



ELSEVIER

Contents lists available at ScienceDirect

Cognition

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

COGNITION

Original Articles

Skilled readers' sensitivity to meaningful regularities in English writing

Anastasia Ulicheva^{a,*}, Hannah Harvey^a, Mark Aronoff^b, Kathleen Rastle^a^a Department of Psychology, Royal Holloway, University of London, UK^b Department of Linguistics, Stony Brook University, USA

ARTICLE INFO

Keywords:

Writing system
Spelling
Meaning
Morphology
English
Lexical category

ABSTRACT

Substantial research has been undertaken to understand the relationship between spelling and sound, but we know little about the relationship between spelling and meaning in alphabetic writing systems. We present a computational analysis of English writing in which we develop new constructs to describe this relationship. Diagnosticity captures the amount of meaningful information in a given spelling, whereas specificity estimates the degree of dispersion of this meaning across different spellings for a particular sound sequence. Using these two constructs, we demonstrate that particular suffix spellings tend to be *reserved* for particular meaningful functions. We then show across three paradigms (nonword classification, spelling, and eye tracking during sentence reading) that this form of regularity between spelling and meaning influences the behaviour of skilled readers, and that the degree of this behavioural sensitivity mirrors the strength of spelling-to-meaning regularities in the writing system. We close by arguing that English spelling may have become fractionated such that the high degree of spelling-sound inconsistency maximises the transmission of meaningful information.

1. Introduction

One, two! One, two! And through and through
The vorpal blade went snicker-snack!

Lewis Carroll, *Jabberwocky*, 1871/1983

Recent research has uncovered substantial patterns of non-arbitrariness in language (see [Dingemanse, Blasi, Lupyán, Christiansen, & Monaghan, 2015](#), for review). Corpus studies reveal subtle phonological differences between words that mean different things: for example, words referring to objects are composed of different sounds and are stressed in different positions than words referring to actions ([Cassidy & Kelly, 1991](#); [Monaghan, Shillcock, Christiansen, & Kirby, 2014](#)). These phonological cues play an important role in language acquisition ([Fitneva, Christiansen, & Monaghan, 2009](#)) and aid syntactic processing in adulthood ([Farmer, Christiansen, & Monaghan, 2006](#); [Monaghan, Christiansen, Farmer, & Fitneva, 2012](#)). These findings go against the traditional perspective that the links between phonological forms and concepts are arbitrary ([De Saussure, 1983](#)).

In this article, we investigate the nature of statistical regularities between the *written* forms of English words and their meanings, focusing on derivational suffixation. These units capture the commonality between known words that are related in meaning (e.g. as in ‘actress’, ‘waitress’) and provide a basis for language productivity, allowing us to label new concepts and phenomena in the ever-changing world (e.g.

‘Trumpists’ and ‘Brexiters’; see [Algeo, 1991](#)). We present a computational analysis of English writing in which we develop new constructs to describe the relationship between spelling and meaning. This analysis demonstrates that particular suffix spellings tend to be *reserved* for particular meaningful functions. We then show across three paradigms that skilled readers are highly sensitive to this form of regularity, and that this sensitivity mirrors the strength of spelling-to-meaning regularities in the writing system. We close by arguing that English spelling may have become fractionated in such a way that the high degree of spelling-sound inconsistency maximises the transmission of meaningful information in the service of skilled reading.

1.1. Phonological cues to lexical categories

This paper focuses on one aspect of word meaning, specifically, lexical category (e.g. noun, adjective, verb; see [Baayen, Milin, Durdevic, Hendrix, & Marelli, 2011](#), for a similar conceptualisation of “meaning”). There is a growing body of research suggesting that statistical information about lexical category is encoded in spoken word forms. This research begins with [Kelly's \(1992\)](#) seminal work demonstrating that stress, syllable number, word duration, voicing, and vowel type provide probabilistic cues to lexical category. For example, in English, nouns tend to have more syllables than verbs and fewer consonants per syllable. Further, bisyllabic nouns are likely to have stress

* Corresponding author at: Department of Psychology, Royal Holloway University of London, Egham TW20 0EX, UK.

E-mail address: Ana.Ulicheva@rhul.ac.uk (A. Ulicheva).

<https://doi.org/10.1016/j.cognition.2018.09.013>

Received 9 February 2018; Received in revised form 31 August 2018; Accepted 22 September 2018

0010-0277/© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

on the first syllable while bisyllabic verbs have stress on the second syllable. In the last decade, about sixteen phonological cues to lexical category in English have been proposed in the literature (for a review, see Monaghan, Chater, & Christiansen, 2005). For example, Onnis and Christiansen (2008) conducted corpus analyses in several languages showing that word-final phonemes could discriminate lexical categories almost as well as suffixes.

Language speakers extract these probabilistic cues from their environment, and exploit them in language learning and language processing. Cassidy and Kelly (1991, 2001) and Fitneva et al. (2009) found that children learn object and action referents better if there is a correspondence between the sounds of the nouns and verbs and their respective lexical categories. Further, using a self-paced reading methodology, Farmer et al. (2006) found that when sentence context generated an expectation for a noun, phonologically-typical nouns (e.g. ‘marble’ whose phonological features cluster with those of other nouns) were read aloud faster than phonologically-atypical nouns (e.g. ‘insect’ whose phonological features cluster with those of verbs). Information about lexical category is critical for comprehending language (e.g. ‘we saw her duck’ is interpreted differently depending on whether ‘duck’ acts as a noun or a verb), and for producing syntactically valid utterances. It is also well known that the ability to pick up and exploit cues to lexical category is very important in language acquisition. Children quickly construct hypotheses about the lexical category of new words to be able to use these new words productively in sentences (Cassidy & Kelly, 1991, 2001; Houston-Price, Plunkett, & Harris, 2005; Markman & Wachtel, 1988).

If writing were a direct transliteration of spoken language, then these regularities between phonological form and meaning would also be present in written language. However, in English (as in many other alphabetic languages), there is not a straightforward relationship between the sounds and spellings of words. Because statistical regularities between printed words and their meanings may impact the acquisition of literacy and the skilled processing of printed words (e.g. Frost, 2012), it is important to consider the nature of these regularities themselves. Our work explores the possibility that adults extract probabilistic morpho-syntactic cues from written language and exploit them in reading and writing. We focus on morphology because it provides the site of the most substantive regularities between English printed words and their meanings (Plaut & Gonnerman, 2000; Rastle, Davis, Marslen-Wilson, & Tyler, 2000).

1.2. Morphological regularities and skilled reading

Four decades of research have demonstrated that skilled readers register and exploit morphological regularities in printed words. Taft and Forster’s (1975) seminal research demonstrated that nonwords with an apparent morphological structure (such as ‘dejuvenate’ that consists of an existing prefix DE- and stem¹ JUVEN as in ‘juvenile’, ‘rejuvenate’) take longer to reject in a lexical decision task compared to nonwords without such structure (e.g. ‘depertoire’ whose “stem” does not exist). This effect is thought to arise because morphologically-structured nonwords like ‘dejuvenate’ contact stored representations of morphemes (in this case, of JUVEN), making it difficult to decide that the stimulus is not an existing word. This finding is interesting because the stem JUVEN does not occur on its own, and it is unlikely that readers are ever explicitly taught that JUVEN has something to do with ‘youth’. Rather, this knowledge likely reflects an accumulation of experience with spoken and written language, which ultimately becomes represented in the reading system.

Since the publication of Taft and Forster (1975), a wealth of converging data from a variety of experimental paradigms has supported

¹ A word or a word segment from which other words are formed through derivational and inflectional suffixation.

the conclusion that the analysis of morphological information is a crucial part of skilled reading. One of the most powerful pieces of evidence for this comes from morphological priming experiments in which primes are masked and presented so briefly that they are not available for conscious report. These experiments have revealed that derivational primes (e.g. ‘dealer’) facilitate recognition of stand-alone stems (e.g. DEAL; Forster, Davis, Schoknecht, & Carter, 1987; Rastle et al., 2000). It is noteworthy that this priming effect cannot be ascribed to simple orthographic overlap between prime and target (e.g. *brothel*-BROTH; –EL is not an existing morpheme and yields no priming; Rastle, Davis, & New, 2004), or to the simple summation of orthographic and semantic overlap (e.g. *sneeze*-SNORE yields no priming; Rastle et al., 2000). Results from studies that use the stem frequency paradigm nicely complement this line of work. It has been shown repeatedly that words with frequent stems are processed more quickly than words with infrequent stems, even when the frequency of the whole word is controlled. For example, Niswander, Pollatsek, and Rayner (2000) manipulated stem frequency while keeping whole word frequency constant, and found early effects of stem frequency on readers’ eye-movement behaviour: first fixation durations were shorter on words with frequent stems such as ‘instalment’ compared to those with infrequent stems such as ‘deferment’ (see also Andrews, 1986; Juhasz, Starr, Inhoff, & Placke, 2003; Pollatsek & Hyönä, 2005). Both types of results suggest that morphologically-structured words are represented in terms of their constituents; they are represented in a componential manner (Plaut & Gonnerman, 2000). These data again indicate that the accumulation of experience with printed words impacts on the nature of stored representations in the reading system.

In contrast to statistical regularities between print and sound, there has been remarkably little work seeking to specify how morphemic regularities between print and meaning are acquired. Recent cross-sectional work investigating groups of children between 7 and 9, 12–13, and 16–17 years old has revealed ongoing accumulation of morphological knowledge throughout this whole period of reading acquisition (Dawson, Rastle, & Ricketts, 2017), although less is known about the mechanisms that underpin the acquisition of this knowledge. Tamminen, Davis, and Rastle (2015) sought to determine whether this information could be discovered implicitly through exposure to morphologically structured stimuli. They presented adults with novel vocabularies comprising an underlying morphological structure (e.g. ‘sleepnule’, ‘buildnule’, ‘teachnule’). Critically, the definitions of these items related back to the stems (in this case, SLEEP, BUILD, and TEACH) in such a way that participants could discover an underlying function of the novel affix (e.g. that it refers to an agent). Following training and a period of memory consolidation, participants were probed for their knowledge about the novel affixes, and invited to generalize this knowledge to untrained words in speeded reading situations. Results showed that participants had extracted the underlying morphological regularities and encoded them in long-term memory in such a way that they could be used to facilitate rapid lexical processing (see also Mirković & Gaskell, 2016; Mirković, Forrest, & Gaskell, 2011).

1.3. Morphological regularities in English writing

This brief review demonstrates that readers learn to appreciate underlying morphological regularities through their experience with words, and that they use this knowledge in the service of rapid skilled reading. However, much less is known about the nature of this knowledge. When readers are accumulating morphological knowledge, what precisely is it that they are accumulating? The answer to this question requires a deep understanding of how morphological relationships impact on the mapping between print and meaning. Previous observations have highlighted the fact that morphology brings a degree of regularity to the otherwise-arbitrary mapping between print and meaning (e.g. Plaut & Gonnerman, 2000). While words that look similar generally are not similar in meaning (e.g. ‘cat’, ‘cut’, ‘can’),

stems occur in words with similar meanings (e.g. ‘darkness’, ‘darkly’), and affixes alter the meanings of words in highly-predictable ways (e.g. ‘cleaner’, ‘teacher’, ‘banker’; Rastle et al., 2000). However, what type of information do affixes possess that leads them to have this impact on the writing system?

It has long been known that affixes can provide powerful cues to lexical category (e.g. Fabb, 1988). For example, the suffix –NESS² attaches to adjectives to form nouns (e.g. ‘kind’ → ‘kindness’; ‘bald’ → ‘baldness’), whereas the suffix –IFY attaches to adjectives and nouns to form verbs (e.g. ‘pure’ → ‘purify’; ‘class’ → ‘classify’). However, the strength with which affixes indicate lexical category has never been precisely articulated. Indeed, most of the previous work on form–category relationships has focussed on phonological regularities in words with a single morpheme (e.g. Farmer et al., 2006; also see Monaghan et al., 2012, Monaghan, Christiansen, & Fitneva, 2011, and Onnis & Christiansen, 2008, for corpus analyses).

Morphological regularities exist in both spoken and written language. However, Berg and Aronoff (2017) have recently provided quantitative evidence that written forms of four English suffixes (–OUS, –AL, –Y, –IC) carry unique information about lexical category that is not available in their phonological forms. Consider the word-final phonological string /–əs/ (Table 1). This string occurs in adjectives and non-adjectives with an approximately equal probability (‘marvellous’, ‘cactus’; compare the rows of the table). However, when we look at the spellings (compare the table columns), these probabilities change greatly. Although in general, there are various ways to spell the /–əs/ sound sequence in English (e.g. ‘bonus’, ‘atlas’, ‘nervous’), adjectives are virtually always spelled –OUS (see Table 1), and other spellings must be used for words of other lexical categories. The suffix spelling –OUS (but not its phonological counterpart /–əs/) therefore appears to be reserved for adjectives.

Similar dependency characterizes English inflectional suffixes. For example, it is well known that spelling –ED in morphologically-complex words is strongly indicative of past tense (e.g. Carney, 1994). This suffix can be pronounced /–əd/. However, what has only recently come to light is that this word-final sound sequence cannot be spelled as –ED when it occurs in morphologically-simple words that could be mistaken for morphologically complex. In such words, a different spelling is typically used (e.g. ‘instead’; Berg, Buchmann, Dybiec, & Fuhrhop, 2014). In fact, there are only a handful of exceptions to this dependency (e.g. seemingly complex but genuinely monomorphemic words ending in /–əd/ and spelled with –ED, ‘moped’, ‘naked’, ‘wicked’; Berg et al., 2014). Thus, the spelling –ED is used almost exclusively to indicate the past. One consequence of this dependency is that the relationship between form and meaning appears to be stronger in written language than in spoken language. In the –ED /–əd/ example, for instance, /–əd/ does not carry the same amount of information about the morphological status of a word as –ED does. This observation goes against a commonly held belief that written language is somewhat impoverished compared to spoken language, in that, for instance, speech is characterized by intonation and stress, whereas this information is lost in writing (e.g. Seidenberg, 2017).

These observations suggest that the relationship between morphemes and meanings may be especially salient in writing. One could speculate that English spelling has evolved in such a way as to maximise the transmission of meaningful information through suffixation. Of the four derivational suffixes studied by Berg and Aronoff (2017), only –OUS was found to be an unambiguous marker of its lexical category, but other suffixes also marked meaning in some way. For example, the spelling –AL communicates adjective or noun status (i.e. not a verb), and adjectives ending in the sound sequence /–əl/ are most likely spelled –AL. Might these observations help us to understand why Lewis Carroll chose the spelling ‘vorpal’ as opposed to ‘vorple’ to describe the

Table 1

The dependency between the spelling of /–əs/ and lexical category (adapted from Berg & Aronoff, 2017).

		–OUS spelling	Other spelling
Number of words	Adjectives	346 (‘marvellous’)	6 (‘citrus’)
	Not adjectives	0	314 (‘cactus’)

young hero’s blade in *Jabberwocky*? Further, is it possible that readers are sensitive to the meaningful information that these different spellings convey, and use this information to assist their reading and spelling behaviour?

In this paper, we present a computational analysis that characterizes the relationship between spelling and lexical category across *all* derivational suffixes in English. This analysis allows us to determine whether this dependency is a general property of the English writing system or an idiosyncrasy limited to a few suffixes. In the light of findings of Berg and Aronoff (2017), we hypothesized that these dependencies are not all-or-nothing, but are graded in nature. We then investigate the possibility that readers capture and exploit probabilistic information about lexical category contained in suffixes when they read and spell (see also Kemp, Nilsson, & Arciuli, 2009).

2. Computational analysis

We propose that the relationship between spelling and lexical category can be described using two concepts – *diagnosticity* and *specificity*. Diagnosticity of a spelling describes how well that spelling predicts lexical category. For example, Berg and Aronoff (2017) observed that the spelling –OUS is strongly associated with adjective status. The specificity of a spelling indicates the extent to which that spelling is the preferred means of designating a particular lexical category for a given sound sequence, relative to other spellings. For example, Berg and Aronoff (2017) observed that given the sound sequence /–əs/, adjectives are virtually always spelled –OUS.

