

Defining cognate similarity in psycholinguistics: A phonological perspective

Isabella Fritz*, Joshua Booth* & Aditi Lahiri

Language and Brain Laboratory, Faculty of Linguistics, Philology & Phonetics, University of Oxford

*Both authors contributed equally to this work.

ABSTRACT

In psycholinguistics, COGNATES are defined as translation equivalents that overlap in meaning, phonology and orthography (for same-script languages). Different approaches have been used to quantify degree of phonological overlap, leading to differences across studies in the categorisation of cognates vs. non-cognates, where diverging experimental results have been found (ranging from facilitation to inhibition). We outline an alternative perspective on L1–L2 phonological correspondences, drawing on insights from theoretical phonology and historical linguistics, highlighting the systematicity of phonological variation across languages: cf. English–German [w]ax–[v]achs, [w]ind–[v]ind (inherited) vs. *crédit*–*Kredit*, *átom*–*Atóm* (loanwords with different stress). We discuss the implications for psycholinguistic studies, particularly the necessity of controlling for prosodic differences. However, we also propose that such systematic phonological differences present opportunities for psycholinguistic studies, allowing the researcher to tease apart nuanced questions of lexical representation and explore the role played by various phonological dimensions in lexical access.

Keywords: Mental lexicon, phonology, L2, cognates, word recognition

1. Introduction

Research on the bilingual mental lexicon and its architecture is often based on experiments comparing cognates with non-cognates (Sánchez-Casas & García-Albea, 2005). In psycholinguistics, the term COGNATE is used to refer to “translation equivalents” that must overlap in meaning, phonological and (for same script-languages) orthographic form, even if not strictly identical; e.g., English *tomato* /tə'mɑ:təʊ/¹ – Spanish *tomate* /to'mate/ (Dijkstra, Grainger, & van Heuven, 1999; note that these words overlap more in ORTHOGRAPHY than phonology). The received view is that cognates are easier to access, produce and process compared to controls which do not have any orthographic or phonological overlap (taken as evidence for an integrated bilingual mental lexicon). This facilitation effect has been found most reliably in the visual modality (see Lijewska, 2020 for a review), whereas auditory word recognition has received much less attention in the literature, with one of the most influential models of bilingual word recognition, the Bilingual Interactive Activation model (BIA: Dijkstra & van Heuven, 2002; later BIA+: Dijkstra et al., 2002; and Multilink: Dijkstra et al., 2019), being based exclusively on visual word recognition.

Similarity is a spectrum, ranging from total identity at one extreme (e.g., English *nest* [nest] – German *Nest* [nest]) to totally unrelated forms at the other (e.g., English *fog* [fɒg] – German *Nebel* [ne:bəl] “fog”). This poses a problem; as highlighted in a recent review paper by Roembke and colleagues (2024), the dichotomised definition of cognate status usually employed in psycholinguistic studies raises the question of a CUT-OFF POINT between cognates and non-cognates, and this has been operationalised differently across studies. For definitions based on form similarity, the choice of where to draw the line seems arbitrary. At what point do words cease to “overlap” or be “similar” and how are orthographic and phonological similarity to be defined?

Various approaches to calculating cognate similarity exist, including similarity ratings by bilingual or monolingual speakers (e.g., used in Muntendam et al., 2022) but also a range of algorithms (Levenshtein Distance: Levenshtein, 1966; ALINE: Kondrak, 2003, PHOR-in-One: Costa et al., 2023). These algorithms provide numerical similarity scores but, even so, the issue of a cut-off point remains. As Roembke et al. (2024) note, the make-up of experimental conditions still depends on an arbitrary delineation which can lead to the same items being assigned to different categories across studies. For example, the English – Dutch word pair *king* – *koning* was classified as cognate by de Groot and Nas (1991) but non-cognate by Muylle and colleagues (2022). Thus, differences in how phonological similarity is measured may explain some of the inconsistencies in findings across word recognition studies.

In this paper, we first provide an overview of the current psycholinguistic approaches to cognate research and quantifying phonological similarity. We then demonstrate how different measures of phonological similarity can lead to inconsistencies in the categorisations of word-pairs vis-à-vis cognate status, including the sub-categorisation of cognates across different conditions according to their degree of phonological overlap (i.e., low vs. high). In order to illustrate this point, we present a case study applying three algorithms proposed to quantify similarity (i.e., ALINE, PHOR-in-One and the Normalised Levenshtein Distance (NLD)) to the same dataset. These data are taken from a recent study by Tiffin-Richards (2024) which explores in detail the role of phonological and orthographic overlap in cognate processing. This paper is particularly useful, as it investigates phonological similarity in cognates and includes a full list of experimental stimuli.

In regard to the algorithms employed to quantify cross-linguistic similarity, this approach has an unfortunate effect of conflating various phonological dimensions which may not equally impact the word recognition. Drawing on examples from different language

combinations, we demonstrate that, although such algorithms certainly can provide a good indicator of overall phonological similarity (when non-phonemic differences are properly controlled), they alone may not be the most appropriate method for assessing which cross-linguistic phonological similarities impact L2 word recognition and processing.

We therefore provide an alternative perspective on phonological similarity. What is often lacking in current approaches is a recognition of the fact that languages are structured systems and phonological variation across languages is highly systematic. We demonstrate the importance of this systematicity with reference to loanwords and inherited cognates. In linguistics, the term “cognate” exclusively refers to words inherited from a common ancestor (e.g., English *daughter* [ˈdɔ:tə] – German *Tochter* [ˈtɔxtə] – Dutch *dochter* [ˈdɔxtər]), in contrast to borrowed LOANWORDS. Loans may of course come from a shared source (cf. English *reptile* [ˈrɛptɪl] – German *Reptil* [ʁɛpˈti:l], both from French), which may be termed SHARED LOANS (cf. Fritz et al., 2023). Alternatively, the first language may have borrowed the relevant item from the second, e.g., Mandarin 巧克力 [tʃʰjaʊ³kʰʰli⁴] or Bengali চকোলেট [ˈtʃɔkolet] from English *chocolate* [ˈtʃɔklət]. Considering these categories separately, it can be seen that historical sound change produces systematic correspondences in cognates (e.g., English *pepper* [pʰɛpə], *pipe* [pʰaɪp], *pan* [pʰæn] – German *Pfeffer* [pʰɛfɐ], *Pfeife* [pʰaɪfə], *Pfanne* [pʰanə]), whereas recent loans might be much closer phonologically (cf. English *hotel* [həʊˈtɛl] – German *Hotel* [hoˈtɛl]). Whilst shared loans may also exhibit systematic correspondences (due to the adaptation of non-native phonemes), phonological adaption to the host language is likely to produce correspondences along different phonological

dimensions to those found in cognates, such as differences across Germanic languages in the placement of stress in French loans (cf. English *moral* ['mɔ:ɪ] – German *Moral* [mo'ʁa:l]).

