	61,078	61,122

that included both the affix-based and nonword-based metrics, using likelihood ratio tests to determine whether adding each nonword-based metric significantly improved model fit. To facilitate comparison across variables, we report *standardised* coefficients for all models.

Results

Analysis 1: Role of affix properties

The output of all models, along with their AIC and BIC values, is presented in Table 1. The table also provides the same statistics for the false alarm penalty model from Korochkina et al. (2026). The effect orthography-based type frequency was significant across all models; however, we omit these statistics for brevity (full model output is available in the analysis code files in this project's repository on OSF) and prioritise the CAOSS-derived metrics, which were the primary focus of this study.

Table 1 shows that all models incorporating affix-based metrics yielded significant effects for both accuracy and response time data.[1] These effects are visualised in Figs. 1 and 2, which illustrate that participants made more errors (Fig. 1) and were slower (Fig. 2) when rejecting nonwords containing affixes whose meanings were richer, more coherent with the meanings of the words they appear in, and less diffuse. Table 1 ranks the models from best to worst based on their AIC and BIC values, and, for both response accuracy and response times, the model including affix richness outperformed all others. The second-best model was the one incorporating affix-word coherence, with a fit nearly identical to that of the false alarm penalty model from Korochkina et al. (2026). The model using affix diffuseness showed the poorest fit, with the highest AIC and BIC values across both response accuracy and response times data.

These findings suggest that the CAOSS models we developed successfully captured meaningful aspects of affix usage that influence how skilled readers process morphologically structured nonwords. This is particularly striking given that the models were not explicitly informed about which affixed words in the training data represented valid and informative uses of affixation, and which were misleading (i.e., false alarms). Moreover, the strong performance of the affix richness model – which substantially outperformed the false alarm penalty model – suggests that this metric captures deeper aspects of affix meaning that go beyond a simple binary classification based on orthographic patterns. We now turn to the results of our second analysis, which examined whether incorporating the nonword-based metrics led to a significant improvement in the fit of these three affix-based models.

Analysis 2: Role of stem-affix combination

Table 2 presents the results of the response accuracy and response time models, each of which included one of the three affix-based metrics along with the three nonword-based metrics. The inclusion of nonword diffuseness did not improve model fit for any of the affix-based models, for either dependent variable. For response accuracy, adding the nonword richness metric improved model fit in both the affix richness and affix diffuseness models. For response times, nonword richness improved model fit only in the affix diffuseness model, whereas nonword neighbourhood density improved the fit across all models. These results indicate that, overall, nonwords were hardest to reject when they were semantically rich, closely related in meaning to their semantic neighbours, and contained affixes whose meanings were richer, more coherent with meanings of the derived words they appear in, and less diffuse. Importantly, in all models except for the affix diffuseness models, the standardised coefficients for the affix-based metrics were larger than those for the nonword-based metrics. This suggests that the meaningfulness of the affixes themselves had a stronger impact on participants' processing than the overall meaningfulness of the nonwords in which those affixes appeared.

Discussion

Research suggests that reading experience plays an important role in the acquisition of derivational knowledge (Rastle, 2019). The present article has examined what can be learned about the meanings of individual derivational affixes in English through recreational reading. We considered for the first time the complex and quasiregular nature of morphological patterns as they occur in natural text, rather than relying on artificial language learning paradigms, simplified simulations of morphological systems, or curated, linguistically annotated training data. We trained a compositional distributional semantic model (CAOSS; Marelli et al., 2017) on vocabulary from 1,200 books that approximate the reading experience of a typical young adult. We then investigated what hypotheses the model can generate about the meanings of prefixes and suffixes based on that vocabulary experience. To translate the model's acquired knowledge about affixes into variables suitable for statistical analysis, we computed a series of metrics and then assessed how well these metrics accounted for readers' behaviour. We found that the model's knowledge of individual affixes (as assessed through the computed metrics) explained patterns seen in skilled readers' processing of novel stem-affix combinations better than metrics that rely on linguistic classifications of morphemic experience into 'useful'

[1] During the review process, a question arose regarding whether the observed effects were driven by the suffix *-ness*. We reran all models (including those in Analysis 2) excluding nonwords containing this suffix, and all key results and conclusions remained unchanged. All code and results from these additional analyses, including data visualisations, are available on OSF in the file 'caoss_analysis_of_empirical_data_remove_ness.html' for readers who wish to inspect them or perform further subset analyses.