If we apply these new constructs to the linguistic analysis of Berg and Aronoff (2017), we see that the written suffix –OUS is both highly diagnostic for adjectives (its written form is *never* found in nouns or verbs; see Table 1) and highly specific for adjectives (if an /–əs/-word is not an adjective, a different spelling should be used, e.g. ‘bonus’; see Table 1). In this case, the written suffix –OUS appears to be reserved for adjectives. Diagnosticity on its own is reminiscent of what is referred to as “cue validity” (conditional probability of a category given a cue; Beach, 1964). Conceptually, diagnosticity and specificity are distinct theoretical constructs, since one describes the mapping from spelling to meaning, while the other characterises the mapping from meaning to spelling. But in practice, are both of them necessary? It is a valid theoretical possibility that diagnosticity and specificity always co-vary, thus dichotomising the English derivational space into two categories (diagnostic and specific vs not diagnostic and not specific). Here we hypothesise that the bidirectional relationship between the form and the meaning cannot be captured using a unitary concept (e.g. cue validity), and needs to be expressed through the *combination* of diagnosticity and specificity (probability of a cue given a category).

Berg and Aronoff (2017) investigated only four suffixes. The aim of our computational analysis is to determine the strength of the relationship between English suffixes and lexical categories in general, and to operationalise this relationship in terms of diagnosticity and specificity. There are multiple potential ways to measure diagnosticity and specificity. One could, for example, use a naïve discriminative learning (NDL) approach, similar to that used by Baayen et al. (2011). These authors used the Rescorla-Wagner rule to estimate how predictive low-level orthographic cues (unigrams and bigrams) were of several meaning outcomes in Serbian and in English. These meaning outcomes included case marking (e.g. “nominative”, “genitive”) and number in Serbian, and

² We use capital letters to represent the written form of a suffix.

localist representations of words (e.g. ‘house’) and affixes (e.g. –NESS) in English. Though Baayen et al. (2011) did not focus on the relationship between orthographic cues and grammatical categories, it may be possible to extend their approach to cover this problem. In our paper, we chose an alternative approach that relies on explicit morpheme representations. We measured the amount of meaningful (i.e. category) information in these morphemes and determined if our estimates explain human behaviour. Whether or not estimates obtained through a different approach, such as NDL modelling, will have a better fit to behavioural data is an open empirical question.

2.1. Method

We extracted all single-word entries from the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993), excluding any entries with spaces or hyphens. Separate entries were created for words that could express different lexical categories (e.g. ‘play’ as a noun or as a verb). Using this list of words as our database, we extracted all spellings annotated as derivational suffixes in CELEX that occur in at least three different words. This process revealed 154 written suffixes in total.

The computation of diagnosticity and specificity required us to extract all words containing the 154 suffixes identified. However, in addition to extracting genuinely suffixed words, we also needed to capture words in which the suffix spellings do not function as genuine morphemes (e.g. the –ER in ‘corner’). We included such pseudoaffixed words because there is evidence that people identify morphemes within these (Rastle et al., 2004). In order to identify these words, we extracted all words containing the spelling of each suffix, as long as the remaining (non-suffix) portion comprised a legitimate stem with at least one phonological vowel (including stems with alternations, i.e. orthographically conditioned variations, such as COMPLI-/COMPLY–). This process meant that words like ‘vanish’ were extracted for the –ISH suffix but that words like ‘dish’, ‘accomplish’, and ‘catfish’ were not, because these do not contain a legitimate stem.³ The number of words extracted for each suffix is presented in Appendix A. Our analyses consider three major lexical categories: nouns, adjectives, and verbs. Only the most frequent pronunciation was considered for each suffix.

We define diagnosticity as the strength with which a suffix spelling predicts lexical category. For a particular suffix, type diagnosticity is calculated by dividing the number of words ending in this suffix and falling into this category by the total number of words that contain the suffix. Diagnosticity is thus expressed as a ratio ranging from 0 to 1. For example, –NIK is not diagnostic for adjectives (diagnosticity is 0), whereas –TUDE unambiguously signals a noun (diagnosticity is 1). A suffix like –IAN is present in nouns (‘magician’), as well as in other categories (‘Amazonian’), and so is not particularly diagnostic for nouns (diagnosticity of 0.54). A second measure of diagnosticity, based on word tokens, was also calculated. This measure was obtained by dividing the summed frequency of words ending in the suffix and falling into a particular category by the summed frequency of all words that contain the suffix.

We define specificity as the extent to which a suffix spelling is preferred for denoting a particular lexical category, amongst all possible spellings of a sound sequence. In order to calculate specificity, we identified the most frequent pronunciation of each suffix, and then extracted all words from our single-word database containing that word-final sound sequence. Note that CELEX lists surface phonological forms (not phonetic/allophonic forms). We stayed as close to the CELEX forms as we could to avoid theoretical controversies. For example, the sound sequence /-ləs/⁴ associated with suffix –LESS also occurs in

words with other word-final spellings (e.g. ‘necklace’, ‘malice’). For words of a given lexical category that end in a given sound (e.g. all /-ləs/-adjectives), we identified what proportion of these contain the suffix spelling (e.g. –LESS). Like diagnosticity, specificity is expressed as a ratio ranging from 0 to 1. Spellings with the specificity of 0 for a given category never go with this category (e.g. –IOUS never occurs in /-əs/-nouns cf. ‘bonus’), whereas the value of 1 indicates that the spelling has full specificity (e.g. when a noun ends in /-ɪtə/, it must be spelled with –ATOR). Medium values of specificity indicate that there are multiple competing spellings for a given category and sound, as with –IER and –EER that can both be used in /-ɪə/-nouns (cf. ‘cashier’, ‘engineer’; specificity of 0.49 and 0.46, respectively). A second measure of specificity, based on word tokens, was obtained by dividing the summed frequency of words ending in the phonological and orthographic sequence and falling into a particular lexical category by the summed frequency of all words that contain the spelling and fall into the category.

Four examples of suffixes that differ on diagnosticity and specificity are given in Fig. 1. In this figure, we can see that –ICAL in the upper-left panel is reserved for communicating adjective status. If a word contains the –ICAL spelling, it is almost certainly an adjective (i.e. highly diagnostic); and if the sound sequence /-ɪkəl/ is an adjective, then it must be spelled –ICAL (i.e. highly specific). However, it is clear from these four examples that these properties can vary orthogonally in English suffixes.

2.2. Results

Fig. 2 plots diagnosticity and specificity values for English derivational suffixes. The majority of suffixes are either diagnostic or specific or both. Mean type diagnosticity for suffixes is 0.78 (the corresponding token value is 0.77), mean specificity is 0.82 (token is 0.83). The correlation between the type variables is 0.53, $p < 0.0001$; that between the token variables is 0.47, $p < 0.0001$.⁵ This indicates that the two constructs are related although they capture distinct properties of suffixes. The full range of diagnosticity and specificity values are contained in Appendix A.

It is immediately apparent that a substantial number of suffixes have diagnosticity and specificity values of 1. This means that these suffix spellings are *always* associated with a particular lexical category (diagnosticity) and that these suffix spellings are the *only means* of communicating a particular lexical category for a given word-final sound sequence (specificity). Is it possible that this very strong relationship between spelling and lexical category is due to the presence of suffixes in our sample that occur only rarely? In order to determine this, we excluded all suffixes that occur in < 20 different words in the CELEX database (Baayen et al., 1993). For the remaining 95 suffixes, the mean diagnosticity and specificity values were virtually unchanged (mean type and token diagnosticity values were 0.76; mean type specificity was 0.83; mean token specificity was 0.82).

2.3. Conclusion

We have characterised spellings of all English suffixes in terms of diagnosticity (the extent to which suffix spellings predict lexical category) and specificity (the extent to which suffix spellings are preferred for denoting a particular lexical category for a given sound sequence). Our analyses demonstrate that the majority of written English

³ We repeated these analyses including words without a plausible morphological structure (e.g. ‘dish’). The results were similar, see Footnote 5.

⁴ Throughout the paper, the IPA transcription is used, which sometimes differs from the coding used in CELEX (e.g. /ləs/ is coded as /lɪs/ in CELEX; /ɪkəl/ is coded as /ɪkP/ in CELEX).

⁵ If all words are included in the calculations of diagnosticity and specificity, the values drop only slightly. Mean diagnosticity is 0.75 for all suffixes (or 0.73 for frequent suffixes only). Mean specificity is 0.74 for all suffixes (or 0.76 for frequent suffixes only). The correlation between the variables is 0.44, $p < 0.0001$. A figure analogous to Fig. 2 that is based on these values can be found on this project’s OSF storage.

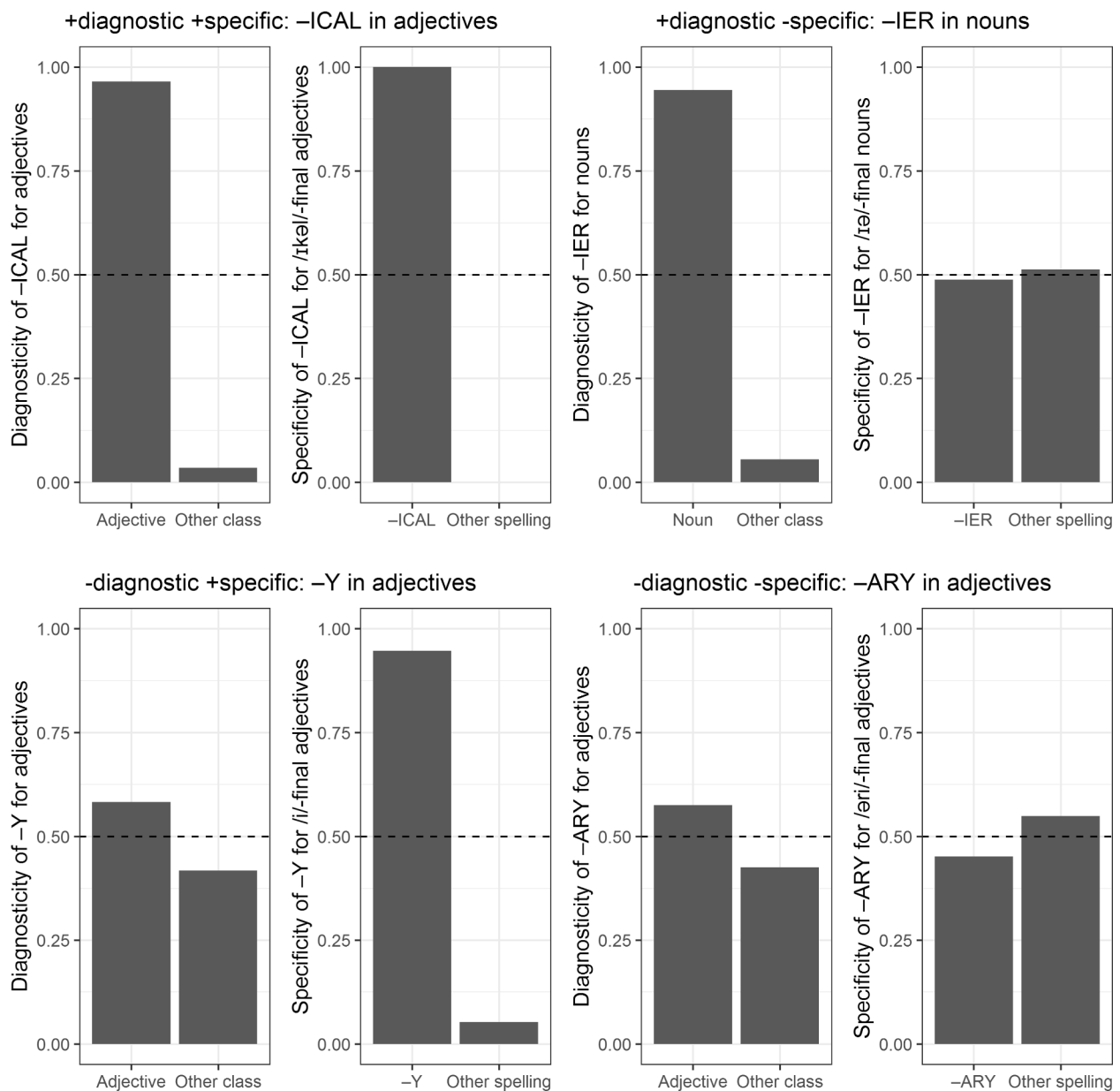


Fig. 1. Diagnosticity and specificity illustrated using four examples.

derivational suffixes are diagnostic and specific for some lexical category. That is, suffix spellings signal the lexical category of their carrier word, and on the other hand, are preferred spellings for denoting this lexical category. This combination of measures is important because it gives rise to the intriguing hypothesis that the proliferation of spellings for particular sound sequences in English has allowed certain spellings to become reserved for the communication of meaningful information.

Our analysis extends and deepens the work of Berg and Aronoff (2017) regarding the properties of four English suffixes. We have demonstrated that our concepts diagnosticity and specificity can be used to describe the bidirectional dependency between suffix spelling and lexical category, and we have shown that this dependency appears to be ubiquitous in English derivational suffixes. In what follows, we investigate to what extent skilled readers extract these regularities from their experience with the writing system, and use this knowledge when they read and spell. In our paper, we focus solely on what are usually referred to as “feedforward” effects. That is, we hypothesise that diagnosticity, a measure of spelling-to-meaning consistency, would impact on tasks that involve the mapping of print to meaning (i.e.

semantic categorisation, reading), while specificity, a measure of meaning-to-spelling consistency, would impact on spelling behaviour. That said, perception and production are intertwined processes, and one might expect “feedback” effects of diagnosticity and specificity as well (e.g. diagnosticity influencing the production of spelling; see Ziegler, Petrova, & Ferrand, 2008; Kessler, Treiman & Mullennix, 2008, for similar findings in reading aloud). This is an issue that needs to be investigated in future studies using appropriate experimental designs (e.g. a two-by-two factorial manipulation).

3. Experiments

Three psychological experiments tested skilled readers’ knowledge and exploitation of the relationship between suffix spelling and lexical category. The three experiments comprised a lexical category judgement task (Experiment 1), a spelling task (Experiment 2), and a sentence reading task (Experiment 3). We further tested whether gradations in diagnosticity and specificity across different suffixes were mirrored in participants’ behaviour. We predicted that in tasks

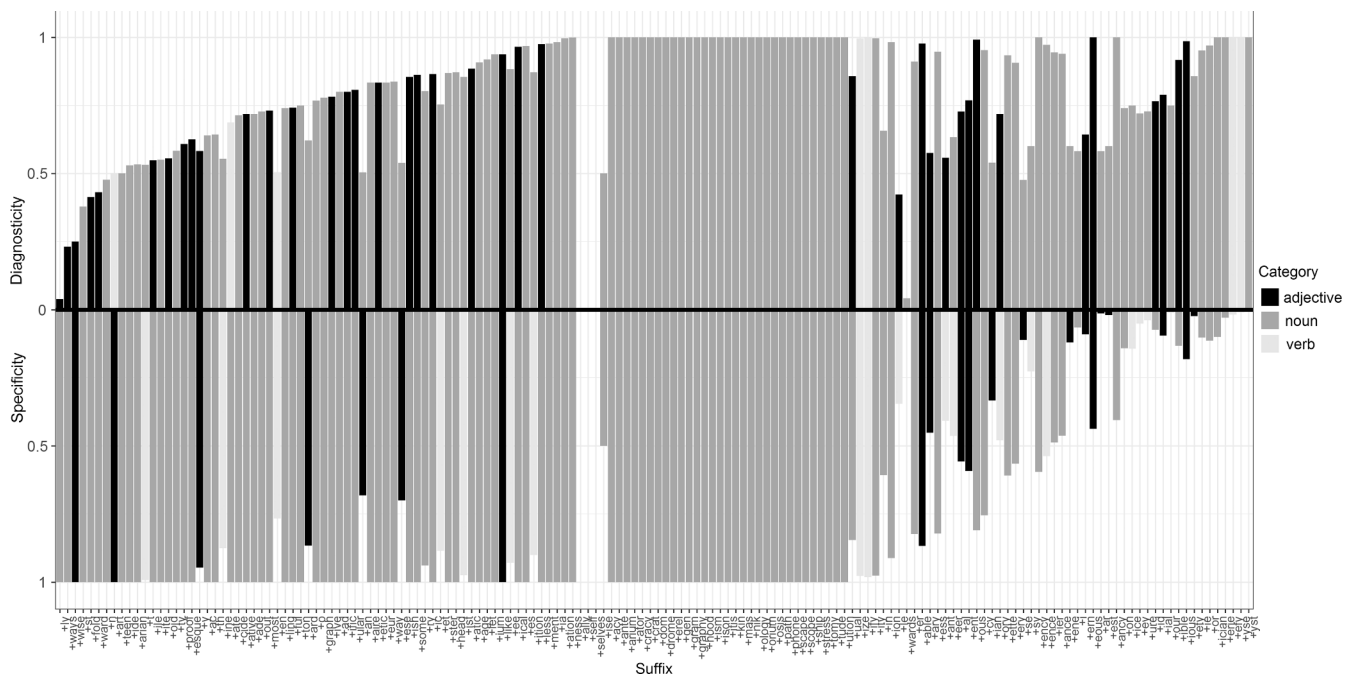


Fig. 2. Type diagnosticity and specificity of English derivational suffixes.

requiring access to grammatical information from printed spellings (lexical category judgment and sentence reading), spellings with higher diagnosticity should yield stronger behavioural effects compared to spellings with weaker diagnosticity. Similarly, we predicted that in tasks requiring production of spellings, stronger behavioural effects would be observed for spellings with higher specificity for predicted categories. Data for all experiments can be found in the OSF storage for this project (<https://osf.io/hac5j/>).