Although native speakers do not acquire or store etymological knowledge in the lexicon, the history of words shapes their current phonological form, resulting in classes of words within the lexicon which share similar properties and will often differ in predictable ways from their cognates in a related language. For the psycholinguist interested in the role of phonology in perception, these patterns can be exploited in experimental design and an awareness of (theoretical and historical) linguistics can be valuable in tackling nuanced questions of lexical representation. We therefore advocate for linguistically-informed study designs, guided by the systematicity of phonological grammars. However, even where phonology is not the direct object of research, these concerns are no less relevant; differences in phonological structure should be carefully controlled, even where there is little orthographic difference (cf. English *ballet* ['bæleɪ] – German *Ballett* [ba'let]). Beyond quantifying phonological overlap, it is necessary to consider factors such as syllabification and stress, as well as critically considering the transcriptions to which algorithms are applied.

2. Cognate similarity

Previous research has demonstrated that, in visual word recognition, same script bilinguals coactivate the corresponding word form in the non-target language(s), activating their shared semantic representation. In turn, this coactivation leads to faster recognition of cognates than non-cognates, resulting in a cognate facilitation effect (see Lijewska, 2020, for an overview). The largest facilitation effects have been found for translation equivalents with precisely the same orthography (referred to as “identical cognates”), e.g., European Portuguese and English *piano* (Arana et al., 2022). Importantly, studies with bilinguals of different-script languages have also found that cognates are processed differently to non-cognates, showing

that phonological overlap impacts bilingual word recognition independent of orthography, as in the English *guitar* and Japanese ギター [gʲita:] (e.g., Spivey & Marian 1999; Voga & Grainger, 2007). Evidence has been found, for example, in masked priming studies (Voga & Grainger, 2007: Greek–French), with phonologically related distractors from the L1 also being shown to impact eye-movement in an L2 visual world paradigm (Spivey & Marian, 1999: English–Russian).

In recent years, same-script bilingual studies have provided further evidence for this phonological coactivation in visual word recognition tasks (see Zhang & Wang, 2025 for a review). In contrast to visual studies, however, results from auditory experiments in fact suggest that the larger the phonological overlap, the greater the competition for lexical access across languages. This has variously led to reduced cognate effects, cognate null-effects or even inhibition in lexical decision tasks, both with auditory priming (e.g., Fritz et al., 2023) and without (e.g., Frances et al., 2023, Cornut et al., 2022). Generally speaking, these effects are the *reverse* of the facilitation found in the visual modality. Given these diverging results, current research findings make it difficult to pinpoint the role of phonology in both visual and auditory word recognition. Taking a more systematic approach to phonology might allow us explore the cause of this variation across studies, where different approaches to quantifying phonological similarity have been used (i.e., algorithms and rating tasks). The conflation of various phonological differences makes it further difficult to disentangle precisely which factors (e.g., phonemic, featural or prosodic) impact bilingual word recognition, and in what ways.

Quantifying the phonological similarity of two words is not a trivial task. As outlined above, previous studies have usually focused on the visual modality and often only considered *orthographic* similarity, using a number of different measures. One such

approach, Van Orden's (1987) graphemic similarity algorithm, focuses on shared letters rather than phonemes, prioritising the first and last letter in the word and controlling for length. This approach is thus less-well suited to auditory studies, where the presentation of sounds is linear. Another approach to graphemic similarity, the Normalised Levenshtein Distance (NLD, Levenshtein, 1966), has also frequently been used in studies involving languages which use the Roman alphabet. The NLD effectively compares letter strings by calculating the proportion (0–1) of identical characters. This is done by counting how many changes (e.g., deletion, substitution or addition) are necessary to convert letter string A into letter string B, whilst controlling for word length. For instance, the orthographic NLD of the English – German word pair *pepper* /pɛpə/ – *Pfeffer* /pʰɛfɐ/ would be 0.43. Using a common cut-off point of 0.5 for the NLD (cf. Roembke et al., 2024), this word-pair would not be classified as cognates, despite the fact that they are intuitively very similar words. The question becomes, what is the basis of this intuition, phonologically speaking?

Although the NLD has also been used to calculate phonological similarity (e.g., Guediche et al., 2020), focusing on phonemes rather than letters, there are several drawbacks to this algorithm (cf. Costa et al., 2023). The NLD operates only on a phonemic level and does not take any other phonological dimensions into account (e.g., word stress, syllabification, phonological features). Most importantly, this approach assumes that each change (i.e., deletion, substitution, addition) is (i) just as meaningful, regardless of position in the word, and (ii) equally affects the similarity status. The latter problem has received greater attention than the former, with a number of alternative algorithms having been proposed which seek to address the issue by including different weighting for different types of changes. Most recently, Costa and colleagues (2023) proposed a modified phonological Levenshtein distance measurement where different types of phonological adjustments (e.g.,

substitution, insertion, deletion) are given different weightings. Their algorithm is based on a similar one put forward by Schepens and colleagues (2013). In their PHOR-in-One algorithm, insertions and deletions are considered to impact similarity more than substitutions, thus resulting in a lower similarity score. In particular, for substitutions, the Euclidean distance between the source and target phonemes is calculated, with smaller distances indicating greater phonological overlap. However, close inspection of the algorithm reveals that distances are calculated based on the positioning of phonemes on the abstract IPA chart (which is idealised and to an extent arbitrary), rather than acoustic measures or natural classes of sounds, as can be seen in Figure 1 (the IPA chart's idealised cardinal vowel quadrilateral).