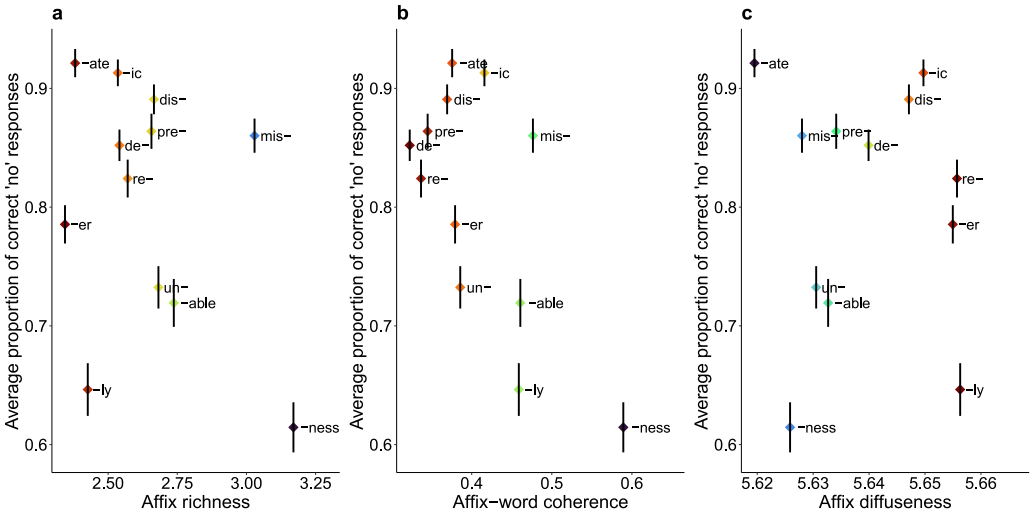

Fig. 1. Skilled readers' accuracy in the morpheme interference task as a function of three affix-based metrics derived from the CAOSS models: affix richness (a), affix-word coherence (b), and affix diffuseness (c). Coloured diamonds represent the average proportion of correct 'no' responses for each of the 12 affixes. The colour gradient indicates affix performance: redder tones represent 'worse' affixes, while bluer tones represent 'better' affixes according to the respective metrics. Vertical black bars denote one standard error from the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

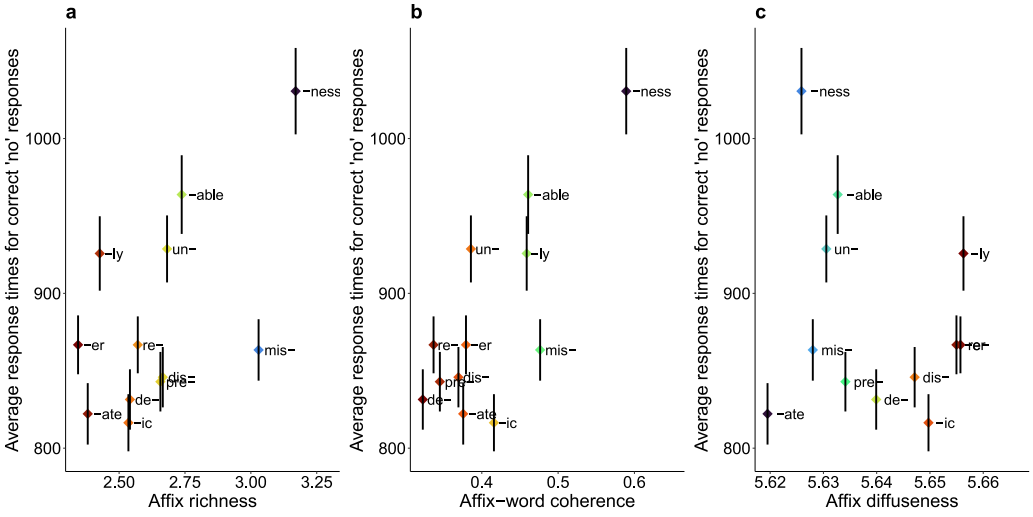

Fig. 2. Skilled readers' response times (for correct responses) in the morpheme interference task as a function of three affix-based metrics derived from the CAOSS models: affix richness (a), affix-word coherence (b), and affix diffuseness (c). Coloured diamonds represent the average response times for correct 'no' responses for each of the 12 affixes. The colour gradient indicates affix performance: redder tones represent 'worse' affixes, while bluer tones represent 'better' affixes according to the respective metrics. Vertical black bars denote one standard error from the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

or 'misleading' cases. In addition, we discovered that the properties of the affixes had a stronger influence on participants' behaviour than the overall meaningfulness of the stem-affix combinations themselves.

These findings break new ground in several ways. Previous approaches to derivational learning have either relied on miniature simulated datasets (e.g., Gonnerman et al., 2007; Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999; Seidenberg & Gonnerman, 2000) or used large corpora while purposely excluding misleading forms from the training, effectively simulating idealised, noise-free learning environments (e.g., Bonandrini et al., 2023; Marelli et al., 2017). Given this reliance on perfectly organised morpheme systems in previous computational investigations, it has been unclear what can actually be learned about derivational affixes when noise is not filtered from the input. We have argued previously that such noise is an inherent feature of the input readers encounter and that, because most readers are not linguists, it is far from a trivial distraction — in fact, it can be detrimental to affix learning (Korochkina & Rastle, 2025). In the present study, we implement this theoretical perspective computationally, demonstrating for the first time that, even without explicit distinctions between useful and misleading cases (e.g., *teacher* vs. *corner*) or guidance on handling varying degrees of informativeness in the input (e.g., *teacher* vs. *dresser*) characteristic of natural reading, a relatively simple model can form plausible hypotheses about the meanings of derivational affixes that align with how human readers process morphological information. More broadly, our research demonstrates the importance of analysing large corpora of naturally occurring texts to inform and build robust theories in language research. This section discusses these contributions and outlines directions for future research.

When we think about derivational morphology, we often default to straightforward examples such as *build* → *build<u>er</u>* or *kind* → *<u>un</u>kind*, where the relationship between the stem and the derived word, as well as the semantic contribution of the affix, is transparent. English

Table 2
Results of Analysis 2 for models incorporating affix-based metrics and the nonword-based metrics derived from the CAOSS models. The letter β denotes the *standardised* coefficient. Metrics that significantly improved the fit of the models in Analysis 1 (which included only affix-based metrics) are marked with an asterisk (*), and those with the highest standardised coefficients are shown in **bold**.