3.1. Experiment 1

Experiment 1 tested skilled readers' explicit knowledge of dependencies between spelling and lexical category. Using our linguistic analysis, we selected suffixes that are diagnostic for adjectives or nouns, and created nonwords ending in these suffixes. We then asked adult skilled readers to make decisions as to whether these nonwords appear more like adjectives or nouns. If people are sensitive to spellings' diagnostic properties, then their responses should be predictable from nonword spelling.

3.1.1. Method

3.1.1.1. Participants. Participants were 46 undergraduate students at Royal Holloway, University of London. They were native English speakers with no history of reading, spelling, or learning difficulties. This experiment was conducted at the end of a longer, unrelated testing session. Participants were given course credit or paid at the rate of £10/h.

3.1.1.2. Stimuli. Twenty suffixes that were diagnostic for the noun or adjective category were selected (see Table 2). These suffixes all had a high type diagnosticity, i.e. greater than or equal to 0.50, and occurred in at least 20 different words.

Eighty non-existing stems with a CVC structure were constructed using the Wuggy pseudoword generator (Keuleers & Brysbaert, 2010; <http://crr.ugent.be/programs-data/wuggy>). Each stem was used only once. Nonwords were orthographically legal and pronounceable, and we did not include any pseudohomophones (e.g. 'sabal' may be pronounced 'sable'). Further, none of the nonwords were close neighbours of existing words; on average, they had 0.41 orthographic neighbours

(with the median of 0, ranging between 0 and 4 orthographic neighbours, Coltheart, Davelaar, Jonasson, & Besner, 1977). See Appendix B for the complete list of nonword stimuli. Nonmorphological endings were not included in this experiment, because its goal was not to adjudicate between types of cues used in lexical category judgement (e.g. orthographic vs morphological).

To minimise potential priming effects, stimuli were divided into two experimental lists with 40 items per list. Participants saw only one list; each suffix occurred twice. Four practice trials were constructed using suffixes that were not used in the main experiment (–IE, –AL, –ULAR, –LING).

3.1.1.3. Procedure. Participants were seated in a quiet room and read the following instructions: “You will be asked to decide if a letter string looks like a noun or an adjective. NOUN is a person, animal, place, thing, or idea: for example, AUNT, CAT, FOREST, CUP, LOVE. ADJECTIVE is an attribute of a noun: for example, SWEET, RED, SIMPLE”. Following practice trials, nonwords were presented one by one in the centre of the screen in Courier New 12-point font, and participants indicated their adjective/noun decisions via a button press. Instructions did not emphasise speed, although participants needed to make their decisions within a two-second timeout period. Experimental items were scrambled and presented in a different random order for each participant. The experiment lasted less than five minutes. The DMDX software (Forster & Forster, 2003) was used for stimulus presentation and data recording.

3.1.2. Results

We used the lme4 package in R for generalised mixed modelling analysis (Bates, Maechler, Bolker, & Walker, 2014) on 1625 data points (11.7% of data points were missing due to timeouts or accidental button presses that did not correspond to one of the two designated response buttons). The analysis considered the impact of suffix condition (adjective or noun suffix) on the binomial response (adjective or noun response). We included two random intercepts, one for subjects and one for suffixes. This analysis was expressed in the model $g\text{mler}(\text{response} \sim \text{condition} + (1|\text{subject}) + (1|\text{suffix}), \text{data}, \text{family} = \text{"binomial"})$. Results showed a main effect of suffix condition such that adjective-diagnostic suffixes elicited more “adjective” responses

Table 2
Characteristics of suffixes used in Experiments 1 and 3.

Noun suffixes					Adjective suffixes				
Suffix	Type frequency in nouns	Token frequency in nouns	Type diagnosticity	Token diagnosticity	Suffix	Type frequency in adjectives	Token frequency in adjectives	Type diagnosticity	Token diagnosticity
–NESS	1241	19,676	1.00	0.98	–OUS	225	18,368	0.99	1.00
–ITY	506	52,672	1.00	1.00	–ABLE	293	18,643	0.98	0.93
–MENT	251	60,300	0.98	0.99	–LESS	192	6274	0.97	0.70
–ENCE	175	27,295	0.97	0.99	–ICAL	166	11,034	0.97	0.98
–LET	34	1138	0.92	0.81	–LIKE	30	272	0.94	0.25
–ER	1091	109,511	0.91	0.62	–ATIC	23	1153	0.88	0.94
–AGE	108	8959	0.91	0.86	–IC	377	22,046	0.86	0.80
–EE	45	4329	0.88	0.89	–ISH	117	10,739	0.85	0.71
–IST	340	13,368	0.85	0.66	–IVE	193	19,956	0.78	0.80
–AN	136	5455	0.50	0.23	–Y	430	32,375	0.58	0.14

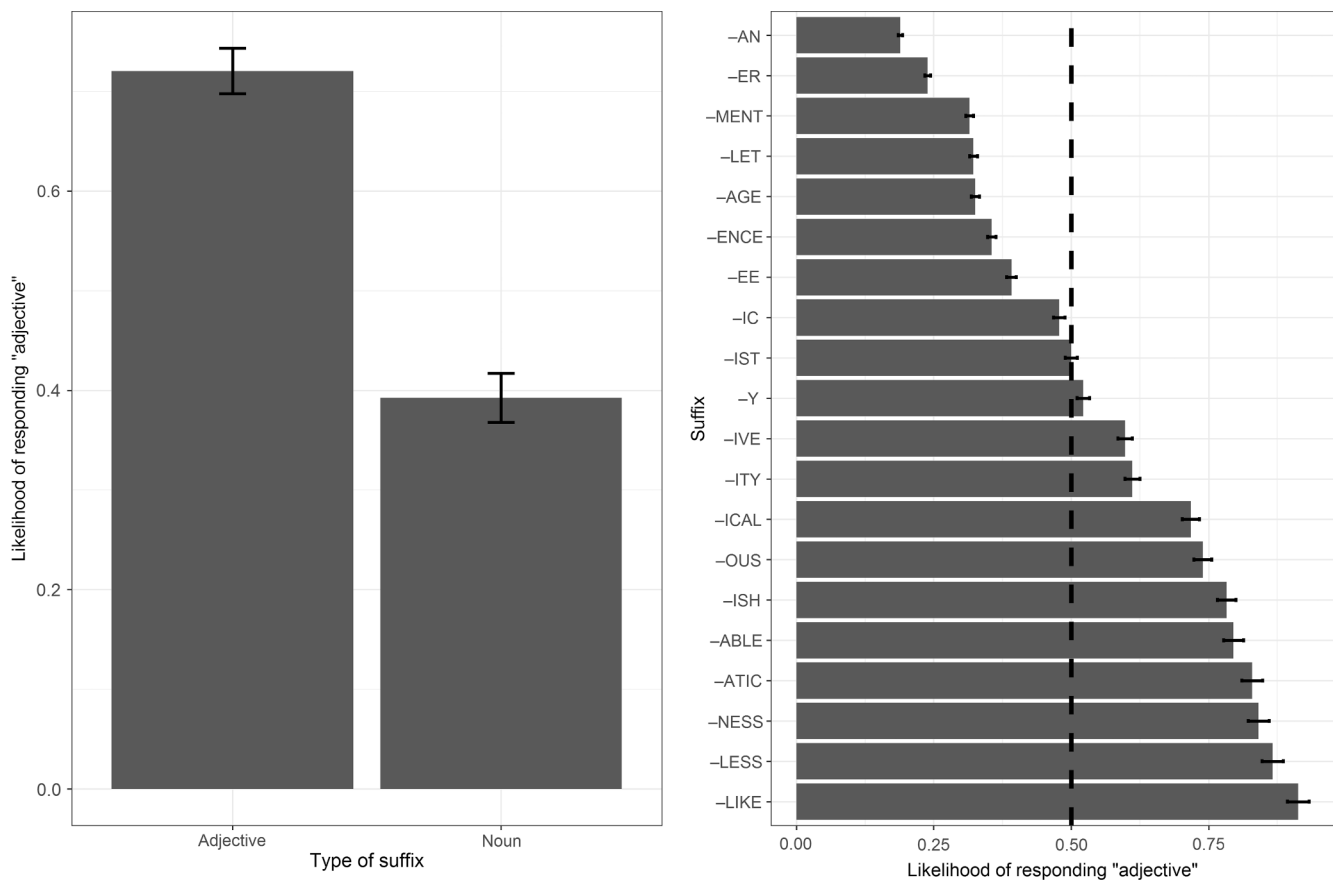


Fig. 3. Results of Experiment 1. The left panel plots the likelihood of the “adjective” response separately for adjective- and noun-diagnostic suffixes. The right panel plots the likelihood of the “adjective” response as a function of suffix.

($B = 1.57$, $z = 4.22$, $p < 0.0001$) than noun-diagnostic suffixes did. Fig. 3 visualises these data, averaged across the whole item set, and as a function of suffix.

Further analyses were conducted to investigate whether participants’ knowledge of the suffix spellings’ diagnosticity mirrors statistics of the writing system. That is, we asked whether participants were better at classifying nonwords that contain strongly diagnostic suffixes than those with weaker diagnosticity. In order to investigate this statistically, we multiplied our diagnosticity measure by -1 for noun-diagnostic suffixes and substituted this measure in our model, in place of the condition variable. This analysis revealed a main effect of diagnosticity on response type: $B = 0.87$, $z = 4.13$, $p < 0.0001$,⁶ such that

more diagnostic suffixes were more likely to be ascribed to the category that they denote. We performed this analysis separately for adjective- and noun- diagnostic suffixes to make sure that this effect was not due to the sign-transformation of the diagnosticity variable, and observed the effect of diagnosticity on both adjective-specific suffixes ($B = 4.56$, $z = 2.88$, $p < 0.01$, i.e. highest absolute values for diagnosticity yield more adjective responses) and noun-diagnostic suffixes ($B = -3.50$, $z = -2.05$, $p < 0.05$, i.e. highest absolute values for diagnosticity

(footnote continued)

in text. Token-based analyses will be reported in the footnotes. Main effect of token diagnosticity: $B = 0.85$, $z = 3.34$, $p < 0.001$; for adjective-diagnostic suffixes only: not significant, $p > 0.8$; for noun-diagnostic suffixes only: $B = -2.27$, $z = -2.20$, $p < 0.05$.

⁶ Analyses based on type specificity and diagnosticity measures are reported

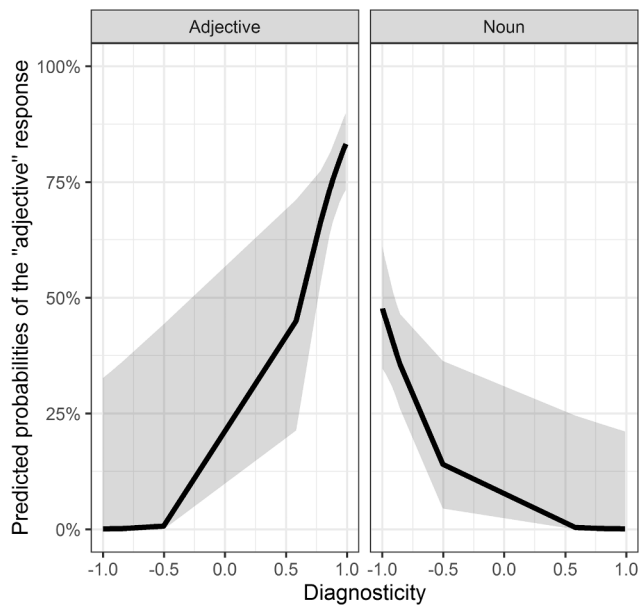


Fig. 4. Explaining variability across suffixes in Experiment 1. The probability of responding “adjective” to a nonword is high if this nonword’s suffix is strongly diagnostic for the “adjective” lexical category, and it is low if this nonword’s spelling is strongly diagnostic for a noun.

yield more noun responses; see Fig. 4). The results based on token measures were similar (main effect of diagnosticity, and an effect of diagnosticity on noun-diagnostic suffixes, see Footnote 6), but adjective-specific suffixes did not show an effect of diagnosticity.

3.1.3. Discussion

Results suggest that skilled readers extract information about lexical category from written language. When nonwords include suffixes that are diagnostic for adjectives, participants classify these as adjectives. Conversely, when nonwords include suffixes that are diagnostic for nouns, participants classify these as nouns. The strength of these effects mirrors the diagnosticity strength of English derivational suffixes. Token-based analyses showed a weaker effect of diagnosticity. We return to this result in the General Discussion.

3.2. Experiment 2

Experiment 2 was a spelling experiment designed to test adults’ knowledge of the extent to which particular spellings are used to denote lexical category for a given sound sequence. We created nonwords with endings that could be spelled in more than one way. We presented these nonwords auditorily in meaningful context conditions that biased participants to interpret them as nouns, adjectives, or verbs, and recorded participants’ spellings of the nonwords. If participants have stored knowledge of the specificity of spellings for lexical category, then context should influence the spellings that they produce, such that these are congruent with the predicted category.

3.2.1. Method

3.2.1.1. *Participants.* Participants were 29 undergraduate students from Royal Holloway, University of London. Sampling conditions were otherwise the same as in Experiment 1.

3.2.1.2. *Stimuli.* Eleven suffix spellings that had a high specificity for noun, adjective, or verb lexical category were selected. These were the *target spellings*. Critically, the most frequent pronunciation of these suffixes could be spelled in more than one way (see Table 3).

Sixty-six non-existing stems were selected from the same pool as in Experiment 1. Stems were joined with suffixes at random; one suffix

was joined with six different stems. None of the nonwords could be pronounced identically or similarly to existing words. The list of nonword transcriptions is available in Appendix C.

The nonwords were pronounced by an adult female native speaker of British English and recorded in stereo at a sampling rate of 22,050 Hz. Items were presented to her out of context in the IPA transcription format (see Appendix C), i.e. one transcription was used for one phonological sequence, regardless of the context the nonword would be assigned to. Recordings were subjected to the noise reduction procedure in Cool Edit Pro (2002, version 2.0, Syntrillium Software Corporation, Phoenix, AZ).

Sixty-six sentences were designed following one of three possible context templates (noun, adjective, or verb template; see Table 4 for an example of each template). Nonwords in the noun context template occupied the syntactic position of a direct/indirect object following an adjective, and so could only be interpreted as nouns. Nonwords in adjective contexts appeared after verbs ‘prove’, ‘stay’, ‘grow’, ‘smell’, ‘appear’, ‘taste’, ‘sound’, ‘seem’, ‘look’, ‘turn’, ‘become’, ‘remain’, that maximised the probability that they would be perceived as adjectives. Nonwords in verb contexts followed an adverb. For each suffix’s target spelling, two context types were selected – one that was congruent with the category that this spelling denoted and one that was incongruent with it. Naturally, the pairing of congruent and incongruent context types was different across suffixes (see Appendix C for the assignment of phonological nonwords to congruent/incongruent contexts). Note that no matching across conditions was required, because the same phonological ending was used in two different context conditions. Thus, any effect could not be attributed to the properties of a phonological sequence or its possible spellings. For instance, had it been the spelling frequency that influenced participants’ performance, we would observe no difference between the context conditions at all (i.e. the most frequent spelling would be used regardless of the predicted lexical category).