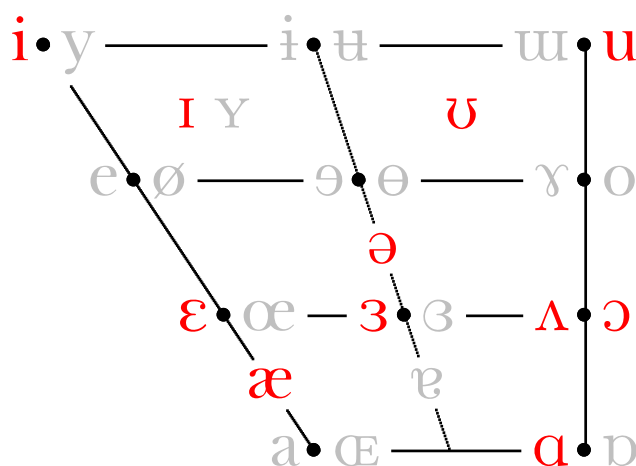


Figure 1: Idealised IPA chart, with placement of English monophthongs highlighted in red.

A discussion of these issues can be found in Ladefoged & Johnson (2015, p. 234ff.), which compares the average F1 and F2 frequencies (the acoustic correlates of tongue height and backness, respectively) of the vowel phonemes of Californian English, (U.S.) northern cities English and (UK) BBC English. Across these varieties, the relative positions of the same two phonemes might differ substantially with respect to height or backness. For instance, /ʌ/ is higher than /æ/ in Californian and BBC English, but lower than /æ/ in

northern cities English. Conversely, /ʌ/ is higher than /ɔ:/ (~625 Hz as opposed to ~650 Hz) in northern cities English, but considerably lower in BBC English (~700 Hz for /ʌ/ as opposed to ~450 Hz for /ɔ:/). See also Ladefoged (1975, p. 194ff.) for an illustration of individual variation regarding relative vowel positions, comparing his own vowels to the average American English vowels.

The ALINE distance (Kondrak, 2003) provides a much more nuanced approach, acknowledging that not all sounds differ equally and considering differences below the level of individual phones. As with the PHOR-in-One algorithm, the ALINE distance incorporates a weighting system which can be optionally adjusted, treating each speech sound as a vector of feature values, based on Ladefoged's (1975) "measurements of distances between vocal organs during speech production" (Kondrak, 2003, p. 14). As such, this approach is based on PHONETIC rather than PHONOLOGICAL features and calculates (articulatory) phonetic similarity. The algorithm is capable of producing nuanced scores of phonetic similarity, but this point must be borne in mind if the object is to investigate PHONOLOGICAL similarity. Furthermore, the accuracy of the scores is highly dependent on the quality of the transcriptions fed into it, requiring researchers to be conscious of the decisions they are making (see Section 3).

In phonological theory, distinctive features are monovalent or binary and define phonemic contrasts and natural classes of sounds which pattern together phonologically, and it is these features which form the basis of underlying mental representations. In contrast, the ALINE distance relies on multivalued features which divide certain phonetic dimensions along detailed articulatory lines, assigning numerical values to each. Compare, for instance, the tertiary distinction of PLACE in phonological feature systems ([LABIAL], [CORONAL] and [DORSAL])² with the eleven place features included in the ALINE algorithm: bilabial,

labiodental, dental, alveolar, retroflex, palato-alveolar, palatal, velar, uvular, pharyngeal and glottal. Phonetic descriptors do not necessarily make predictions about what is being activated phonologically and, for two given speech sounds, it is not necessarily the case that there is a linear relationship between differences in the articulatory posture of the speech organs and distance in the resulting acoustic signals or phonological perception. This raises two important points: (i) the two algorithms (ALINE and PHOR-in-One), due to their differing approaches, can produce highly divergent scores, even when comparing the same word pairs; and (ii) although these algorithms (to a greater or lesser degree of granularity) provide good general indicators of surface phonetic similarity, phonological representations must also be considered if the aim is to make predictions about what is or is not activated in perception. We will illustrate the former issue in Section 3, where scores from different algorithms are compared, before moving to a discussion of the latter in Section 4.

Costa et al. (2023, p. 3700) argue that the results of their modified phonological NLD are “easily interpretable and allow for an objective characterization of translation equivalents as identical cognates, non-identical cognates or noncognates, with a threshold of .50 typically considered for cognate inclusion.” The problem remains, however, that there is a great deal of variety in approaches to categorising cognate status across studies. The cut-off point between cognate and non-cognate remains arbitrary and, as a number of studies additionally categorise cognates depending on degree of phonological overlap, so does the distinction between cognates with high and low phonological similarity. To achieve this, the similarity threshold is adjusted to a given score, for example by using a median split of the similarity scores across items (e.g., Frances et al. 2021). By doing so, the English–Spanish word pair *object* /'abʒekt/ and *objeto* /ob'xeto/ was classified as having a low phonological overlap while

band /bænd/ and *banda* /'banda/ was classified as high phonological overlap (US English and Spanish transcriptions are taken from Frances et al., 2021).

The same is true for studies relying on speakers' rating of word-pairs and their phonological similarity, either by bilinguals (Van Assche et al., 2011) or by speakers who only know one of the two languages (e.g., Degani et al., 2018). Although such ratings can also provide a good overall indicator of phonological similarity, they cannot tell us which factors raters consider in their judgments. Roembke et al.'s (2024) review paper on cognates includes 15 studies that used such similarity ratings (out of 91). If mentioned in the paper, the similarity scales were either from 1–5 or 1–7, lower scores indicating less phonological (or orthographical) overlap. The ratings were then used either to confirm that chosen word-pairs had been assigned to the appropriate categories (i.e., cognate vs. non-cognate, e.g., Groot & Nas, 1991), to test statistically whether ratings differ across experimental conditions (e.g., Elias & Degani, 2022, Muntendam et al., 2022) or were otherwise included as a factor in the statistical analysis (e.g., Cai et al., 2011; Dijkstra et al., 2010). The question here is therefore not so much about selecting stimuli according to a clear cut-off point as confirming the allocation of items which have been selected based upon researcher intuitions.

Although also a good general indicator of phonological similarity, naive informants are unlikely to be able to formalise in linguistic terms how they approached the ratings. In other words, we do not know which factors influence these decisions (e.g., orthography, phonology, semantics, frequency etc.). In other words, it is unclear how to tease apart precisely which linguistic characteristics underpin these intuitions, or whether speakers of different languages rely on the same criteria, making it difficult to formalise any notion of cognate relatedness based on such ratings.