Dependent variable	Affix-based metrics in the model	Predictor	β	*SE*	*z*/*t*	*p*
Response accuracy	Affix richness	**Affix richness**	−0.46	0.05	−9.38	<.001
		Nonword richness*	−0.10	0.05	−2.11	.03
		Nonword neighb. density	−0.07	0.04	−1.72	.09
		Nonword diffuseness	−0.02	0.04	−0.41	.69
	Affix-word coherence	**Affix-word coherence**	−0.40	0.05	−8.24	<.001
		Nonword richness	−0.08	0.05	−1.67	.09
		Nonword neighb. density	−0.03	0.04	−0.79	.43
		Nonword diffuseness	−0.004	0.04	−0.10	.92
	Affix diffuseness	Affix diffuseness	0.23	0.05	4.34	<.001
		Nonword richness*	−0.25	0.04	−5.86	<.001
		Nonword neighb. density	−0.05	0.04	−1.07	.29
		Nonword diffuseness	−0.01	0.04	−0.25	.80
Response times	Affix richness	**Affix richness**	0.14	0.01	11.35	<.001
		Nonword richness	0.01	0.01	0.83	.41
		Nonword neighb. density*	0.05	0.01	4.50	<.001
		Nonword diffuseness	−0.02	0.01	−1.40	.16
	Affix-word coherence	**Affix-word coherence**	0.12	0.01	9.73	<.001
		Nonword richness	−0.001	0.01	−0.06	.95
		Nonword neighb. density*	0.03	0.01	2.72	.007
		Nonword diffuseness	−0.02	0.01	−1.82	.07
	Affix diffuseness	**Affix diffuseness**	−0.06	0.01	−4.68	<.001
		Nonword richness*	0.06	0.01	4.77	<.001
		Nonword neighb. density*	0.04	0.01	3.23	.001
		Nonword diffuseness	−0.02	0.01	−1.52	.13

morphology in particular is frequently described as relatively impoverished compared to that of other languages, reinforcing the impression that these predictable patterns are all it has to offer. However, the reality is more complex. A substantial body of research has highlighted that morphological regularities in English are often far less systematic and predictable than such simplified examples suggest (e.g., Aronoff, 1976; Berg & Aronoff, 2017; Chomsky, 1970; Korochkina & Rastle, 2025). For example, many English words are spelled in ways that obscure the relationship between their constituents and so require specialised linguistic knowledge to recognise the connection between the affix and the derived word (e.g., *sorcerer*, *recognise*, *deplore*, *gravity*). Conversely, some words have surface forms that suggest derivational relationships that are outright misleading (e.g., there is no *tail* in *detail*, *cord* in *record*, or *batter* in *battery*). Recognising this complexity, earlier research suggested that readers' knowledge of derivational regularities must reflect the graded nature of mappings between spelling and meaning, and attempted to simulate how these gradations are acquired using miniature approximations of existing morphological systems (e.g., Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999; Seidenberg & Gonnerman, 2000).

Recent analyses of large natural language corpora have advanced these earlier accounts by specifying how different aspects of real-world reading experience may contribute to derivational learning. Korochkina and Rastle (2025) argued that readers learn primarily from units they *perceive* as morphemes, and that this process is guided by an *orthographic* analysis of word structure, which may not align with a linguistically informed definition of morphemes. Under this view, words like *sorcerer*, whose constituent morphemes are not readily identifiable from spelling, do not aid in learning the *meanings* of affixes and so need to be excluded from consideration, leaving only words where morphological segmentation arises naturally from orthography. Among these, words where segmentation is meaningful (e.g., *decode*) will facilitate learning, whereas morphological false alarms where segmentation is misleading (e.g., *detail*) will hinder it. Korochkina and Rastle (2025) showed that affixes vary substantially in their frequency and false alarm rates, suggesting corresponding variation in readers' affix knowledge. In a follow-up study, they demonstrated that these distributional properties of individual affixes effectively predict variation in skilled readers' affix knowledge (Korochkina et al., 2026). This is an important step, as it demonstrates that the principles first proposed by distributed-connectionist approaches (e.g., Plaut & Gonnerman, 2000; Seidenberg & Gonnerman, 2000) and subsequently expanded by Korochkina and Rastle (2025) align with how readers process lexical information. However, this step remains a workaround: the metrics used in Korochkina et al. (2026) are essentially frequency counts, manually adjusted by the experimenter based on the presumed learning effect (beneficial or detrimental) associated with each word's morphological status (genuinely complex or a false alarm). A key limitation of this approach is that it requires expert linguistic intervention unavailable in real-world reading and relies on a binary distinction that fails to capture the nuanced semantic contributions of affixes.

The critical innovation of the present study is that it moves beyond these limitations, offering a more principled and scalable test of Korochkina and Rastle (2025)'s account of what drives derivational learning in the wild. Our study provides the first demonstration that a more naturalistic morpheme learning environment — one in which meaningful 'signal' co-occurs with misleading 'noise' — can be successfully simulated computationally. We achieved this using distributional semantic modelling, a framework long established as a reliable approximation of how readers derive meaning from words (e.g., Jones et al., 2006, Marelli, 2017, Marelli and Amenta, 2018, and see Mandera et al., 2017, for a review). We adopted a specific instantiation of this framework — a *compositional* distributional semantic model known as CAOSS (Marelli et al., 2017) — which is designed to capture the compositionality of meaning within words. While the use of CAOSS itself is not novel, as it has been successfully applied to morphological processing across various types of derivation and languages (e.g., Bonandrini et al., 2023; Günther & Marelli, 2023; Hsieh et al., 2025), the novelty of our approach lies in the training regime. Previous applications have systematically excluded 'noise' from their training data, thereby simulating an idealised learning environment in which the learner already knows what to attend to. In contrast, our model was trained on unfiltered data, thereby more accurately reflecting the uncertainty learners face in natural reading.