Each participant heard six different nonwords containing each suffix: three times in a context that was congruent with the target spelling of the suffix, and three times in a context that was incongruent with it. For instance, three different /-əs/-nonwords were embedded in sentences where they functioned as adjectives (congruent context where the target spelling –OUS is expected), and three different /-əs/-nonwords were embedded in sentences where they functioned as nouns (any other spelling is expected but not –OUS). See Appendix C for the complete list of stimulus materials.

Each subject saw 66 sentences. Two additional practice trials were constructed using nonwords and sentences that were not used in the main experiment.

3.2.1.3. *Procedure.* The experiment was programmed using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA, USA). Trials began with a 1000-ms blank screen. Participants then saw a printed sentence frame with a gap denoting placement of the target nonword. This sentence frame stayed on screen for 4 s, after which a nonword was played to participants through headphones. After the nonword finished playing, participants were immediately prompted to type their spelling. Time for responding was unrestricted, and participants were allowed to make corrections to their responses prior to moving to the next trial. Sentences were presented in a different random order for each participant. The duration of the experiment was under 10 min.

3.2.2. Results

Suffix spellings were manually extracted from whole nonword spellings (1914 analysable data points). The dependent variable was binary and coded whether the response was the *target spelling* for this suffix or not (see Table 3 for the list of target spellings associated with spoken endings). The analysis considered the impact of context condition (congruent or incongruent context) on the binomial response (target spelling or other spelling). The analyses were otherwise identical to those in Experiment 1. As predicted, results showed that the

Table 3
Suffixes used in Experiment 2.

Target spelling	Pronunciation	Type specificity	Token specificity	Possible alternative spellings	Congruent context	Incongruent context
–MENT	/mənt/	1.00	1.00	adamant	Noun	Adjective
–ER	/ət/	0.82	0.70	modular	Noun	Adjective
–LET	/lət/	1.00	1.00	moonlit	Noun	Adjective
–IST	/ɪst/	0.97	0.86	modest	Noun	Adjective
–NESS	/nɪs/	1.00	1.00	reminisce	Noun	Verb
–AL	/əl/	0.56	0.52	battle, tremble, marvel	Adjective	Verb
–ICAL	/ɪkəl/	1.00	1.00	cubicle, nickel, bicycle	Adjective	Verb
–LESS	/ləs/	1.00	1.00	malice, trellis, necklace, accomplice	Adjective	Noun
–OUS	/əs/	0.81	0.87	crocus, embarrass, atlas, genius	Adjective	Noun
–Y	/i/	0.95	0.98	calorie, devotee	Adjective	Noun
–ATE	/eɪt/	1.00	1.00	bait, portrait, great	Verb	Noun

Table 4

Examples of sentence contexts used in Experiments 2 and 3. The list of IPA transcriptions for the recordings that were played to participants in Experiment 2 can be found in [Appendix C](#). In Experiment 3, nonword spellings were presented in place of gaps; see [Appendix D](#).

Context type	Example
Noun	The candidate showed incredible < ____ > in a difficult situation.
Adjective	The unpleasant officer remained < ____ > throughout his entire service.
Verb	The worker was asked to carefully < ____ > the product for sale.

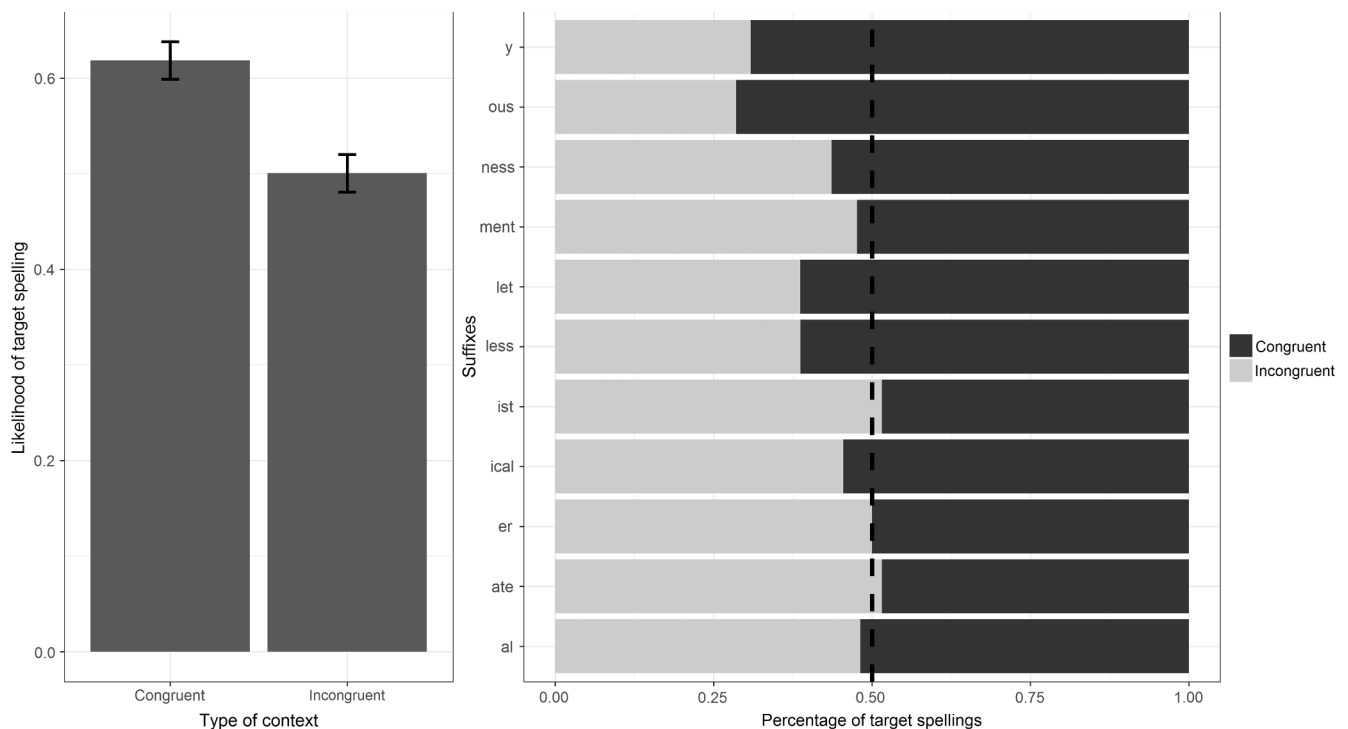


Fig. 5. Results of Experiment 2. The left panel plots the likelihood of the target spelling separately for congruent and incongruent sentence contexts. The right-hand panel plots the percentage of target spellings as a function of suffix and context.

target spelling was less frequent in the incongruent context condition ($B = -0.60$, $z = -5.79$, $p < 0.0001$) than in the congruent context condition. [Fig. 5](#) visualises these data.

Further analyses were conducted to investigate whether participants' knowledge of suffix specificity for lexical category mirrors statistics of the writing system. In other words, is the effect of condition (congruent vs incongruent) larger for suffixes whose spelling is highly specific for lexical category than for those with a weaker specificity? In order to investigate this hypothesis, we added the continuous specificity metric for the target spelling, as well as its interaction with the condition variable, to the linear model. Note that the values of specificity

for the same spelling are different for different context types (e.g. –AN is specific for adjectives and therefore congruent with adjectives, its specificity for adjectives is 0.76; but –AN cannot be used in verbs, its specificity for verbs is 0). As predicted, the interaction between specificity and context was significant, such that the target spelling was used increasingly more often in the congruent condition as its specificity increased ($B = 1.9$, $z = 2.8$, $p < 0.01^7$); there was no effect of

⁷ The corresponding values for the token analyses are $B = 2.59$, $z = 0.60$, $p < 0.01$.

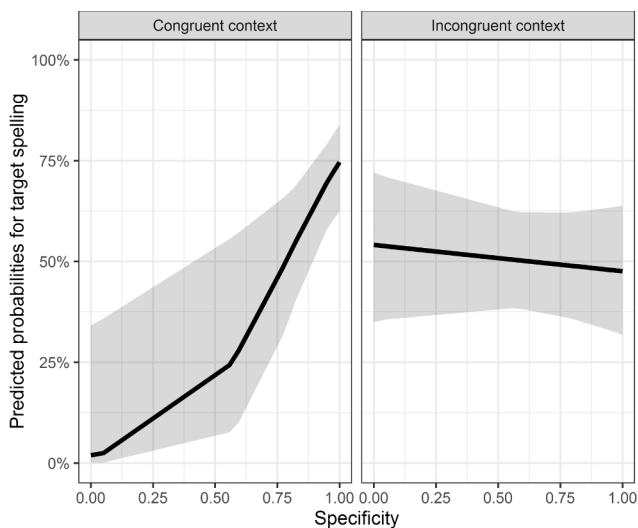


Fig. 6. The effect of specificity on the probability of the target spelling in Experiment 2. More target spellings were observed in congruent contexts for suffixes that are highly specific for their lexical categories than for those that are less specific for their lexical categories. When the context was incongruent, there was no effect of suffix specificity on the probability of the target spelling.

specificity in the incongruent condition. These results are visualised in Fig. 6.

3.2.3. Discussion

These results indicate that skilled readers possess knowledge that the different possible spellings for a sound sequence convey different information. When presented with a spoken word in a sentence context, they produced spellings that are appropriate for the predicted lexical category. Further, the strength of this effect mirrored the degree of specificity of suffix spellings for lexical category in the writing system. The higher the specificity of spellings, the more frequently they were produced by our participants in the congruent context condition. This held true for the congruent condition because the target spellings were strongly linked with their respective lexical categories. When there was no such link, as was the case for the incongruent condition, the proportion of target spellings was stable and was not influenced by target spellings' specificity.

3.3. Experiment 3

Experiments 1 and 2 assessed skilled readers' knowledge of the relationship between suffix spelling and lexical category using paradigms that required an explicit response. Experiment 3 investigates this knowledge in a more natural sentence reading task using eye-tracking methodology. We designed nonwords similar to those in previous experiments, and placed these in sentence contexts that were congruent or incongruent with the nonword suffix spellings. We anticipated that nonwords placed in incongruent contexts would yield processing difficulties, and that these would be reflected in eye-movement behaviour. We did not have any concrete predictions as to which eye-movement measures would be affected by the diagnosticity manipulation, so we chose to monitor a variety of typically studied measures that characterise skips, fixations, reading time, and regressions (see Rayner, Warren, Juhasz, & Liversedge, 2004; Warren, White, & Reichle, 2009). Some of these measures are interpreted as tapping into early processing (skipping rate, first fixation duration), while others index later processing stages involving contextual integration (regressions, go-past times; Rayner & Liversedge, 2011).

3.3.1. Method

3.3.1.1. Participants. 47 undergraduate students at Royal Holloway, University of London were recruited following the same procedure as in Experiments 1 and 2.

3.3.1.2. Stimuli. Eighty experimental nonwords with suffix spellings diagnostic for adjectives (10 suffixes) or nouns (10 suffixes) were selected from Experiment 1. Forty filler nonwords whose suffix spellings were diagnostic for verbs were added so that sentence contexts would be more varied. These filler nonwords were not analysed, because the majority of verb suffixes do not discriminate lexical category (i.e. type diagnosticity is less than 0.5). The ten verb suffixes that we used as fillers were –ATE, –ADE, –EER, –LE, –YSE, –IDE, –OUR, –ICE, –URE, –ART. Orthographic neighbourhood size for experimental nonwords ranged between 0 and 4 (mean was 0.45, median was 0).

One hundred and twenty sentences were designed following noun, adjective, and verb sentence templates used in Experiment 2. Some of these sentences were identical to sentences in that experiment. The assignment of nonwords to congruent/incongruent context types can be found in Appendix D. Sentences had an average of 10.5 words each (ranging between 8 and 15 words). For each sentence there were always at least four words in the sentence before the target, and at least three words post-target. Sentence parts preceding targets were matched across conditions on the number of letters (congruent: 30.48 letters on average, incongruent: 31.70 letters, $p > 0.3$). Note that the number of words pre-target was slightly different across conditions (congruent: 4.75, incongruent: 5.23, $t(64) = 2.07$, $p < 0.05$). Post-target contexts were matched on the number of letters (25.02 vs 26.33, respectively, $p > 0.05$) and the number of words (4.38 vs 4.56, respectively, $p > 0.05$). Pre-target words were always at least five characters long (with no difference across conditions: 7.33 vs 7.15, $p > 0.6$), to reduce the likelihood of these words being skipped.

Each of the 30 suffixes appeared with four different stems, twice in a congruent sentence context, and twice in an incongruent one. We used two experimental lists to counterbalance the assignment of nonwords to the congruent/incongruent conditions across participants (List 1 is presented in Appendix D). This counterbalancing procedure was implemented to ensure that any effects of the congruency variable were not the result of potential item-level differences across nonwords. Six practice sentences were designed following the same constraints as in the main experiment.

3.3.1.3. Procedure. Participants were seated at a viewing distance of 70 cm. This position was maintained with a table-mounted head and chin rest. The eye-tracking task was conducted on an LED monitor (dimensions 1920 × 1080, running at a 100-Hz refresh rate). The eye-tracker was an EyeLink 1000 (SR Research Ltd., Ottawa, Ontario, Canada) recording eye position (right eye) every 1 ms during sentence reading. A 9-point calibration and validation protocol was performed at the beginning of the session. A drift-correction point was displayed at the beginning of every trial, positioned at the location of the sentence onset. An additional check on recording accuracy was performed by displaying a gaze-contingent square of 2.5 characters' width (0.8°) in this same location, following completion of the drift correction. Participants were required to fixate this square for at least 40 ms to trigger the onset of the sentence. If a participant failed to do this, the trial was ended and discarded from further analysis, and the calibration procedure was repeated before the next trial was presented.

All sentences were displayed in black Courier font (size 14pt, horizontal character width 0.3°) on a light grey background, using SR Research ExperimentBuilder software (SR Research, Ottawa, Ontario, Canada). The experiment lasted 30–40 min. Participants could pause the experiment at any point.

Each participant read six practice sentences at the start of the experiment, and then proceeded immediately onto the 120 experimental

sentences. The order of presentation of trials was randomised for each participant. Participants were warned that each sentence would contain a nonword and were asked to read for comprehension. Single YES/NO comprehension questions were presented immediately after the relevant sentence on 50% of trials to ensure engagement with the task. One third of these questions referred to the pre-target context (e.g. ‘Is the sentence about a young man?’), one-third referred to the post-target context (‘Was the scholar studying something old?’), and one-third tested the retention of the nonwords (‘Was the solution cugity?’). When participants finished reading a sentence, they pressed a key to move on to the next trial.

3.3.2. Results

Eye-tracking data were processed using the SR Research DataViewer software (SR Research, Ottawa, Ontario, Canada). Fixations data were cleaned as follows. Short fixations of less than 80 ms that fell within 0.5° of another fixation were allocated to this longer fixation. Similarly, fixations of less than 40 ms were grouped with larger fixations within a 1.25° area. Fixations falling outside of interest areas that were drawn around each word in the sentence (with vertical limits of 5° above and below each word) were discarded, as were any remaining fixations shorter than 80 ms or longer than 1500 ms. All trials were also visually inspected in order to identify those where participants had not read the sentences in full, or where any considerable deviation in the eye-tracking trace was evident during the trial. As a result of these checks, 4% of trials were discarded. Average accuracy on comprehension trials was 0.86% (ranging from 0.66% to 0.98%).

The region of interest for analysis was the nonword. We considered word-skipping, first fixation durations, first run dwell time (sums of all fixation durations during first pass), go-past times, regressions to the target from later parts of the sentence, regressions from the target to earlier regions of the sentence. The meaning of each measure will be explained below. We implemented linear mixed models or generalised linear models as appropriate (the latter type of models were used where the dependent variable was binary, i.e. in the case of skips and regressions). All models included trial order as a main effect, and random intercepts for subjects and for suffixes. However, in the interest of space we report only those effects pertaining to our main variable of interest, i.e. congruency. The full report on all variables can be found in the OSF storage for this project.