3. Comparing algorithm similarity scores

To demonstrate the impact of different algorithms on phonological similarity scores, we applied three different algorithms (ALINE, PHOR-in-One and NLD) to the same stimuli set. The stimuli were taken from a recently published study by Tiffin-Richards (2024) which investigated cognate facilitation in both a lexical decision and a reading task with German L1 speakers who are proficient in their L2 English. Relevant for our analysis is the study's experimental manipulation of the cognates' phonological and orthographic overlap. The experiments comprised three conditions: orthographically identical cognate pairs (e.g., E *minute* ['mɪnɪt] – G *Minute* [mi'nu:tə]; note that these are far from identical *phonologically*), close cognate pairs (e.g., E *elephant* ['ɛlɪfənt] – G *Elefant* [ele'fant]) and non-cognate pairs (e.g., E *song* [sɒŋ] – G *Lied* [li:t]). To calculate phonological similarity, the author used the Normalised Levenshtein Distance (also included in our analysis). As in Tiffin-Richards (2024), we used the DISC system (cf. Baayen et al., 1995) where a phoneme is represented with a single ASCII character. We calculated the NLD using the *stringsim* function (method = “lv”) which is part of the *stringdist* package (van der Loo, 2014). Secondly, we used the *Natural Language Toolkit* (NLTK, Bird et al., 2009) which includes a function for calculating the ALINE score. The feature matrix is based on the original weighting system proposed by Kondrak (2002). We normalised the ALINE score by the average score of word self-comparisons, as suggested by Downey and colleagues (2008). The Levenshtein distance was normalized by maximum string length (a commonly used approach). As input strings, we used the transcriptions provided by our third measurement, i.e., the adjusted NLD put forward by Costa and colleagues (2023). For the PHOR-in-One, no package or algorithm was available to calculate the distance scores, however, the authors provide a corpus of scores for

over 6,000 words, including comparisons of German – British English pairs. Our meta-analysis and the used stimuli plus their transcriptions can be found on (<https://osf.io/fsyg9>).

As the ALINE matrix does not include all phonemes from the British English inventory, a number of further adjustments to our transcriptions were necessary. For example, the phoneme /ɪ/ was not available in the algorithm and was therefore replaced by another phoneme, the short tense high front vowel /i/ (which is not a contrastive phoneme in English). We made some further informed decisions on which transcriptions to feed into the algorithm. For example, we replaced various phonetic relations of /r/ (i.e., [ʁ], [ɹ]) with the phonemic /r/. However, we retained the devoiced allophones of the German obstruents in final position (where the voicing contrast is neutralised). Finally, we removed the stress marks from the transcriptions, as they are not considered in the default ALINE algorithm (but could be added as a new phonological feature). A full list of the necessary adjustments made to the transcriptions in order to be able to run the package is available on our OSF page.

Although the PHOR-in-One corpus included most of the 162 items from the Tiffin-Richards (2024) stimuli set, 27 word-pairs were either absent from the corpus or the translation equivalents provided in the corpus did not match those chosen by Tiffin-Richards (e.g., *Schurke* – *villain* vs. *Bösewicht* – *villain*). We therefore do not have a complete set of scores for the PHOR-in-One algorithm, making a comparison with both the originally reported values and our computed scores difficult. We transcribed the missing items to be included for the ALINE score by adhering as closely as possible to the transcriptions provided in the PHOR-in-One. Table 1 shows the descriptive results of our analyses for each of the three algorithms across the three experimental conditions tested in Tiffin-Richards

(2024) as well as the author's similarity scores (note the missing values for the PHOR-in-ONE algorithm).

Table 1: Similarity scores across different algorithms.

Condition	Phonological NLD (as in Tiffin-Richards) ¹	NLD (based on PHOR-in-One transcription)	PHOR-in-One	ALINE
Identical cognate	0.61 (0.17)	0.585 (0.18)	0.813 (0.14)	0.930 (0.08)
Close cognate	0.49 (0.22)	0.499 (0.20)	0.815 (0.12)	0.916 (0.09)
Non-cognate	0.05 (0.10)	0.095 (0.13)	0.421 (0.13)	0.549 (0.14)

Generally, the mean scores across conditions and algorithms show that the weighting underlying the ALINE and the PHOR-in-ONE has a large impact on the similarity scores compared to the non-weighted normalised NLD which, overall, shows much lower scores. Differences are also apparent when zooming into individual word pairs. For example, the word pair *circus* [sɜ:kəs] – *Zirkus* [tsɪkəs] scored 0.904 in the ALINE, 0.815 in the PHOR-in-One, but 0.286 in the NLD. Another example is the identically-spelt word-pair *museum* [mju:ziəm] – *Museum* [muze:ʊm], which had a high ALINE score (0.840) but low PHOR-in-One (0.567) and NLD (0.500) scores. Importantly, depending on the algorithm applied, this word-pair may have been put into different conditions (cognate with high vs. low overlap, or even non-cognate).

The issue with relying on a hard cut-off point in order to quantify cognate similarity (particularly as the sole criterion) lies in their arbitrariness: crucially, any cut-off point would necessarily vary considerably from algorithm to algorithm. For instance, depending on the precise aims of the researcher (for instance, is the aim to differentiate between cognates and noncognates, or cognates with higher and lower overlap?), a cut-off point of 0.5 might be

¹ In Tiffin-Richards' (2024), identical cognates were coded as 0 and 1 was maximally distant. Since we report the values vice versa, we also subtracted the values in the original paper from 1.

reasonable when applied to NLD scores but, for other measures, a much higher cut-off point might be warranted, as the ALINE algorithm, for instance, results in scores which are considerably inflated and crowded in the upper end of the scale. For example, totally unrelated words might still get a score above 0.5 (e.g., English *nag* [næg] – German *meckern* [mɛkɐn]: 0.672, due to their initial sequences of nasal+front vowel+dorsal stop). As a result, a cut-off point more like 0.75 might be more apt for near-cognates (e.g., English *tongue* [tʌŋ] – German *Zunge* [tsʊŋə]: 0.764) or even 0.9 for near-identical cognates (e.g., English *milk* [mɪlk] – German *Milch* [mɪʃ]: 0.969). Ultimately, our recommendation is that researchers use their linguistically-informed judgement to categorise words, as categorisations may very well vary depending on the precise issue being investigated and use algorithms as an objective measure to confirm such linguistically-guided intuitions, particularly in terms of statistical analysis.