Our second key innovation concerns the inclusion of prefixes in our training set, making our study the first, to our knowledge, to apply the CAOSS model to prefixation. Prefixes differ from suffixes in several important ways, making the learning of prefixes arguably even less straightforward than that of suffixes. Prefixed words occur much less frequently in text than suffixed words, and although prefixes rarely trigger changes in the spelling of the words to which they attach, the most common English prefixes are of Latinate origin and tend to combine with Latinate roots (Korochkina & Rastle, 2025). As a result, a smaller proportion of prefixed than suffixed words can be recognised without specialised linguistic knowledge, limiting the number of words from which readers can infer prefix meanings through exposure to written language (Korochkina & Rastle, 2025). The exact meanings of prefixes are also often more difficult to pinpoint, as their contribution to word meaning is frequently more subtle and more complex than that of suffixes (Marchand, 1960). This is partly because suffixes often alter the lexical category of the base word (e.g., *kind* → *kindness*, *do* → *doable*), making their function more transparent. In contrast, prefixes typically modify meaning without changing grammatical class, and their conceptual contribution may align with that of adjectives (e.g., *co-host*, *ex-king*), adverbs (e.g., *unconscious*, *informal*), or even prepositions (e.g., *prewar*, *postgraduate*) (Marchand, 1960).

Our finding that the metrics derived from the CAOSS model can explain variance in participants' processing of novel morphemic combinations thus constitutes a significant innovation both in computational modelling of affix semantics and in research on human morphological processing more broadly. On the computational side, our finding demonstrates that the CAOSS model can handle ambiguous and noisy input and learn to position affixes along a continuum of semantic meaningfulness that aligns with patterns observed in human readers' processing of these affixes. Notably, the *affix-word coherence* metric derived from our models was as informative as the linguistically informed false alarm penalty metric proposed by Korochkina et al. (2026), while the *affix richness* metric even outperformed it. This result suggests that the model captured nuances of affix semantics beyond the adjusted-frequency workaround based on binary classification of affixed words as beneficial or detrimental, as used in our earlier work Korochkina et al. (2026). This achievement is particularly notable for prefixes, given their more complex interaction with stem meaning and the fact that our training data – reflecting the patterns observed in popular books – contained nearly twice as much data for suffixes as for prefixes.

In summary, the advantage of the CAOSS-derived metrics over the false alarm penalty metric is that the CAOSS model does not require a linguistic expert to determine which complex words are informative or misleading. Moreover, it can assess the contribution of an affix to the meaning of each complex word in a graded manner (e.g., *dresser* vs. *teacher*), something the false alarm penalty metric cannot do. Consequently, the modelling approach we used in this study provides a closer approximation of how derivational learning arises through reading experience and offers stronger support for the validity of the theoretical framework proposed by Korochkina and Rastle (2025). Because this approach more closely mirrors the psychological reality of derivational learning, it also raises the question of whether it could be refined to approximate that reality even more closely. In our study, the number of triplets provided for each affix naturally varied based on their availability in the 1,200 books we analysed, meaning that what the model learned about derivational affixation was directly shaped by the variability present in text. In other words, affixes that appeared in a wider range of word types in the training input exerted a greater influence on learning than affixes occurring in only a few types. While this aligns with research suggesting that type frequency is more important than token frequency for generalisation (e.g., Raviv et al., 2022; Tamminen et al., 2015), it remains an open empirical question whether the model's performance could be further improved by weighting the input according to the token frequency of individual types.

One aspect that distinguishes our approach from much of the existing work in morphology research is that, rather than beginning with linguistically defined morpheme categories, we adopt the perspective of a naïve reader and ask what can be inferred directly from the orthographic patterns available in the input. This perspective allows us to examine what readers might learn about different derivational morphemes – and what categories or distinctions emerge spontaneously from the input – without imposing expert linguistic assumptions from the outset. To illustrate, psycholinguistic research has a long tradition of classifying affixes as *productive* or *non-productive* and attempting to link these distinctions to lexical processing (e.g., Baayen, 1991, 1992; Hay & Baayen, 2002). Morphological productivity denotes the extent to which language users can create new morphologically complex words using derivational affixes (e.g., Baayen & Lieber, 1991; van Marle, 1985; Schultink, 1961). Productive affixes are thought to enable a potentially unlimited number of possible forms; for example, the suffix *-ness* is often described as a productive affix. In contrast, unproductive affixes (e.g., the prefix *en-* or the suffix *-ity*) permit only a fixed, countable set of forms.