The effects of congruency on all dependent variables studied are summarised in Table 5.

Word-skipping. The analysis considered the binary variable of whether the target nonword was skipped (i.e. no fixation occurred in first-pass reading). Overall, 9% of target nonwords were skipped (each participant made 7 target skips on average, ranging from 0 to 25 target skips; note that there were 80 potentially analysable trials per participant). There was no effect of congruency on skipping behaviour ($B = -0.11, z = -0.92, p > 0.1$).

Table 5

Results of Experiment 3. Means and standard deviations (SD) for eye-movement measures and critical values from their corresponding linear mixed effects models ($B, z/t$ values, significance levels).

Eye-tracking measures	Mean for congruent condition (SD)	Mean for incongruent condition (SD)	Congruency effect	Estimate (B)	z/t value
<i>Binary probability measures</i>					
Skips	0.09 (0.29)	0.08 (0.28)	-0.01	-0.11	$z = -0.92$
Regressions-in	0.33 (0.47)	0.35 (0.48)	0.02	0.12	$z = 0.14$
Regressions-out	0.13 (0.34)	0.17 (0.38)	0.04	0.30	$z = 3.12^{**}$
<i>Continuous measures (in ms)</i>					
First fixation duration	291 (139)	295 (148)	4.51	4.89	$t = 1.07$
First run dwell time	386 (239)	392 (250)	6.88	6.73	$t = 0.92$
Go-past time	477 (320)	519 (374)	42.00	41.22	$t = 3.84^{***}$

Note: Double asterisk (**) corresponds to $p < 0.001$, triple asterisk (***) corresponds to $p < 0.0001$.

First fixation duration. This analysis considered the impact of congruency on the duration of the first fixation event on the target (in ms). There was no effect of congruency on first fixation durations ($B = 4.89, t = 1.07, p > 0.1$).

First run dwell time. This analysis considered the impact of congruency on the sum of durations of all fixations on the target during the first pass (in ms). There was no effect of congruency on the first run dwell time ($B = 6.73, t = 0.92, p > 0.1$).

Go-past time. This analysis considered the impact of congruency on the summed fixation duration (in ms) from when the target was first fixated until the eyes move to the right, including any fixations made following regressions to earlier words in the sentence. There was an effect of congruency on go-past time ($B = 41.22, t = 3.84, p < 0.0001$) such that go-past times were longer for incongruent trials than for congruent ones.

Regressions-in. This analysis considered the impact of congruency on whether the target received at least one regression from later parts of the sentence. Thirty-four percent of all trials available for analysis contained such regressions. There was no effect of congruency on this measure ($B = -0.12, z = 0.14, p > 0.1$).

Regressions-out. This analysis considered the impact of congruency on whether regression(s) were made from the target to previous parts of the sentence prior to leaving in a forward direction. Fifteen percent of all trials available for analysis contained such regressions. The effect of congruency on regressions-out was significant in the expected direction, i.e. incongruent targets were more likely to elicit regressions to earlier regions of the sentence ($B = 0.30, z = 3.12, p < 0.001$).

In sum, we found significant effects of congruency on go-past times and regressions made from the target nonword back to the preceding part of the sentence. These two eye-movement measures in particular are associated with disambiguation processes resulting from difficulty in integrating an encountered word into the sentence context. We did not find any significant effect of congruency on any of the other measures reported, all of which are thought to reflect initial lexical processing (e.g. Rayner & Liversedge, 2011). These results therefore indicate that the category information becomes important following initial lexical processing, when the reader has enough contextual information to understand the sentence and detect any incongruities. To check if these effects could be explained by a difference in the number of words pre-target (more words pre-target in incongruent contexts than in congruent ones), we added the interaction of this number with the congruency variable to the corresponding mixed models. The addition of extra terms did not impact on the fit of the models (go-past times: $\chi(2) = 2.87, p = 0.24$; regressions-out: $\chi(2) = 0.23, p = 0.89$), and so we conclude that the observed congruency effects cannot be explained by the difference in the number of words pre-target across the congruency conditions.

Further analyses investigated whether the size of the congruency effects on the go-past and regressions-out measures was influenced by the degree to which suffix spellings were diagnostic of lexical category. In order to assess this, we tested whether the addition of the interaction between diagnosticity and congruency would improve the fit of the models in these cases. The fit of the go-past model improved significantly when the interaction and the main effect of diagnosticity were added: $\chi(2) = 8.58, p < 0.05$.⁸ The interaction between diagnosticity and condition was significant ($B = 28.24, t = 2.60, p < 0.01$), reflecting a larger effect of diagnosticity in the incongruent condition than in the congruent condition. Note, however, that the diagnosticity effect was not statistically significant when the data were analysed separately within conditions (congruent condition: $B = 8.30, t = 0.63, p = 0.53$; incongruent condition: $B = 36.35, t = 1.58, p = 0.13$). These results are visualised in Fig. 7.

Similarly, the fit of the regressions-out model improved significantly when the interaction between diagnosticity and condition was added: $\chi(2) = 14.37, p < 0.001$.⁹ The original model produced a convergence warning when the variables were added. One way to assess if this warning is a false positive is to refit this model using a different optimisation procedure, such as “bobyqa” (Powell, 2009). We did this, and observed that the fits of these two models and the significance levels for the fixed effects were identical, indicating that the initial warning was indeed a false positive (Bates, Maechler, Bolker, & Walker, 2015). We report the results from the refitted model below.

The effects of diagnosticity on regressions-out were as follows. First, there was a main effect of diagnosticity: the probability of regressing to previous regions of the sentence decreased as diagnosticity increased in the congruent condition ($B = -0.15, z = -2.31, p < 0.05$), and increased as diagnosticity increased in the incongruent condition (marginal significance, $B = 0.21, z = 1.76, p = 0.08$). Second, there was an interaction between diagnosticity and congruency such that the probability of regression became higher as diagnosticity increased in the incongruent condition ($B = 0.29, z = 3.01, p < 0.001$), whereas it became lower as diagnosticity increased in the congruent condition. These results are visualised in Fig. 8.

3.3.3. Discussion

Experiment 3 demonstrated that adults are slower to progress past a target nonword whose spelling is incongruent with the predicted lexical category. This is likely due to the fact that such situations cause participants to regress to earlier regions of the sentence. Both of these effects suggest that participants experience difficulties in integrating such nonwords into incongruent sentence contexts (Rayner et al., 2004; Warren et al., 2009). Further, the strength of the relationship between suffix spelling and lexical category modulates this integration difficulty. Reading is facilitated when suffix spellings that are strongly diagnostic of a particular lexical category are used in contexts that are congruent with that category. Conversely, reading is more effortful when such spellings are used in contexts that are incongruent with that category.

4. General discussion

English has an alphabetic writing system in which letters map relatively consistently to sounds. However, it is well known that additional regularities between spelling and meaning arise as a result of morphological relationships between words. Stems occur repeatedly in

⁸ The token diagnosticity measure did not explain additional variance in the data ($p > 0.3$).

⁹ Analogous results were observed when the token measure of diagnosticity was used: $\chi(2) = 6.942, p < 0.01$. This effect was in the same direction as in the type analysis, but weaker. Note that the effects were not significant when the conditions were analysed separately (congruent: $B = -0.09, z = -1.25, p = 0.21$; incongruent: $B = 0.14, z = 1.19, p = 0.24$).

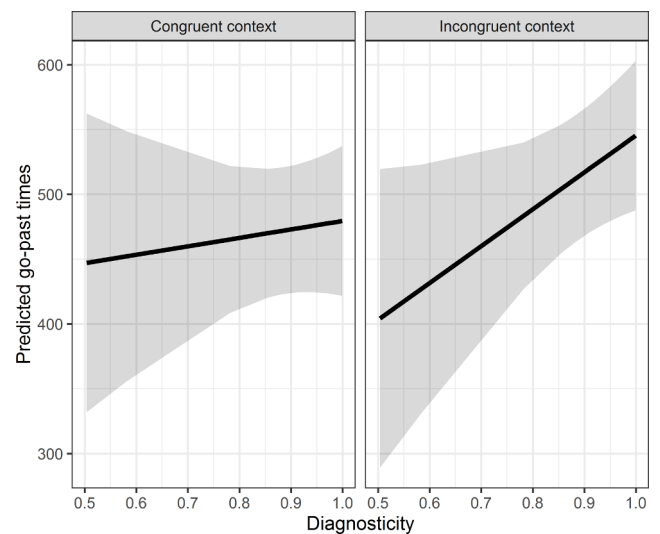


Fig. 7. Diagnosticity had a significant impact on the effect of congruency on go-past times in Experiment 3. The right panel corresponds to the incongruent condition: participants took longer to progress past the nonword if its spelling strongly conflicted with the sentence context. This effect was smaller in the congruent condition (left panel).

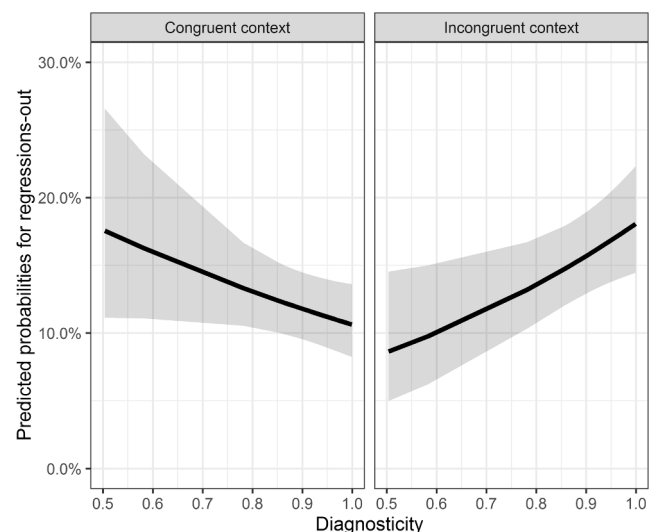


Fig. 8. Diagnosticity modulates the effect of congruency on regressions-out (regressions from the target nonword to previous regions of the sentence) in Experiment 3. The probability of regressing back to the beginning of the sentence decreases for suffixes that are highly diagnostic for lexical category cued by the sentence. Conversely, this probability increases when the suffix spelling that is specific for one category comes into conflict with a different syntactic role that is ascribed to it by the sentence context.

words of similar meanings (e.g. ‘builder’, ‘building’, ‘rebuild’) and affixes alter the meanings of words in predictable ways (e.g. ‘teacher’, ‘builder’, ‘banker’; Rastle et al., 2000). A wide variety of research suggests that skilled readers acquire knowledge of these regularities, and use this knowledge in the recognition of words (e.g. Rastle & Davis, 2008; Taft & Forster, 1975).

However, linguistic analyses over the past 50 years provide hints that the impact of these regularities in the writing system may not have been fully recognized by experimental psychologists studying visual word processing (e.g. Carney, 1994; Venezky, 1972). Indeed, it has long been known that sometimes morphological regularities take priority over spelling-sound regularities. For example, stems are preserved across derivations (e.g. ‘magic’, ‘magical’, ‘magician’), even though this

preservation can create spellings that do not provide a good reflection of their spoken forms (e.g. Treiman & Bourassa, 2000). Similarly, the spellings of meaningful affixes can be preserved in cases where the realisation of the spoken forms differs (as in the case of the past tense spelling –ED; Carney, 1994). Recently, Berg and Aronoff (2017; also Berg et al., 2014) have gone further than this, and suggested that some suffix spellings may actually be reserved for communicating particular grammatical functions. For example, the spelling –ED is strongly indicative of the past, and is very rarely used to spell ostensibly complex words comprising a single morpheme.

The purpose of our computational analysis was to quantify these relationships between spelling and lexical category, and to determine whether these regularities describe English derivational suffixes in general. We introduced two new concepts: diagnosticity and specificity. Diagnosticity refers to the extent to which a particular spelling predicts lexical category, while specificity refers to the extent to which a particular spelling is the preferred means of denoting a particular lexical category for a given sound sequence. Though diagnosticity refers to a property of English suffixes that linguists have long been familiar with (e.g. Carney, 1994), specificity refers to a property that has been virtually unstudied (but see Berg & Aronoff, 2017). Further, these regularities have never been quantified on a corpus scale. Our analyses demonstrate definitively that the 154 English suffixes studied are characterised by a high degree of diagnosticity and specificity, indicating that these forms of regularity are ubiquitous in the present-day English derivational system. These data confirm the notion that spellings of derivational English suffixes tend to be reserved for particular meaningful functions.

The acquisition of skilled reading involves the accumulation of years of experience with the writing system. In light of the large proportion of the English lexicon characterised by derivational suffixation, we hypothesized that these regularities might be encoded in the long-term knowledge of skilled readers. Results of our investigations that tested this hypothesis were unambiguous. Across three experiments using different paradigms, we found that these regularities impacted on the reading, spelling, and decision-making behaviour of skilled readers. These results suggest that skilled readers learn and subsequently exploit morphological cues to lexical category even without instruction to do so. Participants showed clear sensitivity to the relationships between suffix spellings and lexical category information. Moreover, the strength of these behavioural effects mirrored the degree of diagnosticity and specificity of the suffixes used. These results suggest that skilled readers' long-term knowledge represents the statistical structure of the writing system. It is unlikely that our participants were ever taught these relationships explicitly; more likely is that this knowledge was acquired through implicit statistical learning processes (see also St. Clair, Monaghan & Christiansen, 2010, for evidence on statistical learning of lexical categories).

We have characterised the English writing system in a manner that departs from the typical focus on spelling-sound relations (e.g. Ziegler, Stone, & Jacobs, 1997). English is typically characterised as a deep orthography as a result of its high degree of inconsistency across this mapping (Katz & Frost, 1992), making it difficult to learn to read (Seymour, Aro, & Erskine, 2003). However, the results of our analyses suggest that the proliferation of spellings for particular sound sequences in English may actually be functional. This possibility is intriguing for two reasons. First, it suggests that irregularity in the English spelling-sound mapping may permit regularity in the spelling-meaning mapping. This fractionation could not occur in a spelling-sound transparent writing system. Second, it suggests that the written signal conveys more disambiguating information than the spoken signal in these cases. These hypotheses are supported by diachronic data on the evolution of suffix spellings from Old English to modern days provided by Berg and Aronoff (2017). For instance, before the 16th Century adjectives were commonly spelled with word-final –OUSE, –US, or –OWS (e.g.

'glorius'), but these spellings were displaced by –OUS over time. Berg and Aronoff's analyses indicate that spellings have evolved in such a way as to express meaningful information more strongly. It is our view that diagnosticity and specificity are characteristic not of the English writing system exclusively but rather reflect general cognitive principles. Thus, these principles likely manifest, in one way or another, in other orthographies. Further research is needed, however, to test this conjecture. We chose English as a starting point for this work, because its spelling evolved under minimal environmental constraints (unlike French and Italian whose development was regulated by language academies, see Berg & Aronoff, 2017).

In order to quantify the extent to which English suffix spellings convey more information about lexical category than English sound sequences, we analysed each suffix in terms of the entropy of its spelling and phonological realisation. Entropy (H) is an information-theoretic measure of uncertainty (Shannon, 1948), which has previously been used to quantify the relationship between spelling and sound (e.g. Mousikou, Sadat, Lucas, & Rastle, 2017). We can calculate the entropy of each suffix spelling/spoken realisation using the formula $\Sigma[-p_i \times \log_2(p_i)]$, where p_i is the proportion of words belonging to a given lexical category. H values of 0 denote that the ending occurs in words of only one category (i.e. the category can be predicted with absolute certainty), whereas high H values indicate that the ending occurs in words of more than one category relatively often (i.e. the category can be predicted with less precision).

Using this approach, we calculated H_{spelling} to provide an estimate of the prediction precision from the suffix spelling, and we calculated H_{sound} to provide an estimate of the prediction precision from the most frequent pronunciation of the suffix. We then examined the slope of the regression equation $\ln(H_{\text{spelling}} \sim H_{\text{phonology}}$, weights = suffix frequency). There are three possible outcomes from this analysis. A slope of 1 indicates that suffix pronunciations and their spellings provide the same degree of information about lexical category; a slope greater than 1 indicates that phonology is more predictive of category compared to spelling; and a slope less than 1 would support our conjecture that spelling is more predictive of category compared to phonology. We found that the regression line has a slope of 0.9 ($SE = 0.04$, $t = 25.32$, $p < 0.0001$), and that this slope is significantly different from 1 ($F(1) = 7.44$, $p < 0.01$). That is, the spellings of derivational suffixes generally yield better prediction of lexical category (mean H is 0.61) than do their spoken realisations (mean H is 0.63).