Another general observation is that scores are highly dependent on whether the algorithm is fed a phonological or a phonetic transcription. The distinction here is between abstract phonological transcriptions which include only meaningful, *contrastive* differences in phonemes (e.g. /p/ vs. /b/, cf. *pad–bad*) and detailed phonetic transcriptions, which include non-contrastive variation, focused more on acoustics and articulation than mental representations (e.g., *butter* which, depending on dialect, might be realised as [bʌʔə], [bʌɾə], [bʌtə] etc.). In other words, a phonological transcription of *coral* in British English might be /kɒrəl/, but a more phonetic transcription would include details such as aspiration, velarisation, e.g., [ˈkʰɒ.ɹəl].

In our meta-analysis, there does not seem to be a clear distinction, with phonological and phonetic transcriptions often being mixed (e.g., allophonic /r/ and /l/ transcriptions in the PHOR-in-One). Thus, it would be helpful if researchers elaborated on the transcriptions that they fed into the algorithms in terms of phonemic or phonetic detail, as the examples we have outlined above demonstrate. Moreover, making decisions on whether to put in phonetic or an underlying transcription is often not trivial. For instance, /l/ might be realised as “clear” [l] or “dark” [ɫ], with the latter restricted to syllable coda position in Standard Southern British English. Modern Standard German does not make this distinction, but there is variation within both languages in the realisation of /l/. Other dialects of English use *only* [l] or [ɫ], while still others vocalise it in coda position (as do many Bavarian dialects of German: cf. Bavarian *Muich* – Modern Standard German *Milch* “milk”). Is such phonetic detail desirable in our transcriptions if we are concerned with phonological representations? As an example, the English transcriptions provided by Costa and colleagues (2023) include allophonic variation between clear [l] and dark [ɫ] (and non-phonemic assimilation of /m/ in words such as *triumph* [tɹaɪəmf]) but do not include similar allophonic variation in the aspiration of syllable-initial voiceless stops before a stressed vowel, e.g., [p^heɪpə], [t^hɛst] (as opposed to /peɪpə/, /tɛst/).

Arguably, all algorithms can provide a good overall indicator of phonological similarity, with varying degrees of nuance (i.e., weighted vs. non-weighted algorithms). However, based on both psycholinguistic and phonological work, we argue that, before running such an algorithm, there are certain phonological considerations which should be

taken into account and controlled for (but which are overlooked by current algorithms). These relate to stress placement and syllable count in word-pairs, and manually checking that the transcriptions used in the algorithm are consistent across stimuli. We will outline examples of these three factors below.

None of the three similarity measures that we have discussed take into account the placement of word stress, which is also related to the word's metrical and syllable structure. Previous studies have shown that stress placement differences in cognates do affect word recognition processes (e.g., Fritz et al., 2023, Muntendam et al., 2022). To avoid the introduction of possible confounds, researchers may wish to control their stimuli in terms of stress. Compare, for instance, English *pilot* ['pʌɪlət] and German *Pilot* [pi'lo:t]. Similarly, cognates may exhibit a different number of syllables across languages, with or without differences in stress placement, which also impacts word recognition (cf. Fritz et al., 2023): e.g., E *giraffe* [dʒɪ.'ɑ:f] – G *Giraffe* [gi.'ʁa.fə]; E *method* ['mɛ.θəd] – G *Methode* [me.'to:.də].

When assessing phonological similarity for a study on auditory processing, one should also take into account theories of L1 auditory word recognition, most prominently the cohort model (e.g., Taft & Hambly, 1986; Marslen-Wilson, 1987).³ Given that auditory processing is linear, word onsets play a crucial role in the activation of the cohort. This fact has often been neglected in auditory word recognition studies, as the above-discussed algorithms do not take into account where phonological overlap occurs within a word. One way of explicitly testing the coactivation of the L1 cohort in L2 word recognition is by employing a cross-modal fragment priming paradigm. In this paradigm, participants hear a fragment cut out from a word, followed by a target word appearing on the screen (usually combined with a lexical decision task). Using only initial word fragments (e.g., the initial

syllable of a disyllabic word, such as ['rɛp] taken from reptile ['rɛptail]), we can control the degree of phonological overlap much more carefully (as additional differences may occur later in the word). This paradigm allows us to test, for example, if the stressed English syllable ['rɛp] also activates the German counterpart which carries stress on the final syllable (*Reptil* [rɛp'ti:l]; cf. Fritz et al., 2023).

With our comparison of algorithms, we have aimed to illustrate not only the phonological levels with which they concern themselves, but also those phonological levels which the various approaches *do not* take into account. Researchers may, however, wish to control for these dimensions (stress and syllable structure) during stimuli selection, *before quantifying phonological overlap*.

4. Linguistic cognates and shared loans

Thus far, our focus has been on methods of assessing cross-linguistic phonological similarity, particularly as a means of quantifying differences between cognates and noncognates or between cognates with higher and lower degrees of phonological overlap. However, despite the many useful applications for similarity algorithms already discussed (particularly in controlling degree of form overlap), it is important to recognise their limitations. The measures provided by such algorithms can be somewhat coarse or generalised, unable to capture differences both above and below the level of the phoneme, and result in a levelling of multiple phonological dimensions. If our interest is to explore which kinds of similarities do (or do not) matter for lexical access in bilingual processing, insights from the fields of phonology and historical linguistics can prove very valuable. This applies if the aim is to explore more fine-grained issues of L2 phonological mental representations. However, a better understanding of the phonologies of the languages being investigated will inevitably be helpful in studying (auditory) word recognition, even if not directly related to phonology.

Fundamentally, languages are structured systems and historical change is both guided and constrained by (i) cognitive processes involved in language processing and (ii) the structure of the underlying grammar. Sound change is therefore essentially phonologically coherent (i.e., it is neither random nor idiosyncratic). The crucial point is that, although cognates may diverge in form across languages over time, differences are *systematic*. Consequently, for two related languages, subsets of the lexicon will typically exist which exhibit regular phonological correspondences across the two languages, due to historical sound change (which applies regularly in all qualifying contexts) having affected one language but not the other. For instance, all *inherited cognates* with initial [pf̩] in German will correspond to English [p], e.g., *Pfad* [pf̩a:t] – *path* [pɑ:θ].⁴

As sound change begins in the synchronic phonological grammar, differences in inherited cognates across languages will usually involve only a small number of phonological features (typically just one or two). Features can be understood as sub-phonemic units which define natural classes of sounds, i.e., the class of voiced stops [b, d, g] would be defined by the features [VOICE] and [PLOSIVE]. This is illustrated in Table 2, which provides examples of related language pairs where a given sound change has affected one language, resulting in classes of cognates which differ in only a single feature (which might have been changed, added or lost).