A range of numerical approaches have been proposed to quantify an affix's productivity (e.g., Aronoff, 1976; Baayen, 1991, 1992; Baayen & Lieber, 1991). Yet, all such metrics rely on dictionary-based classifications to determine which derived forms to count, considering only those words that dictionaries label as morphologically complex, regardless of whether readers themselves recognise this complexity (e.g., *teacher* vs. *grocer*). They also fail to account for noise introduced by words that merely look affixed (e.g., *corner*). A further common feature of these measures is that they "reflect the linguist's intuitions concerning productivity" (e.g., Baayen & Lieber, 1991, p. 809) — that is, they are grounded in expert judgments about what is productive in the language. What these metrics do *not* take into account is that typical readers are not linguists, and their intuitions about morphological structure may diverge substantially from expert analyses. Nor do these metrics address *how* or *why* differences in affix processing and use might arise in the first place. A correspondence between the CAOSS model's assessment of affix meaningfulness and traditional notions of morphological productivity (e.g., *-ness* emerging as a 'good' affix from both perspectives) suggests that our approach may offer a *data-driven* foundation for the linguistically informed distinctions, grounded in real language use and human behaviour. Specifically, if an affix occurs with many distinct stems and few false alarms, readers are more likely to form robust meaning representations, which may make the affix more readily available for forming new words. We view this as a theoretically grounded hypothesis that warrants systematic investigation in future research.

This issue of setting aside the traditional linguistic lens directly bears on how we treat words whose stems are not free-standing English words (i.e., words with bound stems). Our modelling approach assumes that such words do not contribute to learning the *meaning* of the affix (and see Korochkina & Rastle, 2025, for further discussion). However, this does not imply that such words play no role in derivational learning more broadly. Repeated exposure to consistent patterns likely enhances general familiarity with the affix: for example, a reader may notice that in words like *preclude*, *include*, and *exclude*, the element *clude* co-occurs with *pre-*, *in-*, and *ex-*, which are also found in words whose stems can stand alone (e.g., *predate*, *insecure*, *extract*). Experience with such patterns may support *orthographic learning* by increasing readers' awareness of the affix and improving their ability to recognise it. In this way, words with bound stems may boost affix knowledge, even if they do not support learning its meaning or functional contribution in a compositional sense. The CAOSS model cannot account for such cases, as it requires explicit stem representations to model the affix meaning, and distributional semantic models cannot generate vector embeddings for bound stems, since these do not occur independently in natural language. However, this limitation is by design: the CAOSS model is intended to capture *compositional meaning*, not facilitative effects

of orthographic exposure. Given evidence that complex words with bound stems are processed differently from those with free-standing stems (e.g., Forster & Azuma, 2000; Pastizzo & Feldman, 2004), understanding how such words contribute to derivational learning remains an important direction for future research. We note, however, that addressing this question may require computational approaches that adopt more permissive assumptions about the basic units of analysis (e.g., Baayen et al., 2019).

In terms of morphological processing more broadly, our findings suggest that readers' experience with derivational affixes in print shapes their perception of both affix meaningfulness and the plausibility of unfamiliar morphemic combinations in which these affixes appear. The richer, more coherent with the meanings of the words in which it occurs, and less diffuse an affix's meaning was, the harder it was for our participants to reject the morphologically structured nonwords containing these affixes. For example, the CAOSS suffix model indicated that the meaning of the suffix *-ness* is richer, more coherent with the meanings of derived words using this suffix, and more specific than the meaning of the suffix *-ate*. Consistent with this modelling result, participants made more errors and took longer to reject nonwords containing *-ness* compared to those containing *-ate*. Our complementary analysis, which examined the meaningfulness of nonwords as a whole in addition to that of their affixes, indicated that readers were also sensitive to the richness of a nonword's meaning and its similarity to the meanings of its closest semantic neighbours. However, compared to the influence of the affix-word coherence and affix richness metrics, the impact of these nonword meaningfulness measures on participants' processing was considerably weaker. This suggests that skilled readers' judgments of affixed nonwords are likely to be driven mainly by the properties of the affixes they contain, rather than by the specific meaning of the stem-affix combination in each individual nonword. This result aligns with previous studies showing that the identification times of complex words in print are influenced not only by stem properties but also by affix characteristics (e.g., frequency or phonological changes introduced to stems upon affixation), which can sometimes outweigh the effects of stems (e.g., Bradley, 1979; Burani & Thornton, 2003; Vannest & Boland, 1999). It is also consistent with prior research indicating that, in the earliest stages of processing of morphologically structured nonwords, the overall interpretability of the item is irrelevant, as long as the nonword can be segmented into meaningful constituents (e.g., *sportation* priming *sport*; Longtin and Meunier (2005), Longtin et al. (2003), Rastle and Davis (2008); but see Meunier and Longtin (2007), for the role of semantic plausibility at later stages of processing).

One might argue at this point that the effects we observed were affected by our choice of stimuli. Since the affixes we used varied both in frequency and involvement in false alarms, the results might differ with a different set of affixes. However, this variation was necessary because our study tested the idea that readers' derivational knowledge varies according to their experience of affix distribution in text. Another concern could be that the nonwords in our study were created by randomly combining affixes and stems, meaning that we did not account for part-of-speech restrictions that apply to some affixes (e.g., the suffixes *-ful* and *-able* typically do not attach to adjectives). While this may have influenced our nonword meaningfulness metrics, this method of creating nonwords is consistent with approaches used in other studies employing the morpheme interference paradigm (e.g., Crepaldi et al., 2010). Whether skilled readers are sensitive to such part-of-speech constraints, and how this sensitivity affects their lexical processing, is a promising avenue for further study.