4.1. Why are some suffixes more diagnostic than others?

We ran preliminary analyses to understand why some morphemes are diagnostic, while others are not. For this analysis, we removed seven suffixes that are indicative of lexical categories that were not considered in this paper (e.g. –WISE, –TEEN). Next, we tested if diagnosticity is a property of more frequent suffixes by correlating type frequency and type diagnosticity. Surprisingly, the correlation was negative, i.e. -0.24 , $p < 0.01$, but this was entirely due to suffixes with diagnosticity of 1. There are 40 such suffixes, and they can be highly frequent (e.g. –ISM), as well as very infrequent (e.g. –NIK, –DROME). In fact, all extremely infrequent suffixes are diagnostic; that is, they appear in a very small number of words of one lexical category, and do not tolerate any exceptions. This might mean that infrequent suffixes are not retained in written language unless they are diagnostic. Once we remove these 40 suffixes, a positive correlation is observed (0.23 , $p < 0.05$), meaning that highly diagnostic suffixes tend to be those that occur in many words. There are exceptions to this (frequent but ambiguous suffixes like –AL, –Y). In line with our behavioural data, we find no relationship between token frequency and diagnosticity.

Similar dependency is characteristic of suffix specificity. As with the diagnosticity analyses, we excluded 20 suffixes that are associated with categories that we did not consider in this paper (e.g. –ALLY,

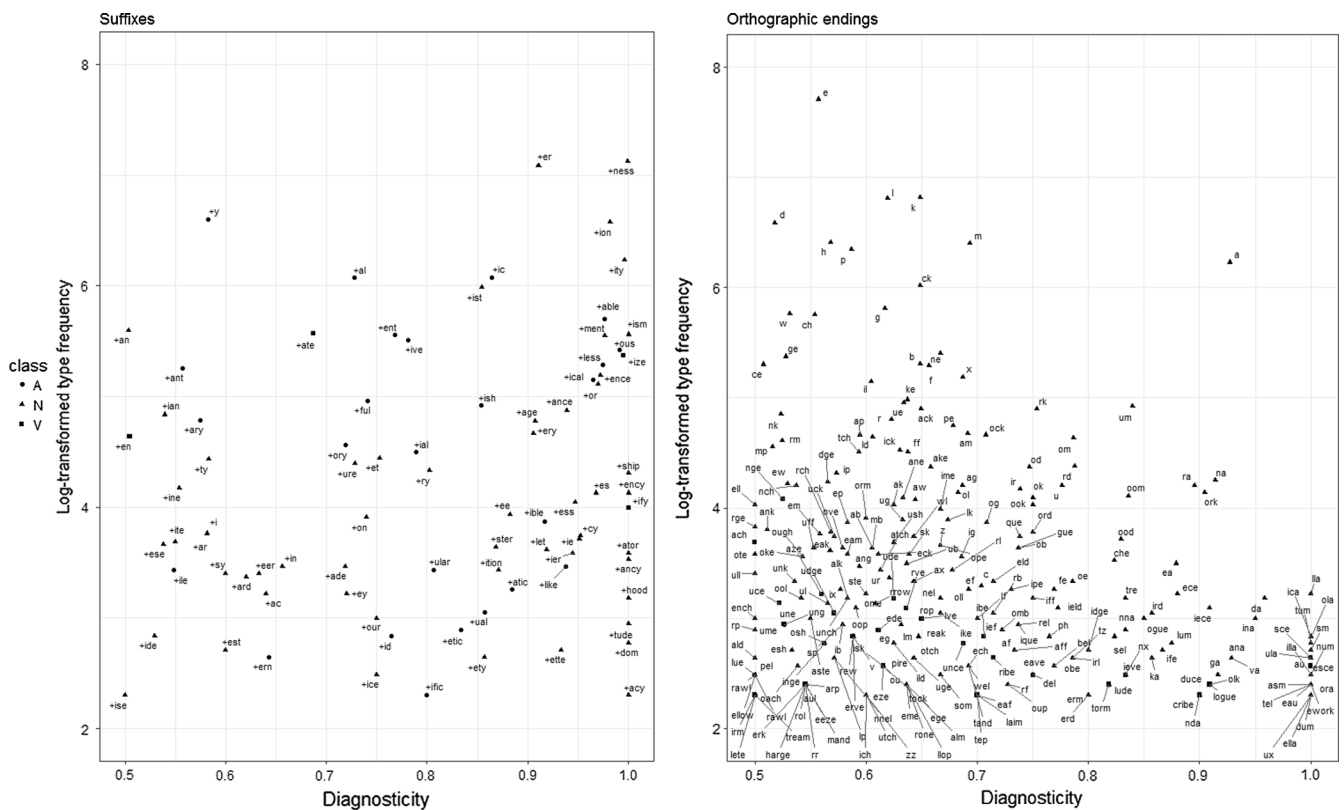


Fig. 9. Diagnosticity and frequency of suffixes and non-morphemic endings in English. A – adjective, N – noun, V – verb.

–WARDS).¹⁰ We also removed all suffixes with the specificity of 1. The analysis of the remaining 36 suffixes revealed a positive relationship between frequency and specificity (correlation 0.53, $p < 0.001$). That is, suffixes with a higher frequency tended to be more specific (e.g. –Y, –ER), and suffixes with a lower frequency tended to be less specific (e.g. –ENE, –ISE).

A noteworthy aspect of the diagnosticity and specificity data is that that token-based measures provided a poorer fit to our behavioural data compared to the type-based measures, in all experiments. Further, token frequency had a weaker correlation with diagnosticity and specificity in the corpus data compared to type frequency, as we have noted above. This finding is in line with studies showing facilitatory effects of contextual diversity (the number of contexts in which a unit or word appear) on the recognition of printed words (Adelman, Brown, & Quesada, 2006), and specifically, in the context of morphology, the effect of morphological family size on online processing (how many different words embed a morpheme; Amenta, Marelli, & Crepaldi, 2015; De Jong, Schreuder & Baayen, 2000; Ford, Davis, & Marslen-Wilson, 2010) and learning (Tamminen et al., 2015).

4.2. Are the effects morphological or orthographic in nature?

Might the relationship that we have observed between the suffix form and its meaning be part of a more general characterisation of English orthography? Several previous studies focussed on orthographic (nonmorphological) cues to lexical categories. For example, Kemp et al. (2009) found that people use orthography to make judgements about lexical categories of novel words and to convey lexical category information (e.g. –IFF is a frequent noun ending, while –ERGE

is a frequent verb ending). The authors used 10 noun-related and 10 verb-related nonmorphological endings, and found that these influence people's behaviour in sentence construction, sentence judgement, and pseudoword judgement tasks. Arciuli and Cupples (2006) also studied orthographic correlates of noun and verb categories. They reported a sentence construction experiment where participants created sentences with given nonwords ('fantern', 'setect', 'feduct'), thus assigning them a lexical category.

Our view is that there are fundamental differences between low-level orthographic and morphological cues, in terms of the degree of precision with which the latter are encoded in written language, and also in terms of their distributional properties. To illustrate some distributional differences, suffixes are a closed class of units, whereas the number of orthographic units is larger and depends on the definition that one adopts. For example, do orthographic units include written word bodies (e.g. –EAL; Arciuli & Cupples, 2006), or one-letter units (e.g. Onnis & Christiansen, 2008; see discussion in Ktori, Mousikou, & Rastle, 2018). Fig. 9 plots the distributions of suffixes and orthographic endings (all one, two, three, four, and five letter endings that are not part of existing suffixes). Not only there are fewer suffixes than other endings, but the former are also higher in diagnosticity and frequency (most diagnostic suffixes are highly frequent, whereas other highly diagnostic endings are rare). The majority of non-suffix endings are indicative of the noun category (88%), with the remaining 11% pointing to verbs. Only two endings (–UNG and –LETE) are strongly associated with the adjective category. On the other hand, suffixes permit discrimination between nouns (65%) and adjectives (30%). Finally, suffixes' combinatorial nature permits a substantial degree of productivity that is absent for non-morphemic sequences.

In summary, we have provided a new way to think about English writing, in which one consequence of the apparent disorder of spelling-sound information is order in the spelling-meaning mapping. We have shown that literate adults are sensitive to these regularities between spelling and meaning, and that their long-term knowledge mirrors the

¹⁰ In the token analyses, 29 suffixes that were specific for a category other than noun, adjective, or verb, were removed. Further, suffixes with the specificity of 1 were removed leaving 26 suffixes for the token-based analysis. There was no relationship between token frequency and token specificity.

strength of these regularities in the writing system. These conclusions suggest that reading the poem *Jabberwocky* provides information about meaning that is not present when listening to the poem. Lewis Carroll's choice of the spelling 'vorpal' to describe the blade (as opposed to 'vorple') was unlikely to be random, even if he didn't know that to be the case. Further, our research suggests that readers take advantage of the relationship between spelling –AL and adjective status to assist their interpretation of this part of the poem. One interesting pathway for future research will be to chart how these regularities emerged through spelling changes in the development of English writing.

Acknowledgements

This work was funded by the ESRC Future Research Leader

Appendix A

Diagnosticity, specificity and entropy values for 154 suffixes. A – adjective, N – noun, V – verb. NA – specificity value is not defined because the spelling never occurs as a potential suffix with words of a given category. H – entropy.

Suffixes	Diagnosticity _A	Diagnosticity _N	Diagnosticity _V	Specificity _A	Specificity _N	Specificity _V	Type frequency	H phonology	H orthography
–ABLE	0.977	0.023	0.000	0.867	0.625	NA	300	0.176	0.160
–AC	0.320	0.640	0.040	1.000	1.000	1.000	25	1.121	1.124
–ACY	0.000	1.000	0.000	NA	1.000	NA	10	0.000	0.000
–AD	0.000	0.800	0.000	NA	1.000	NA	5	0.811	0.722
–ADE	0.000	0.719	0.281	NA	1.000	1.000	32	0.869	0.857
–AGE	0.025	0.908	0.067	1.000	1.000	1.000	119	0.588	0.523
–AIRE	0.167	0.833	0.000	1.000	1.000	NA	6	1.096	0.650
–AL	0.727	0.252	0.014	0.557	0.392	0.051	433	1.413	0.976
–ALLY	0.000	0.000	0.000	NA	NA	NA	397	0.000	0.000
–AN	0.496	0.504	0.000	0.758	0.681	0.000	270	1.382	1.000
–ANCE	0.008	0.939	0.038	NA	0.462	0.333	131	0.165	0.411
–ANCY	0.000	1.000	0.000	NA	0.405	NA	34	0.000	0.000
–ANT	0.557	0.427	0.016	0.408	0.644	0.667	192	0.971	1.088
–ANTE	0.000	1.000	0.000	NA	1.000	NA	1	1.500	0.000
–AR	0.349	0.581	0.070	0.351	0.014	0.033	43	0.603	1.253
–ARD	0.069	0.621	0.241	1.000	0.867	1.000	29	0.993	1.523
–ARIAN	0.467	0.533	0.000	1.000	1.000	NA	30	0.997	0.997
–ARIUM	0.000	1.000	0.000	NA	1.000	NA	4	0.000	0.000
–ART	0.250	0.000	0.500	1.000	NA	1.000	4	1.918	1.500
–ARY	0.575	0.425	0.000	0.451	0.274	0.000	120	1.018	0.984
–ATE	0.126	0.187	0.687	1.000	1.000	1.000	262	0.643	1.201
–ATIC	0.885	0.115	0.000	1.000	1.000	NA	26	0.592	0.516
–ATION	0.000	0.996	0.002	NA	1.000	1.000	507	0.057	0.041
–ATIVE	0.718	0.282	0.000	1.000	1.000	NA	71	0.858	0.858
–ATOR	0.000	1.000	0.000	NA	1.000	NA	36	0.000	0.000
–CIDE	0.000	0.714	0.286	NA	1.000	1.000	7	0.863	0.863
–CRACY	0.000	1.000	0.000	NA	1.000	NA	7	0.000	0.000
–CRAT	0.000	1.000	0.000	NA	1.000	NA	7	0.000	0.000
–CY	0.048	0.952	0.000	0.222	0.755	0.000	42	0.830	0.276
–DOM	0.000	1.000	0.000	NA	1.000	NA	16	0.000	0.000
–DROME	0.000	1.000	0.000	NA	1.000	NA	2	0.000	0.000
–EE	0.020	0.882	0.059	0.500	0.929	1.000	51	0.771	0.733
–EER	0.033	0.633	0.333	0.250	0.463	0.909	30	1.338	1.109
–EFY	0.000	0.000	1.000	NA	NA	0.018	1	0.000	0.000
–EN	0.272	0.194	0.505	0.412	0.016	0.766	103	0.650	1.663
–ENCE	0.000	0.972	0.028	NA	0.538	0.667	180	0.165	0.183
–ENCY	0.000	1.000	0.000	NA	0.595	NA	62	0.000	0.000
–ENE	0.400	0.600	0.000	0.400	0.120	0.000	5	0.976	0.971
–ENT	0.768	0.208	0.019	0.592	0.356	0.333	259	0.971	0.905
–EOUS	1.000	0.000	0.000	0.438	0.000	NA	14	0.310	0.000
–ER	0.018	0.911	0.059	0.595	0.823	0.750	1198	0.603	0.568
–EREL	0.000	1.000	0.000	NA	1.000	NA	3	0.000	0.000
–ERIE	0.000	1.000	0.000	0.000	0.030	0.000	7	1.018	0.000
–ERN	0.643	0.286	0.071	0.091	0.035	0.000	14	1.382	1.198
–ERY	0.094	0.906	0.000	0.069	0.565	0.000	106	1.018	0.451
–ES	0.016	0.968	0.000	NA	1.000	NA	62	0.000	0.238
–ESE	0.462	0.538	0.000	1.000	0.700	NA	39	0.954	0.996
–ESQUE	0.625	0.375	0.000	1.000	1.000	NA	8	0.954	0.954
–ESS	0.000	0.947	0.053	NA	0.821	0.000	57	0.345	0.297
–EST	0.133	0.600	0.200	0.034	0.020	0.000	15	0.667	1.555
–ET	0.024	0.753	0.224	0.143	0.885	1.000	85	1.136	0.919
–ETIC	0.833	0.167	0.000	1.000	1.000	NA	18	0.523	0.650
–ETTE	0.000	0.933	0.067	NA	0.609	0.167	15	0.722	0.353