Table 2: Classes of corresponding initial consonants in cognates across a range of languages, with the relevant differentiating phonological features indicated.

/s/–/h/ [OBSTRUENT], [CONTINUANT]		/w/–/v/ [CONTINUANT], [VOICE], [LABIAL]		/f/–/p/ [OBSTRUENT], [LABIAL]	
Latin	Ancient Greek	English	German	English	Italian
<i>sāl</i> /sa:l/	ἅλς /hals/	<i>wind</i> /wɪnd/	<i>Wind</i> /vɪnd/	<i>father</i> /fa:ðə/	<i>padre</i> /padre/
<i>sex</i> /seks/	ἕξ /heks/	<i>wax</i> /wæks/	<i>Wachs</i> /vaks/	<i>fish</i> /fɪʃ/	<i>pesce</i> /peʃ:e/
<i>septem</i> /septem/	ἑπτὰ /hepta/	<i>wolf</i> /wʊlf/	<i>Wolf</i> /vɔlf/	<i>ford</i> /fɔ:d/	<i>porto</i> /pɔrto/
/s/ = [CORONAL] /h/ has no PLACE feature		/w/ = [SONORANT] /v/ = [OBSTRUENT]		/f/ = [CONTINUANT] /p/ = [PLOSIVE]	

When sub-classes of the lexicon share these phonological relationships, this may have an impact on lexical access for bilingual speakers. Such patterns may be exploited in experimental design in order to explore in more subtle ways the nature of mental representations and the processes underlying bilingual lexical activation, particularly in contexts where algorithms provide too coarse a measure of phonological similarity.

This point is illustrated by a recent EEG experiment by Booth and colleagues (2026), which demonstrates the ways in which experimental design might be approached using the systematicity of cross-language correspondences. The experiment, employing a cross-modal fragment priming paradigm, investigated the extent to which German–English bilingual speakers’ L1 representations were co-activated in the processing of L2 lexical items and compared noncognates (e.g., *meat* [mi:t] – *Fleisch* [flaɪʃ]) with historical cognates of three different types: (i) cognates with “identical” word-initial sequences (e.g., *nest* [nɛst] – *Nest* [nɛst]), (ii) cognates where the onset was a fricative in English and a plosive in German (e.g., *thing* [θɪŋ] – *Ding* [dɪŋ]) and (iii) cognates where the onset was a plosive in English and an affricate in German (e.g., *pan* [pæn] – *Pfanne* [pfanə]). For the non-identical cognates, the

relevant consonants share place features and only differ in a single phonological feature relating to manner (as a result of sound change). As the mental representation of plosives is considered to be underspecified for CONstriction (cf. Lahiri & Reetz, 2002, 2010; Lahiri, 2018), the English fricative was expected to partially co-activate the plosive-initial German cognate, but the English plosive was not expected to coactivate the mismatching German affricate. This was borne out in the results, with the [θɪŋ] – [dɪŋ] condition behaving like near cognates and resulting in no attenuation of the N400 component (in line with the identical cognates, interpreted as reduced facilitation due to L1 competition). In contrast, attenuation of the N400 component was found for the mismatching [pæn] – [pfanə] condition, which thus behaved as though no cognate existed, patterning with the noncognate condition (e.g., [mi:t] – [flaɪ], where the N400 was also attenuated and priming effects were greatest).

The key point here is that, although the word-initial consonants in both the fricative–plosive and plosive–affricate conditions only differ in a single phonological feature, only the former resulted in cross-language co-activation, providing further experimental evidence that not all features are necessarily fully specified in mental representations at all levels of phonological representation (see also Lahiri & Marslen-Wilson, 1992; Friedrich et al., 2008; Cornell et al., 2013; Althaus et al., 2024). Even more strikingly, algorithms would not necessarily differentiate between the two classes of non-identical cognates (their ALINE scores are identical) or might even indicate that the [pæn] – [pfanə] condition's cognates show greater form overlap, particularly in the case of the NLD measure (and to a lesser degree, PHOR-in-One). Nor would these cognates necessarily qualify as cognates at all, with NLD scores below the permissive 0.5 cut-off point (and PHOR-in-One scores below 0.7).

This is illustrated in Table 3, which compares the similarity scores for Booth and colleagues’ (2026) experimental conditions when applying the various algorithms.

Table 3: Similarity scores for the different experimental conditions in Booth et al. (2026) across different algorithms.

	Matching (<i>nest–Nest</i>)	Non-mismatching (<i>thing–Ding</i>)	Mismatching (<i>pan–Pfanne</i>)	Noncognates (<i>meat–Fleisch</i>)
Phonological NLD	0.70	0.36	0.48	0.09
PHOR-in-One	0.83	0.60	0.69	0.47
ALINE	0.93	0.81	0.81	0.51

Importantly, it is not only phonemes or features which may mismatch in perception and play a role in cognate processing and it is thus just as important to consider phonological structure above the level of the phoneme as below it. Again, systematic correspondences are not restricted to the segmental or featural level, but may also affect prosodic structure. For instance, monomorphemic Germanic words will be invariably short and initially stressed, as in Dutch *wéer* [vɛ:r], English *weather* [ˈwɛðə], German *Wétter* [ˈvɛtɐ], Danish *véjr* [vɛʔɐ], Norwegian *vær* [væ:r], Swedish *väder* [ˈvɛ:dɛɪ]. In contrast, shared loans from Romance are typically longer and stress placement may differ across languages, e.g., English *colony* [ˈkɒləni], Dutch *kolónie* [koˈlo:ni], German *Kolonie* [koloˈni:], Danish/Norwegian/Swedish *koloní* [koloˈni², koloˈni:, kɒlɔˈni:] (cf. Booth & Lahiri 2023). Thus, any differences in word stress across Germanic languages will be restricted to loanwords and experimental stimuli would have to be drawn from this class.