Another valuable direction for future research prompted by our study is to investigate how *affix-level* sensitivity emerges across development. The aim of the research presented here was to train a model on input characteristic of the reading experience of a typical young adult and to examine what can be learned about individual affixes based on this reading experience, validated against lexical processing data from adults. Consequently, our findings cannot inform understanding of derivational learning *over the course of development*. Research indicates that general sensitivity to derivational information in online processing tasks – such as morpheme interference (used in this study) or priming tasks – does not even emerge until late adolescence (e.g., Beyersmann et al., 2012; Dawson et al., 2018, 2021b). It therefore remains an open empirical question at what stage of development these effects first appear, and whether individual differences in text exposure influence the rate at which sensitivity to the meanings of individual affixes develops.

In conclusion, our work contributes to the field by providing the first demonstration of how derivational learning can emerge through natural reading. Our findings show that even under the noisy and variable conditions of reading 'in the wild' – that is, without explicit linguistic guidance about which aspects of the input should be prioritised and which disregarded – there is sufficient structural consistency in texts to support learning about the core functions of different affixes. Moreover, our results provide compelling evidence that this acquired sensitivity to the intricate patterns of affix usage in written language influences how readers process novel multimorphemic words. It is important to highlight that our work focused on what can be learned about derivational affixes through reading alone. Yet, readers will almost certainly approach this task with some prior knowledge of derivational morphemes, meaning that our estimates should be regarded as conservative. Further research is needed to evaluate the contribution of spoken language input to derivational learning, so as to supplement our estimates and provide a fuller account of how derivational knowledge develops. Our study focused on English, and we do not intend to make claims that extend beyond this language, as other languages and writing systems will naturally exhibit different patterns. However, we believe that our general approach offers a useful framework for cross-linguistic morphology research. Corpus-based methods are uniquely positioned to reveal how frequently specific patterns occur in natural language and therefore provide a crucial quantitative dimension for developing theories grounded in real language use, thereby enhancing the ecological validity of psycholinguistic research. This article illustrates how the integration of corpus analysis, computational modelling, and empirical research can meaningfully advance the field, and we hope this approach will inspire further work.

CRediT authorship contribution statement

Maria Korochkina: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marco Marelli:** Writing – review & editing, Methodology, Conceptualization. **Kathleen Rastle:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

Funding

MK and KR were supported by a research grant from the Economic and Social Research Council, United Kingdom (ES/W002310/1). MM was supported by a research grant from the European Union (ERC-COG-2022, BraveNewWord, 101087053). The views and opinions expressed in this article are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Declaration of competing interest

The authors have no competing interests to declare.

Supplementary material

The supplementary material for this article provides a detailed breakdown of all triplet counts per affix used to train the CAOSS models. It is available on this project's page on the Open Science Framework (https://osf.io/sf2bh/).

Data availability

All data, code, and materials associated with this article are available on this project's page on the Open Science Framework: https://osf.io/sf2bh/.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov, & F. Csáki (Eds.), *2nd international symposium on information theory* (pp. 267–281). Akadémiai Kiadó.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705.

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, *3*, 1–12. http://dx.doi.org/10.3389/fpsyg.2012.00232.

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. http://dx.doi.org/10.1075/ml.00014.ame.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge: MIT Press.

Baayen, H. R. (1991). Quantitative aspects of morphological productivity. In G. Booij, & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 109–149). Kluwer Academic Publisher.

Baayen, H. R. (1992). On frequency, transparency and productivity. In G. Booij, & J. van Marle (Eds.), *Yearbook of Morphology 1992* (pp. 181–208). Springer.

Baayen, H. R., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, *2019*(1), Article 4895891. http://dx.doi.org/10.1155/2019/4895891.

Baayen, H. R., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, *29*(5), 801–843. http://dx.doi.org/10.1515/ling.1991.29.5.801.

Baayen, H. R., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Data Base on CD-ROM*. Linguistic Data Consortium.

Baroni, M., Bernardi, R., & Zamparelli, R. (2014). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies*, *9*(6), 5–110, URL https://hdl.handle.net/11572/98426.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious mixed models. http://dx.doi.org/10.48550/arXiv.1506.04967.

Behzadnia, A., Ziegler, J. C., Colenbrander, D., Bürki, A., & Beyersmann, E. (2023). The role of morphemic knowledge during novel word learning. *Quarterly Journal of Experimental Psychology*, *77*(8), 1620–1634. http://dx.doi.org/10.1177/17470218231216369.

Berg, K., & Aronoff, M. (2017). Self-organization in the spelling of English suffixes: The emergence of culture out of anarchy. *Language*, *93*(1), 37–64. http://dx.doi.org/10.1353/lan.2017.0000.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177. http://dx.doi.org/10.1080/00437956.1958.11659661.

Beyersmann, E., Castles, A., & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *The Quarterly Journal of Experimental Psychology*, *65*(7), 1306–1326. http://dx.doi.org/10.1080/17470218.2012.656661.

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. http://dx.doi.org/10.1162/tacl_a_00051.

Bonandrini, R., Amenta, S., Sulpizio, S., Tettamanti, M., Mazzucchelli, A., & Marelli, M. (2023). Form to meaning mapping and the impact of explicit morpheme combination in novel word processing. *Cognitive Psychology*, *145*, Article 101594. http://dx.doi.org/10.1016/j.cogpsych.2023.101594.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *26*(2), 211–243. http://dx.doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Bradley, D. C. (1979). Lexical representation of derivational relation. In M. Aronoff, & M. L. Kean (Eds.), *Juncture* (pp. 37–55). MIT Press.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (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), http://dx.doi.org/10.3389/fpsyg.2016.01116.