-ETY	0.143	0.857	0.000	0.500	0.024	NA	14	0.037	0.592
-EUR	0.000	0.833	0.167	NA	1.000	0.000	6	0.863	0.650
-EY	0.240	0.720	0.040	0.013	0.051	0.067	25	1.269	1.021
-FOLD	0.414	0.138	0.103	1.000	1.000	1.000	29	1.789	1.789
-FUL	0.741	0.252	0.000	1.000	1.000	NA	143	0.882	0.871
-GEN	0.000	1.000	0.000	NA	1.000	NA	2	0.000	0.000
-GRAM	0.000	1.000	0.000	NA	1.000	NA	4	0.000	0.000
-GRAPH	0.000	0.778	0.222	NA	1.000	1.000	9	0.764	0.764
-GRAPHY	0.000	1.000	0.000	NA	1.000	NA	5	0.000	0.000
-HEAD	0.032	0.871	0.065	1.000	1.000	1.000	31	0.715	0.748
-HOOD	0.000	1.000	0.000	NA	1.000	NA	24	0.000	0.000
-I	0.419	0.581	0.000	0.037	0.065	0.000	43	1.269	0.981
-IA	0.018	0.982	0.000	NA	1.000	NA	56	0.000	0.129
-IAL	0.789	0.211	0.000	0.095	0.042	0.000	90	1.413	0.744
-IAN	0.460	0.540	0.000	0.369	0.333	NA	126	1.000	0.995
-IBLE	0.917	0.083	0.000	0.133	0.375	NA	48	0.176	0.414
-IC	0.865	0.133	0.002	1.000	1.000	1.000	436	0.637	0.589
-ICAL	0.965	0.035	0.000	1.000	1.000	NA	172	0.214	0.218
-ICE	0.000	0.750	0.250	NA	0.143	0.500	12	0.345	0.811
-ICIAN	0.000	1.000	0.000	NA	0.100	0.000	3	0.513	0.000
-ID	0.765	0.235	0.000	0.073	0.667	NA	17	0.308	0.787
-IDE	0.000	0.529	0.412	0.000	1.000	1.000	17	1.757	1.253
-IE	0.024	0.951	0.000	0.002	0.102	0.000	41	1.269	0.330
-IER	0.028	0.944	0.028	0.250	0.488	0.091	36	1.338	0.365
-IFIC	0.800	0.200	0.000	1.000	1.000	NA	10	0.722	0.722
-IFY	0.000	0.000	1.000	NA	NA	0.982	54	0.000	0.000
-ILE	0.548	0.323	0.129	1.000	1.000	1.000	31	1.344	1.383
-IN	0.031	0.656	0.094	0.250	0.607	0.500	32	1.620	1.656
-INE	0.323	0.554	0.123	0.944	0.875	1.000	65	1.459	1.371
-ION	0.001	0.982	0.011	0.000	0.912	0.128	717	0.650	0.158
-IOUS	0.985	0.015	0.000	0.181	0.000	NA	68	0.040	0.111
-ISE	0.000	0.500	0.500	NA	0.500	0.014	10	0.176	1.000
-ISH	0.854	0.080	0.058	1.000	1.000	1.000	137	0.878	0.778
-ISM	0.000	1.000	0.000	NA	1.000	NA	260	0.000	0.000
-ISON	0.000	1.000	0.000	NA	1.000	NA	1	0.000	0.000
-IST	0.141	0.854	0.005	0.966	0.974	1.000	398	0.667	0.631
-ITE	0.325	0.550	0.075	1.000	1.000	1.000	40	1.224	1.548
-ITION	0.000	0.871	0.129	NA	0.900	1.000	31	0.513	0.555
-ITIS	0.000	1.000	0.000	NA	1.000	NA	16	0.000	0.000
-ITY	0.004	0.996	0.000	0.500	0.976	NA	508	0.037	0.037
-IUM	0.063	0.938	0.000	NA	1.000	NA	16	0.000	0.337
-IVE	0.781	0.206	0.012	1.000	1.000	1.000	247	0.804	0.825
-IZE	0.000	0.005	0.995	NA	0.250	0.977	215	0.176	0.043
-KIN	0.000	1.000	0.000	NA	1.000	NA	6	0.000	0.000
-LE	0.423	0.341	0.228	0.345	0.566	0.949	355	1.413	1.607
-LESS	0.975	0.010	0.000	1.000	1.000	NA	197	0.156	0.209
-LET	0.027	0.919	0.054	NA	1.000	1.000	37	0.179	0.480
-LIKE	0.938	0.000	0.000	1.000	NA	NA	32	0.382	0.400
-LING	0.240	0.740	0.000	1.000	1.000	NA	50	0.752	0.928
-LY	0.039	0.006	0.002	1.000	1.000	NA	2717	0.297	0.323
-MAS	0.000	1.000	0.000	NA	1.000	NA	3	0.000	0.000
-MENT	0.000	0.977	0.023	NA	1.000	NA	257	0.000	0.160
-MOST	0.731	0.077	0.000	1.000	1.000	NA	26	1.049	1.073
-N	0.182	0.477	0.273	1.000	1.000	1.000	88	1.769	1.831
-NESS	0.000	0.999	0.001	NA	1.000	1.000	1242	0.009	0.009
-NIK	0.000	1.000	0.000	NA	1.000	NA	2	0.000	0.000
-O	0.068	0.767	0.041	1.000	1.000	1.000	73	1.133	1.233
-OID	0.556	0.444	0.000	1.000	1.000	NA	18	1.000	0.991
-OLOGY	0.000	1.000	0.000	NA	1.000	NA	31	0.000	0.000
-ON	0.020	0.740	0.140	0.010	0.142	0.111	50	1.382	1.301
-OR	0.012	0.970	0.018	0.054	0.113	0.033	166	0.603	0.224
-ORIMUM	0.000	1.000	0.000	NA	1.000	NA	2	0.000	0.000
-ORY	0.719	0.260	0.021	0.479	0.131	1.000	96	1.018	0.964
-OSIS	0.000	1.000	0.000	NA	1.000	NA	6	0.000	0.000
-OUR	0.000	0.750	0.250	0.000	0.011	0.054	20	0.603	0.811
-OUS	0.991	0.009	0.000	0.810	1.000	NA	227	0.040	0.073
-OUT	0.061	0.727	0.000	1.000	1.000	NA	33	1.284	1.361
-PATH	0.000	1.000	0.000	NA	1.000	NA	5	0.000	0.000
-PHONE	0.000	1.000	0.000	NA	1.000	NA	2	0.000	0.000
-PROOF	0.609	0.087	0.304	1.000	1.000	1.000	23	1.265	1.265
-RY	0.132	0.803	0.039	0.818	0.938	0.667	76	1.036	0.988
-SCAPE	0.000	1.000	0.000	NA	1.000	NA	2	0.811	0.000
-SCOPE	0.000	1.000	0.000	NA	1.000	NA	4	0.000	0.000
-SE	0.127	0.476	0.317	0.667	0.111	0.533	63	0.975	1.824
-SELF	0.000	0.000	0.000	NA	NA	NA	7	0.000	0.000
-SELVES	0.000	0.000	0.000	NA	NA	NA	3	0.000	0.000
-SHIP	0.000	1.000	0.000	NA	1.000	NA	74	0.000	0.000
-SOME	0.862	0.138	0.000	1.000	1.000	NA	29	0.567	0.579
-ST	0.091	0.379	0.288	1.000	1.000	1.000	66	2.282	2.295
-STER	0.026	0.868	0.105	1.000	1.000	1.000	38	0.613	0.657

–STRESS	0.000	1.000	0.000	NA	1.000	NA	3	0.000	0.000
–SY	0.333	0.600	0.067	0.778	0.226	1.000	30	0.830	1.231
–T	0.105	0.531	0.305	0.291	0.992	1.000	239	1.862	1.721
–TEEN	0.063	0.500	0.000	1.000	1.000	NA	16	1.253	1.272
–TH	0.000	0.643	0.020	NA	1.000	1.000	98	1.229	1.229
–TOMY	0.000	1.000	0.000	NA	1.000	NA	1	0.000	0.000
–TON	0.083	0.750	0.167	1.000	1.000	1.000	12	1.149	1.041
–TUDE	0.000	1.000	0.000	NA	1.000	NA	19	0.000	0.000
–TY	0.298	0.583	0.024	1.000	1.000	1.000	84	1.469	1.477
–UAL	0.857	0.143	0.000	0.846	1.000	NA	21	0.629	0.592
–ULAR	0.806	0.194	0.000	1.000	1.000	NA	31	0.650	0.709
–URE	0.049	0.728	0.222	0.000	0.039	0.120	81	0.603	1.030
–UTION	0.000	1.000	0.000	NA	1.000	NA	8	0.000	0.000
–WARD	0.431	0.086	0.034	1.000	1.000	1.000	58	1.404	1.620
–WARDS	0.000	0.042	0.000	NA	NA	NA	24	0.000	0.497
–WAY	0.047	0.837	0.000	1.000	1.000	NA	43	0.892	0.781
–WAYS	0.231	0.077	0.000	1.000	1.000	NA	13	1.095	1.140
–WISE	0.250	0.000	0.000	1.000	NA	NA	20	0.792	0.811
–Y	0.583	0.370	0.024	0.947	0.773	0.933	738	1.269	1.273
–YSE	0.000	0.000	1.000	NA	0.000	0.009	2	0.176	0.000
–YST	0.000	1.000	0.000	0.000	0.006	0.000	2	0.667	0.000

Appendix B

Nonword stimuli used in Experiment 1 (lexical category judgement task) and their predicted category based on diagnosticity. Each participant saw 40 nonwords from this list.

Nonwords with suffixes that are diagnostic for adjectives: hilable, cevable, gufable, hixable, gewous, dolous, raxous, domous, molless, cedless, jevless, halless, mibical, vefical, temical, vebical, jaflike, surlike, vawlike, joxlike, caxy, jebly, romy, moxy, talic, dohic, vevic, simic, cusive, horive, rurive, higive, terish, dodish, tumish, gomish, milatic, huxatic, salatic, vepatic.

Nonwords with suffixes that are diagnostic for nouns: rabness, juzness, tobness, mogness, vupity, tamity, cugity, cagity, hirmment, tilment, cilment, vixment, mubage, harage, hebage, jepage, cuxlet, cixlet, jixlet, cavlet, rogist, vusist, cumist, silist, gofan, coman, ruxan, dofan, vuder, gefer, jumer, vozer, sogee, vaxee, rilee, jufee, nadence, jofence, ralence, sigence.

Appendix C

Sentences and nonword stimuli used in Experiment 2 (spelling task). CON – congruent context where the target spelling is expected to occur, INC – incongruent context where the target spelling is not expected to occur. Target spellings for each phonological sequence are listed in Table 3.

Sentence	Context type	Congruency	Nonword played (IPA)
The candidate showed incredible <_> in a difficult situation.	noun	CON	ræbmənt
The lawyer demonstrated strong <_> when defending her client.	noun	CON	fɪpmənt
The girl explained the intelligent <_> of her new teacher.	noun	CON	ɔftmənt
The unpleasant officer remained <_> throughout his entire service.	adjective	INC	pefmənt
The tallest mountain peak seemed <_> to the inexperienced climbers.	adjective	INC	bɔvmənt
The new comedienne appeared <_> as she stepped onto the stage.	adjective	INC	rɔpmənt
The mayor thanked the amazing <_> for organising the parade.	noun	CON	fepər
The proposal was rejected by a bitter <_> holding a personal grudge.	noun	CON	vəumer
A campaign was launched by a fierce <_> to challenge the controversial changes.	noun	CON	mepər
Their uncle seemed <_> when he arrived at the party.	adjective	INC	kæksər
His unreliable friend proved <_> when asked to take responsibility.	adjective	INC	dʒu:bər
Her nervous kitten stayed <_> when the dog left the room.	adjective	INC	maɪgər
The traveller bought an unusual <_> for his collection.	noun	CON	nɔdlət
The chemist used a delicate <_> for the experiment.	noun	CON	sɔtlət
The historian explained the notorious <_> in a long article.	noun	CON	sæblət
The glistening water seemed <_> by the colourful lights.	verb	INC	tɔzlət
The opinions of the group seemed <_> by the contentious issue.	verb	INC	pɔflət
The quiet bay looked <_> in the soft moonlight.	verb	INC	gefɔlət
The patient had a horrible <_> with unusual symptoms.	noun	CON	hibɪst
The minister apologised to the innocent <_> for making the false accusation.	noun	CON	ɔktɪst
The security guard spotted the sneaky <_> on the CCTV monitor.	noun	CON	rebɪst
The whole situation appeared <_> to the new recruit.	adjective	INC	dʒætɪst
The popular film's plot seemed <_> to the film critic.	adjective	INC	mɪju:pɪst
The small child appeared <_> after he broke the expensive ornament.	adjective	INC	ɔmbɪst
The presentation recognised the impressive <_> of the protestors.	noun	CON	sedʒnɪs
The lighthouse keeper hated the lonely <_> of his occupation.	noun	CON	mɪbnɪs
The described steps had a logical <_> to their sequence.	noun	CON	fæznɪs
The pupils always quietly <_> when they sit in the back row.	verb	INC	dædʒnɪs
The mourners began to sadly <_> as the coffin disappeared.	verb	INC	kɔgnɪs
The spies were instructed to secretly <_> the enemy's government.	verb	INC	ɔɪnɪs
The fresh loaves and rolls smelled <_> as they entered the bakery.	adjective	CON	dʒəʊpəl
The snow on the hills looked <_> as the sun began to set.	adjective	CON	vəʊzəl
The expired food tasted <_> but they ate it anyway.	adjective	CON	geksəl
The worker was asked to carefully <_> the product for sale.	noun	INC	dʒu:fəl
The students were told to quickly <_> their homework assignments.	noun	INC	fəʊvəl

The musicians were known to creatively < __ > the famous tunes.	noun	INC	vu:təl
The impatient customer became < __ > when the manager appeared.	adjective	CON	hɒvɪkəl
The cloudy skies turned < __ > as the day progressed.	adjective	CON	sæfɪkəl
The dour professor remained < __ > as the student explained.	adjective	CON	vəʊdɪkəl
The journalist was ordered to quickly < __ > to the scene of the accident.	noun	INC	nebɪkəl
The composer was asked to skillfully < __ > the grand overture for the coronation.	noun	INC	veɪsɪkəl
The librarian needed to sternly < __ > everyone who returned their books late.	noun	INC	gæɪnɪkəl
The unfavourable outcome seemed < __ > to the experienced clerk.	adjective	CON	gæksləs
The strong allegations proved < __ > once the investigation was complete.	adjective	CON	ɔ:tɪləs
The grumpy bachelor stayed < __ > throughout his adult life.	adjective	CON	vɪksləs
The adventurer searched for the magical < __ > in the hidden cave.	noun	INC	dʒʌnləs
The strangers struck up an amiable < __ > over their shared predicament.	noun	INC	nɪmləs
Her father bought a pretty < __ > for her eighteenth birthday.	noun	INC	gə:ləs
The morning air smelled < __ > as they walked through the countryside.	adjective	CON	wɪvəs
The young messenger appeared < __ > when he brought them the letter.	adjective	CON	ɜ:vəs
The undercooked steak tasted < __ > so he sent it back to the kitchen.	adjective	CON	dəʊdɪs
The driver asked the conductor to have a quiet < __ > with the disruptive passenger.	noun	INC	reləs
The pilot warned that significant < __ > was likely during the landing.	noun	INC	ɪptəs
The cook made some tasty < __ > for the special menu.	noun	INC	lɪ:məs
The keynote speech sounded < __ > to the budding journalist.	adjective	CON	hefɪ
The latest problem seemed < __ > to the anxious leader.	adjective	CON	ʌmzɪ
The promotional materials looked < __ > but there was no time to reprint them.	adjective	CON	deɪpɪ
The representative was a dignified < __ > with many years of experience.	noun	INC	u:di
The gardener planted some green < __ > around the borders.	noun	INC	wezɪ
A press release about the illegal < __ > damaged the nominee's reputation.	noun	INC	gəmi
The shopkeeper started to wrongly < __ > the youth for stealing.	verb	CON	fju:beɪt
The ambassador was able to cheerfully < __ > in six different languages.	verb	CON	ɔ:seɪt
The weightlifter could easily < __ > the heavy stone onto the cart.	verb	CON	rəmeɪt
The auctioneer offered the precious < __ > to the highest bidder.	noun	INC	pɪ:meɪt
The protestors tied their official < __ > to the tree.	noun	INC	ɜ:pɪt
The vet rescued the injured < __ > from the ditch.	noun	INC	nemeɪt

Appendix D

Sentences and nonword stimuli used in Experiment 3 (eye-tracking task), List 1. List 2 was identical, but nonwords that were presented in congruent contexts in List 1 were presented in their corresponding incongruent contexts, and vice versa. CON – congruent context where the nonword spelling is congruent with the nonword's syntactic function, INC – incongruent context where nonword spelling is incongruent with its syntactic role in the sentence. N – noun, A – adjective, V – verb contexts.