This insight formed the basis of a recent EEG study by Fritz and colleagues (2023), who demonstrated that, for shared loans (e.g., English *signal* [ˈsɪgnəl] – German *Signal* [zɪgˈna:l]), the existence of a corresponding word and prosodic differences in metrical structure and stress placement affected the ease of lexical access and had significant

consequences for lexical activation and integration. Where loans only existed in the L2 (in this case English), attenuation of the N400 component was found and reaction times were fastest. However, this was only the case when the metrical structure of the English word did not violate German metrical structure rules, e.g., *pigeon* ['pɪdʒɪn] (in contrast to *curfew* ['kɜ:.fju:], where the English stress placement is in conflict with German parameters, which would predict final stress). Again, these results suggest interference from the L1, as words existing in both languages (shared loans, e.g., *reptile* ['rɛptɪl] – G *Reptil* [ʁɛp'ti:l]) also result in reduced facilitation, interpreted as resulting from increased competition for selection. Note that the words used in this experiment were necessarily all loans, as native words (i.e., true cognates) *never* differ in stress placement in Germanic. Differences in stress are restricted to a subset of the lexicon which consists exclusively of longer loanwords.

Most of the examples provided thus far draw on closely related (Germanic) languages. For more distantly related languages, cognates' phonological representations may have diverged to such an extent that modern speakers would not be expected to link the two in their mental lexicons. Compare, for example, English *hundred* [hʌndrɪd] and German *Hundert* [hʊndɐt] with Russian сто [sto], Italian *cento* [tʃɛnto], Greek *εκατό* [ekato], Bengali শত [ʃɔtɔ], Persian صد [sæd]. This does, however, depend on the individual sounds involved, as phonemes with certain features have never changed. For instance, unlike plosive consonants, (fully specified and sonorous), Proto-Indo-European nasals have been much more resistant to change, e.g., E *nose* [nəʊz], G *Nase* ['na:zə], Russian нос [nos], Lithuanian

nósis ['no:s'is] Latin *nāsus* ['na:sus], Italian *naso* ['nasɔ], Sanskrit नस् [nas], Bengali নাক [na:k].

In such languages, or even totally unrelated languages, recent loanwords (either shared loans from a third language or from one language into the other) are much more likely to demonstrate phonological form-overlap (cf. English *radio* and Russian радио [radʲio]). Such words can therefore offer similar opportunities to cognates, as language contact and large-scale lexical borrowing from one language into another may provide an additional source of both phonemic and prosodic correspondences (although the nature of the correspondences will often be different to that of sound change in historical cognates). Novel phonemes are rarely (if ever) borrowed, especially where the featural contrast does not exist elsewhere in the native system. Instead, loanwords are adapted according to the native contrasts, constrained by the structure of the L1 phonological system and the underlying features of the relevant phoneme. Again, this adaptation is utterly systematic (see Kennard & Lahiri 2020). For instance, English [t] is borrowed into Bengali as /t/ (and not /t̪/), e.g., টেবিল ['tebil] “table”, টিকিট ['tikit] “ticket”.

Similarly, many [ps]-initial Greek words (spelt ⟨ps⟩) have been borrowed into the Germanic languages. Although English disallows this sequence syllable initially (but not finally, e.g., *eclipse* [əklɪps]), [ps] is a permissible onset in German. As a result, English adapts such loans with [s], e.g., *psyche* [saɪki], producing [s] – [ps] correspondences between English and Greek, *but also* English and German (where these loans are shared): e.g., E *pseudonym* ['sjʊdənɪm] – G *Pseudonym* [psɔːdoˈnyːm]; E *psoriasis* [səˈlɪɪəsɪs] – G

Psoriasis [psɔ'ʁi:azis]; E *psychology* [saɪ'kɒlədʒi] – G *Psychologie* [psyçolo'gi:] (note that stress placement additionally differs in many of these words).

For non-related languages, such as English and Japanese (where inherited cognates do not exist), loanwords are possible (and common). Loans may be shared (i.e., both languages have independently borrowed a word from a third language) or one language might have borrowed a word from the other (cf. English *tsunami*, *sushi* and Japanese ドレス [doresi] “dress”, バイク [baiku] “motorbike”). In both cases, these loans are adapted to the host language in a systematic way, not only phonemically but also prosodically. For example, English loans often contain consonant clusters or coda consonants which are not permitted in Japanese. These loans are thus adapted to the CV syllabic structure of Japanese, invariably via the insertion of epenthetic vowels, e.g., [masiku] < English *mask* [mɑ:sk] (see Nakayama et al., 2012).

5. Implications for psycholinguistic studies

Drawing on insights from phonological theory, we advocate for a more systematic approach to phonology in the study of the bilingual mental lexicon, particularly in relation to research questions and study design. Modality-specific processing (i.e., visual vs. auditory) should also be considered when the object is a more careful investigation of the role of phonology, as it is clear that cognates and loanwords may differ along a number of dimensions, particularly orthographic and phonological form (and these two need not necessarily go hand in hand).

In the vast majority of past visual studies, the interest has been in orthographic overlap and visual word recognition. Increasingly, however, attention is being turned to the role of phonology, either directly (in auditory studies) or indirectly (i.e., phonological activation via reading in visual studies, e.g., Tiffin-Richards, 2024). This being the case, it is

increasingly important to be clear about the object of inquiry. If our interest is in orthographic overlap, it is critically important to control for phonological representations as well (in order to avoid the introduction of unnecessary confounds). If, however, the object is to explore phonological representations specifically (via either the visual or auditory modality), then a more systematic approach may be beneficial, as outlined in this paper.

Algorithms providing numerical similarity scores can be usefully employed by researchers and are capable of providing good overall indications of degree of overlap (particularly the weighted approaches, such as ALINE and PHOR-in-One). However, even before the point of applying these scores, it is important to reflect critically on the stimuli which are being fed in, particularly in light of variation which the algorithms are unable to capture. For instance, researchers may want to control for prosodic issues, such as stress placement and number of syllables, as well as considering where in the words' linear string variation arises (i.e., word initially, or towards the right edge, where variation is less pivotal in terms of word recognition, cf. Marslen-Wilson 1987). Finally, decisions must be made as to the nature of the transcriptions used, whether phonemic or more phonetic (and transparency is important here, cf. Roembke et al. 2024).