Burani, C., & Thornton, A. M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. In H. R. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 157–208). Mouton de Gruyter.

Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review*, *40*(2), 287–309. http://dx.doi.org/10.17763/haer.40.2.y7u0242x76w05624.

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

Dawson, N., Hsiao, Y., Tan, A. W. M., Banerji, N., & Nation, K. (2023). Effects of target age and genre on morphological complexity in children's reading material. *Scientific Studies of Reading*, *27*, 529–556. http://dx.doi.org/10.1080/10888438.2023.2206574.

Dawson, N., Rastle, K., & Ricketts, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(4), 645–654. http://dx.doi.org/10.1037/xlm0000485.

Dawson, N., Rastle, K., & Ricketts, J. (2021a). Bridging form and meaning: Support from derivational suffixes in word learning. *Journal of Research in Reading*, *44*(1), 27–50. http://dx.doi.org/10.1111/1467-9817.12338.

Dawson, N., Rastle, K., & Ricketts, J. (2021b). Finding the man amongst many: A developmental perspective on mechanisms of morphological decomposition. *Cognition*, *211*, Article 104605. http://dx.doi.org/10.1016/j.cognition.2021.104605.

De Simone, E., Moll, K., & Beyersmann, E. (2024). Cross-linguistic differences in morphological processing: Evidence from English and Italian. *Scientific Studies of Reading*, *29*(2), 181–200. http://dx.doi.org/10.1080/10888438.2024.2413108.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. http://dx.doi.org/10.48550/arXiv.1810.04805.

Dinu, G., Pham, N. T., & Baroni, M. (2013). DISSECT — DIStributional SEmantics Composition Toolkit. In *Proceedings of the 51st annual meeting of the association for computational linguistics: system demonstrations* (pp. 31–36). Association for Computational Linguistics, URL https://aclanthology.org/P13-4006/.

Forster, K. I., & Azuma, T. (2000). Masked priming for prefixed words with bound stems: Does *submit* prime *permit*. *Language and Cognitive Processes*, *15*(4–5), 539–561. http://dx.doi.org/10.1080/01690960050119698.

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, *136*(2), 323–345. http://dx.doi.org/10.1037/0096-3445.136.2.323.

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. http://dx.doi.org/10.1177/1747021820902019.

Günther, F., & Marelli, M. (2023). CAOSS and transcendence: Modeling role-dependent constituent meanings in compounds. *Morphology*, *33*(4), 409–432. http://dx.doi.org/10.1007/s11525-021-09386-6.

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. http://dx.doi.org/10.1177/1745691619861372.

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

Hay, J., & Baayen, H. R. (2002). Parsing and productivity. In G. Booij, & J. van Marle (Eds.), *Yearbook of Morphology 2001* (pp. 203–235). Springer.

Hsieh, C.-Y., Marelli, M., & Rastle, K. (2025). Compositional processing in the recognition of Chinese compounds: Behavioural and computational studies. *Psychonomic Bulletin & Review*, 1–12. http://dx.doi.org/10.3758/s13423-025-02668-8.

Jones, M. N., Kintsch, W., & Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552. http://dx.doi.org/10.1016/j.jml.2006.07.003.

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In *Oxford handbook of mathematical and computational psychology* (pp. 232–254). Oxford University Press.

Korochkina, M., Cooper, H., Brysbaert, M., & Rastle, K. (2026). Morpheme knowledge is shaped by information available through orthography. *Psychonomic Bulletin & Review*, *33*, 3. http://dx.doi.org/10.3758/s13423-025-02830-2.

Korochkina, M., Marelli, M., Brysbaert, M., & Rastle, K. (2024). The Children and Young People's Books Lexicon (CYP-LEX): A large-scale lexical database of books read by children and young people in the United Kingdom. *Quarterly Journal of Experimental Psychology*, *77*(12), 2418–2438. http://dx.doi.org/10.1177/17470218241229694.

Korochkina, M., & Rastle, K. (2025). Morphology in children's books, and what it means for learning. *Npj Science of Learning*, *10*, 22. http://dx.doi.org/10.1038/s41539-025-00313-6.

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), 211–240. http://dx.doi.org/10.1037/0033-295X.104.2.211.

Lazaridou, A., Marelli, M., Zamparelli, R., & Baroni, M. (2013). Compositionally derived representations of morphologically complex words in distributional semantics. *vol. 1*, In *Proceedings of the 51st annual meeting of the association for computational linguistics* (pp. 31–36). Association for Computational Linguistics.

Longtin, C.-M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language*, *53*(1), 26–41. http://dx.doi.org/10.1016/j.jml.2005.02.008.

Longtin, C.-M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*(3), 313–334. http://dx.doi.org/10.1080/01690960244000036.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. http://dx.doi.org/10.1016/j.jml.2016.04.001.

Marchand, H. (1960). *The categories and types of present-day english word-formation: A Synchronic-Diachronic approach*. Otto Harrassowitz, Wiesbaden.

Marelli, M. (2017). Word-embeddings Italian semantic spaces: A semantic model for psycholinguistic research. *Psihologija*, *50*(4), 503–520. http://dx.doi.org/10.2298/PSI161208011M.