Sentence beginning	Nonword	Sentence ending	Congruency	Context Type
The candidate showed incredible	rabness	in a difficult situation.	CON	N
The lawyer demonstrated strong	juzness	when defending her client.	CON	N
The girl explained the intelligent	vupity	of her new teacher.	CON	N
The man gushed about the exciting	tamity	of his proposed trip.	CON	N
The lighthouse keeper hated the lonely	hirment	of his occupation.	CON	N
The patient had a horrible	tilment	with unusual symptoms.	CON	N
The presentation recognised the impressive	mubage	of the protestors.	CON	N
The described steps had a logical	harage	to their sequence.	CON	N
The traveller bought an unusual	cuxlet	for his collection.	CON	N
The chemist used a delicate	cixlet	for the experiment.	CON	N
The historian explained the notorious	rogist	in a long article.	CON	N
The child knocked his favourite	vusist	off of the mantelpiece.	CON	N
Her husband bought a fragrant	gofan	for their anniversary.	CON	N
The textbook explained the historical	coman	that led to the riots.	CON	N
The newspaper printed a distorted	vudence	of the true events.	CON	N
The composer wrote a special	jofence	for the coronation.	CON	N
The librarian found an offensive	sogee	tucked inside the book's cover.	CON	N
The adventurer searched for the magical	vaxee	in the hidden cave.	CON	N
The strangers struck up an amiable	vuder	over their shared predicament.	CON	N
The security guard spotted the sneaky	gefer	on the CCTV monitor.	CON	N
The politician missed the urgent	dovate	after losing his phone.	INC	V
The apprentice visited the clever	fimate	to ask for advice.	INC	V
The museum hired an experienced	celade	to run the exhibition.	INC	V
The minister apologised to the innocent	vuxade	for making the false accusation.	INC	V
The committee produced a short	vizeer	summarising the key points.	INC	V
The driver asked the conductor to have a quiet	gudeer	with the disruptive passenger.	INC	V
The representative was a dignified	muxle	with many years of experience.	INC	V
The hospital saw many cases of the terrible	juxle	in the winter months.	INC	V
The cook made some tasty	vibyse	for the special menu.	INC	V
Her father bought a pretty	cesyse	for her eighteenth birthday.	INC	V
The gardener planted some green	rurive	around the borders.	INC	A
A press release about the illegal	higive	damaged the nominee's reputation.	INC	A
The mayor thanked the amazing	vevic	for organising the parade.	INC	A
The proposal was rejected by a bitter	simic	holding a personal grudge.	INC	A
A campaign was launched by a fierce	tumish	to challenge the controversial changes.	INC	A
The auctioneer offered the precious	dopish	to the highest bidder.	INC	A

The protestors tied their official	romy	to the tree.	INC	A
The vet rescued the injured	moxy	from the ditch.	INC	A
The train made a noisy	jaflike	as it pulled into the station.	INC	A
The pilot warned that significant	surlike	was likely during the landing.	INC	A
The unpleasant officer remained	gufable	throughout his entire service.	CON	A
The tallest mountain peak seemed	hixable	to the inexperienced climbers.	CON	A
The new comedienne appeared	raxous	as she stepped onto the stage.	CON	A
The fresh loaves and rolls smelled	domous	as they entered the bakery.	CON	A
The snow on the hills looked	jevless	as the sun began to set.	CON	A
The expired food tasted	halless	but they ate it anyway.	CON	A
The impatient customer became	temical	when the manager appeared.	CON	A
The cloudy skies turned	vebical	as the day progressed.	CON	A
The dour professor remained	vawlike	as the student explained.	CON	A
The experienced bank clerk became	joxlike	when he heard the outcome.	CON	A
The strong allegations proved	caxy	once the investigation was complete.	CON	A
The grumpy bachelor stayed	jeby	throughout his adult life.	CON	A
The keynote speech sounded	talic	to the budding journalist.	CON	A
The latest problem seemed	dobic	to the anxious leader.	CON	A
The promotional materials looked	cusive	but there was no time to reprint them.	CON	A
The morning air smelled	horive	as they walked through the countryside.	CON	A
The young messenger appeared	terish	when he brought them the letter.	CON	A
The undercooked steak tasted	dodish	so he sent it back to the kitchen.	CON	A
The tense atmosphere became	milatic	when his ex-wife arrived.	CON	A
The prevailing mood turned	huxatic	when the results were announced.	CON	A
The experienced presenter sounded	migide	during their conversation.	INC	V
Their drunk uncle seemed	guxide	when he arrived at the party.	INC	V
His unreliable friend proved	rolour	when asked to take responsibility.	INC	V
Her nervous kitten stayed	murour	when the dog left the room.	INC	V
The unlikely story sounded	hurice	to the suspicious policeman.	INC	V
The eager student seemed	tixice	to the older cohort.	INC	V
The proud lioness looked	maxure	to the circus master.	INC	V
The saucepan of stew smelled	vugure	to the hungry children.	INC	V
The whole situation appeared	vturat	to the new recruit.	INC	V
The local cuisine tasted	vavart	to many of the tourists.	INC	V
The keen organiser became	tobness	when she heard about the plan.	INC	N
The watchful guard dog turned	mogness	when he heard the intruder.	INC	N
The ancient scroll remained	hebage	as the scholar tried to decode it.	INC	N
The retired teacher became	jepage	in his old age.	INC	N
The proposed solution proved	cugity	and hard to understand.	INC	N
The shop assistant appeared	cagity	but was actually very helpful.	INC	N
The speaker's voice sounded	cilment	after the long lecture.	INC	N
The winning display seemed	vixment	compared to the other entries.	INC	N
The brilliant pathologist became	jixlet	when examining the evidence.	INC	N
The seeping wound smelled	cavlet	when they removed the dressing.	INC	N
The worker was asked to carefully	jizate	the product for sale.	CON	V
The students were told to quickly	haxate	their homework assignments.	CON	V
The musicians were known to creatively	morade	the famous tunes.	CON	V
The cleaners needed to neatly	cefade	their employer's belongings.	CON	V
The couple decided to mutually	gozeer	their shared assets.	CON	V
The spies were instructed to secretly	toxeer	the enemy's government.	CON	V
Her mother started to anxiously	virle	after seeing the missed phonecall.	CON	V
The thief managed to successfully	tible	capture for a whole week.	CON	V
The accountant tried to discretely	turyse	the company's regulations.	CON	V
The farmer dashed to swiftly	vazyse	his small flock.	CON	V
The dancer began to happily	cevide	in time to the music.	CON	V
The mourners began to sadly	faxide	as the coffin disappeared.	CON	V
The pupils always quietly	goxour	when they sit in the back row.	CON	V
The market traders started to loudly	tawour	for the best positions.	CON	V
The revellers began to merrily	cirice	towards the exit when the bar closed.	CON	V
The host started to grandly	jaxice	the guests towards the ballroom.	CON	V
The recipe instructed chefs to finely	gebure	the vegetables before cooking them.	CON	V
The bride began to slowly	somure	down the aisle towards the altar.	CON	V
The loud noise caused them to abruptly	vufart	around to see what was happening.	CON	V
The sleepy driver began to slightly	jefart	over to the wrong side of the road.	CON	V
The directive was written to firmly	mibical	the conditions of the deal.	INC	A
The small fee could barely	vefical	the necessary expenses.	INC	A
The baker used a heated knife to evenly	gewous	the icing over the cake.	INC	A
The chief surgeon was selected to expertly	dolous	the important operation.	INC	A
The ambassador was able to cheerfully	fawer	in six different languages.	INC	N
The weightlifter could easily	vozer	the heavy stone onto the cart.	INC	N
The shopkeeper started to wrongly	molless	the youth for stealing.	INC	A
The friends began to stupidly	cedless	as they left the party.	INC	A
The children always meanly	salatic	the substitute when their teacher is away.	INC	A
The nurse began to softly	vepatic	the baby to soothe it.	INC	A
The receptionist offered to helpfully	cumist	a taxi to take them home.	INC	N
His father needed to sternly	silist	him about his behaviour.	INC	N
The scientist needed to accurately	ruxan	the liquid with the measuring cylinder.	INC	N
The adjudicator began to truthfully	dofan	the entries in the competition.	INC	N
The entertainer liked to wittily	hilable	random members of the audience.	INC	A
The leader of the opposition encouraged people to openly	cevable	the new legislation.	INC	A

The helicopter was forced to rapidly	rilee	to the nearest airfield.	INC	N
The ice caused the car to randomly	jufee	all over the road.	INC	N
The members were not allowed to overtly	ralance	for any political party.	INC	N
The scholars liked to logically	sigence	about broad topics over lunch.	INC	N

References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17(9), 814–823. <https://doi.org/10.1111/j.1467-9280.2006.01787.x>
- Algeo, J. (1991). *Fifty years among the new words: A dictionary of neologisms, 1941–1991*. Cambridge, UK: Cambridge University Press.
- Amenta, S., Marelli, M., & Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: Early morpho-orthographic and morpho-semantic effects in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1587–1596. <https://doi.org/10.1037/xlm0000104>
- Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language*, 25(6), 726–740. [https://doi.org/10.1016/0749-596X\(86\)90046-X](https://doi.org/10.1016/0749-596X(86)90046-X)
- Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, 59(5), 920–948. <https://doi.org/10.1080/02724980443000782>
- Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3), 438–481. <https://doi.org/10.1037/a0023851>
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database, (CD-ROM)*. University of Pennsylvania, PA: Linguistic Data Consortium.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.7. <https://cran.r-project.org/web/packages/lme4/index.html>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beach, L. R. (1964). Cue probabilism and inference behavior. *Psychological Monographs: General and Applied*, 78(5–6), 1–20. <https://doi.org/10.1037/h0093853>
- 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. <https://doi.org/10.1353/lan.2017.0000>
- Berg, K., Buchmann, F., Dybiec, K., & Fuhrhop, N. (2014). Morphological spellings in English. *Written Language and Literacy*, 17(2), 282–307. <https://doi.org/10.1075/wll.17.2.05ber>
- Carney, E. (1994). *A survey of English spelling*. London: Routledge.
- Carroll, L. (1983). Through the looking-glass, and what Alice found there. In J. Prelutsky & A. Lobel (Eds.), *The random house book of poetry for children*. Retrieved from <http://www.poetryfoundation.org> (Original work published 1871).
- Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, 30(3), 348–369. [https://doi.org/10.1016/0749-596X\(91\)90041-H](https://doi.org/10.1016/0749-596X(91)90041-H)
- Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin and Review*, 8(3), 519–523. <https://doi.org/10.3758/BF03196187>
- Coltheart, M., Davelaar, E., Jonassen, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.
- Dawson, N., Rastle, K., & Ricketts, J. (2017). Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 645–654. <https://doi.org/10.1037/xlm0000485>
- De Jong, N. H., IV, Schreuder, R., & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, 15(4–5), 329–365. <https://doi.org/10.1080/01690960050119625>
- De Saussure, F. (1983). *Course in general linguistics*. LaSalle: Open Court.
- Dingemans, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>
- Fabb, N. (1988). English suffixation is constrained only by selectional restrictions. Retrieved from *Natural Language and Linguistic Theory*, 6(4), 527–539. <http://www.jstor.org/stable/4047592>
- Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, 103(32), 12203–12208. <https://doi.org/10.1073/pnas.0602173103>
- Fitneva, S., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language*, 36(5), 967–997. <https://doi.org/10.1017/S0305000908009252>
- Ford, M. A., Davis, M. H., & Marslen-Wilson, W. D. (2010). Derivational morphology and base morpheme frequency. *Journal of Memory and Language*, 63(1), 117–130. <https://doi.org/10.1016/j.jml.2009.01.003>
- Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with grammatically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology Section A*, 39(2), 211–251. <https://doi.org/10.1080/14640748708401785>
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, 35(1), 116–124. <https://doi.org/10.3758/BF03195503>
- Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, 35(5), 263–279. <https://doi.org/10.1017/S0140525X11001841>
- Houston-Price, C., Plunkett, K., & Harris, P. (2005). 'Word-learning wizardry' at 1;6. *Journal of Child Language*, 32(1), 175–189. <https://doi.org/10.1017/S0305000904006610>
- Juhász, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, 94(2), 223–244. <https://doi.org/10.1348/000712603321661903>
- Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in Psychology*, 94, 67–84. [https://doi.org/10.1016/S0166-4115\(08\)62789-2](https://doi.org/10.1016/S0166-4115(08)62789-2)
- Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99(2), 349–364. <https://doi.org/10.1037/0033-295X.99.2.349>
- Kemp, N., Nilsson, J., & Arciuli, J. (2009). Noun or verb? Adult readers' sensitivity to spelling cues to grammatical category in word endings. *Reading and Writing*, 22, 661–685. <https://doi.org/10.1007/s11445-008-9140-z>
- Kessler, B., Treiman, R., & Mullenix, J. (2008). Feedback-consistency effects in single-word reading. In E. L. Grigorenko, & A. J. Naples (Eds.), *New directions in communication disorders research: Integrative approaches. Single-word reading: Behavioral and biological perspectives* (pp. 159–174). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627–633. <https://doi.org/10.3758/BRM.42.3.627>
- Ktori, M., Mousikou, P., & Rastle, K. (2018). Cues to stress assignment in reading aloud. *Journal of Experimental Psychology: General*, 147(1), 36–61. <https://doi.org/10.1037/xge0000380>
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20(2), 121–157. [https://doi.org/10.1016/0010-0285\(88\)90017-5](https://doi.org/10.1016/0010-0285(88)90017-5)
- Mirković, J., Forrest, S., & Gaskell, G. (2011). Semantic regularities in grammatical categories: Learning grammatical gender in an artificial language. In 33rd annual meeting of the Cognitive Science Society, 20–23 July 2011, Boston, MA, USA. Retrieved from <http://ray.yorksj.ac.uk/id/eprint/2179>
- Mirković, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PLoS ONE*, 11(4), e0152489. <https://doi.org/10.1371/journal.pone.0152489>
- Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, 96(2), 143–182. <https://doi.org/10.1016/j.cognition.2004.09.001>
- Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (2012). *Methodological and analytic frontiers in lexical research* (pp. 13–31). <https://doi.org/10.1075/bct.47.02mon>
- Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, 140(3), 325–347. <https://doi.org/10.1037/a0022924>
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130299. <https://doi.org/10.1098/rstb.2013.0299>
- Mousikou, P., Sadat, J., Lucas, R., & Rastle, K. (2017). Moving beyond the monosyllable in models of skilled reading: Mega-study of disyllabic nonword reading. *Journal of Memory and Language*, 93, 169–192. <https://doi.org/10.1016/j.jml.2016.09.003>
- Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, 15(4–5), 389–420. <https://doi.org/10.1080/01690960050119643>
- Onnis, L., & Christiansen, M. H. (2008). Lexical categories at the edge of the word. *Cognitive Science: A Multidisciplinary Journal*, 32(1), 184–221. <https://doi.org/10.1080/03640210701703691>
- 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. <https://doi.org/10.1080/01690960050119661>
- Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, 20(1–2), 261–290. <https://doi.org/10.1080/01690960444000098>
- Powell, M. J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report. Department of applied mathematics and theoretical physics, University of Cambridge.
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23(7–8), 942–971. <https://doi.org/10.1080/01690960802069730>
- Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and

- semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, 15(4–5), 507–537. <https://doi.org/10.1080/01690960050119689>.
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review*, 11(6), 1090–1098. <https://doi.org/10.3758/BF03196742>.
- Rayner, K., & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.). *Oxford library of psychology. The Oxford handbook of eye movements* (pp. 751–766). New York, NY, US: Oxford University Press.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1290–1301. <https://doi.org/10.1037/0278-7393.30.6.1290>.
- Seidenberg, M. (2017). *Language at the speed of sight: How we read, why so many can't, and what can be done about it*. New York: Basic Books.
- Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143–174. <https://doi.org/10.1348/000712603321661859>.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- St. Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, 116(3), 341–360. <https://doi.org/10.1016/j.cognition.2010.05.012>.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 638–647. [https://doi.org/10.1016/S0022-5371\(75\)80051-X](https://doi.org/10.1016/S0022-5371(75)80051-X).
- Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, 79, 1–39. <https://doi.org/10.1016/j.cogpsych.2015.03.003>.
- Treiman, R., & Bourassa, D. C. (2000). The development of spelling skill. *Topics in Language Disorders*, 20(3), 1–18. <https://doi.org/10.1097/00011363-200020030-00004>.
- Venezky, R. L. (1972). *The structure of English orthography, Vol. 82*. Berlin, Boston: De Gruyter Mouton.
- Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition*, 111(1), 132–137. <https://doi.org/10.1016/j.cognition.2008.12.011>.
- Ziegler, J. C., Petrova, A., & Ferrand, L. (2008). Feedback consistency effects in visual and auditory word recognition: Where do we stand after more than a decade? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(3), 643–661. <https://doi.org/10.1037/0278-7393.34.3.643>.
- Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers*, 29(4), 600–618. <https://doi.org/10.3758/BF03210615>.