Marelli, M., & Amenta, S. (2018). A database of orthography-semantics consistency (OSC) estimates for 15,017 English words. *Behavior Research Methods*, *50*, 1482–1495. http://dx.doi.org/10.3758/s13428-018-1017-8.

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*(3), 485–515. http://dx.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. http://dx.doi.org/10.1016/j.cognition.2017.05.026.

van Marle, J. (1985). *On the paradigmatic dimension of morphological creativity*. Foris.

Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. *Quarterly Journal of Experimental Psychology*, *64*(6), 1200–1220. http://dx.doi.org/10.1080/17470218.2010.538211.

Meunier, F., & Longtin, C.-M. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, *56*(4), 457–471. http://dx.doi.org/10.1016/j.jml.2006.11.005.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *vol. 26*, In *Advances in neural information processing systems* (pp. 3111–3119). Curran Associates, http://dx.doi.org/10.48550/arXiv.1310.4546.

Mirkovic, J., Vinals, L., & Gaskell, M. G. (2019). The role of complementary learning systems in learning and consolidation in a quasi-regular domain. *Cortex*, *116*, 228–249. http://dx.doi.org/10.1016/j.cortex.2018.07.015.

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, *34*(8), 1388–1429. http://dx.doi.org/10.1111/j.1551-6709.2010.01106.x.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, (320), 1191–1195. http://dx.doi.org/10.1126/science.1152876.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*(3), 304–330. http://dx.doi.org/10.2307/747823.

Nathaniel, U., Eidelsztein, S., Girsh Geskin, K., Yamasaki, B. L., Nir, B., Dronjic, V., Booth, J. R., & Bitan, T. (2024). Neural mechanisms of learning and consolidation of morphologically derived words in a novel language: Evidence from Hebrew speakers. *Neurobiology of Language*, *5*(4), 864–900. http://dx.doi.org/10.1162/nol_a_00150.

Nunes, T., & Bryant, P. (2006). *Improving literacy by teaching morphemes*. Routledge.

van Paridon, J., & Thompson, B. (2021). *subs2vec*: Word embeddings from subtitles in 55 languages. *Behavior Research Methods*, *53*(2), 629–655. http://dx.doi.org/10.3758/s13428-020-01406-3.

Pastizzo, M. J., & Feldman, L. B. (2004). Morphological processing: A comparison between free and bound stem facilitation. *Brain and Language*, *90*(1–3), 31–39. http://dx.doi.org/10.1016/S0093-934X(03)00417-6.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*(4–5), 445–485. http://dx.doi.org/10.1080/01690960050119661.

R. Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, URL https://www.R-project.org.

Rastle, K. (2019). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54. http://dx.doi.org/10.1016/j.cortex.2018.02.008.

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

Rastle, K., & Taylor, J. (2018). Print-sound regularities are more important than print-meaning regularities in the initial stages of learning to read: Response to Bowers & Bowers (2018). *Quarterly Journal of Experimental Psychology*, *71*(7), 1501–1505. http://dx.doi.org/10.1177/1747021818775053.

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, *26*, 462–483. http://dx.doi.org/10.1016/j.tics.2022.03.007.

Rueckl, J. G., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language*, *68*(1–2), 110–117. http://dx.doi.org/10.1006/brln.1999.2106.

Schultink, H. (1961). Produktiviteit als morfologisch fenomeen. *Forum der Letteren*, *2*, 110–125.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464, URL https://www.jstor.org/stable/2958889.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*(9), 353–361. http://dx.doi.org/10.1016/S1364-6613(00)01515-1.

Stevens, P., & Plaut, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychonomic Bulletin & Review*, *29*(5), 1673–1702. http://dx.doi.org/10.3758/s13423-022-02086-0.

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

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition*, *125*(1), 107–112. http://dx.doi.org/10.1016/j.cognition.2012.06.014.

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, *79*, 1–39. http://dx.doi.org/10.1016/j.cogpsych.2015.03.003.

Treiman, R., & Cassar, M. (1996). Effects of morphology on children's spelling of final consonant clusters. *Journal of Experimental Child Psychology*, *63*(1), 141–170. http://dx.doi.org/10.1006/jecp.1996.0045.

Treiman, R., Wolter, S., & Kessler, B. (2020). How sensitive are adults to the role of morphology in spelling? *Morphology*, *31*, 261–271. http://dx.doi.org/10.1007/s11525-020-09356-4.

Tucker, R., Castles, A., Laroche, A., & Deacon, S. H. (2016). The nature of orthographic learning in self-teaching: Testing the extent of transfer. *Journal of Experimental Child Psychology*, *145*, 79–94. http://dx.doi.org/10.1016/j.jecp.2015.12.007.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188. http://dx.doi.org/10.1613/jair.2934.

Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2020). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition*, *195*, Article 103810. http://dx.doi.org/10.1016/j.cognition.2018.09.013.

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

Vannest, J., & Boland, J. E. (1999). Lexical morphology and lexical access. *Brain and Language*, *68*(1–2), 324–332. http://dx.doi.org/10.1006/brln.1999.2114.

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*, 1371–1398. http://dx.doi.org/10.3758/s13428-018-1118-4.

Zwitserlood, P., Bölte, J., & Dohmes, P. (2000). Morphological effects on speech production: Evidence from picture naming. *Language and Cognitive Processes*, *15*(4–5), 563–591. http://dx.doi.org/10.1080/01690960050119706.