Throughout this paper, we have sought to highlight the systematicity of phonological grammars by providing examples across different phonological levels (featural, phonemic and metrical), stressing too that not all differences are equal from a processing perspective. Much of the focus has been on Germanic, but the same principle also holds for other language families, such as Romance, which will have their own systematic correspondences. For instance, if Italian has initial $/tʃ/$, cognates will likely have initial $/θ/$ in Spanish, $/s/$ in French and $/tʃ/$ in Romanian: e.g., Italian *città* $/tʃit̪a/$ “city”, *cielo* $/tʃɛlo/$ “sky”; Spanish *ciudad* $/θjuˈdad/$, *cielo* $/θjelo/$; French *cit * $/site/$, *ciel* $/sjel/$; Romanian *cetate* $/tʃeta.te/$,

cer /tʃer/. Such systematicity can be exploited in experimental design in psycholinguistic studies investigating the bilingual mental lexicon. Phonological knowledge and historical analyses can therefore provide a rich literature for psycholinguistics working on bilingual word recognition.

Advocating for grounding experimental designs in linguistic theory is not new. Pylkkänen et al. (2011), for example, advocate for a consideration of theoretical semantics when investigating the neuroscientific mechanisms underlying semantic processing. Cognates do not have any *a priori* special status in the lexicon (there is no [+cognate] or [+loan] feature stored in underlying representations); indeed, some (older) loans might pattern with cognates and show similar correspondences (e.g., G *Pflaster* [pflastə] “plaster”, *Pflanze* [pflantsə] “plant” < Latin *plastrum*, *planta*), whilst other cognates might not have diverged so much in form, thus being in one sense more loan-like (e.g., E *nest* [nɛst], *hand* [hænd] – G *Nest* [nɛst], *Hand* [hant]). However, due to their history, cognates are likely to demonstrate special structural characteristics which will be true of the whole class of cognates. It is these structural patterns which matter for modern speakers rather than the historical development of the words *per se*. As these differences are phonemic, defined in terms of phonological features, incorporating these factors into study design enables us to tease apart these issues and explore the nature of lexical representations and phonological activation in a structured and theoretically-informed way. Similarly, structural characteristics of (shared) loanwords can be exploited, as in the case of metrical stress.

Drawing on insights from phonological work, it is possible to look beyond orthographic overlap and think differently about the ways in which words might be similar (or indeed dissimilar) across languages. When investigating word recognition, it is

worthwhile to adopt a consistent approach to similarity and differences, considering not only phonological overlap, but also correspondences in terms of features and prosody (including syllable structure and stress) *in addition to* orthography. We argue for the importance of approaching languages as complete systems, characterised by systematic patterns and correspondences. Such correspondences may go across languages and impact processing and lexical access in an L2 context.

Data availability: Data and analyses scripts generated as part of this article can be found on OSF (<https://osf.io/fsyg9>).

References

- Althaus, N., Lahiri, A., & Plunkett, K. (2024). Coronal underspecification as an emerging property in the development of speech processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(12), 1932–1953. <https://doi.org/10.1037/xlm0001367>
- Arana, S. L., Oliveira, H. M., Fernandes, A. I., Soares, A. P., & Comesaña, M. (2022). The cognate facilitation effect depends on the presence of identical cognates. *Bilingualism: Language and Cognition*, 25(4), 660–678. <https://doi.org/10.1017/S1366728922000062>
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *CELEX2 LDC96L14*. Philadelphia, PA: Linguistic Data Consortium.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc. <https://www.nltk.org/index.html>
- Booth, J., Stich, F., Fritz, I., & Lahiri, A. (2026). Through *dick* and *dünn*: cognates and the legacy of diachronic change in synchronic language processing. *Language, Cognition and Neuroscience*, 41(7), 972–996. <https://doi.org/10.1080/23273798.2026.2683473>
- Booth, J., & Lahiri A., (2023). 'Foot Structure in Germanic'. *Oxford Research Encyclopedia of Linguistics*. 20th September 2023. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.962>
- Cai, Z. G., Pickering, M. J., Yan, H., & Branigan, H. P. (2011). Lexical and syntactic representations in closely related languages: Evidence from Cantonese–Mandarin bilinguals. *Journal of Memory and Language*, 65(4), 431–445. <https://doi.org/10.1016/j.jml.2011.05.003>
- Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 757–772. <https://doi.org/10.1037/a0030862>

- Cornut, C., Mahé, G., & Casalis, S. (2022). L2 word recognition in French–English late bilinguals: Does modality matter?. *Bilingualism: Language and Cognition*, 25(1), 121–136.
<https://doi.org/10.1017/S1366728921000511>
- Costa, A. S., Comesaña, M., & Soares, A. P. (2023). PHOR-in-One: A multilingual lexical database with PHonological, ORthographic and PHonographic word similarity estimates in four languages. *Behavior Research Methods*, 55(7), 3699–3725. <https://doi.org/10.3758/s13428-022-01985-3>
- Degani, T., Prior, A., & Hajajra, W. (2018). Cross-language semantic influences in different script bilinguals. *Bilingualism: Language and Cognition*, 21(4), 782–804.
<https://doi.org/10.1017/S1366728917000311>
- Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. *Journal of Memory and Language*, 41, 496–518.
<https://doi.org/10.1006/jmla.1999.2654>
- Dijkstra, T., & van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175–197.
<https://doi.org/10.1017/S1366728902003012>
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62(3), 284–301. <https://doi.org/10.1016/j.jml.2009.12.003>
- Dijkstra, T., Buytenhuijs, A. W. F., van Halem, N., al-Jibouri, Z., de Korte, M., & Rekké, S. (2019). Multilink: A Computational Model for Bilingual Word Recognition and Word Translation. *Bilingualism: Language and Cognition*, 22(4), 657–679.
<https://doi.org/10.1017/S1366728918000287>
- Downey, S. S., Hallmark, B., Cox, M. P., Norquest, P., & Lansing, J. S. (2008). Computational feature-sensitive reconstruction of language relationships: Developing the aline distance for comparative historical linguistic reconstruction. *Journal of Quantitative Linguistics*, 15(4), 340–369. <https://doi.org/10.1080/09296170802326681>

- Elias, M., van Hell, J. G., Prior, A., & Degani, T. (2025). Cognate facilitation in different-script trilinguals as a function of task demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 51(12), 2009–2032. <https://doi.org/10.1037/xlm0001442>
- Elias, M., & Degani, T. (2022). Cross-language interactions during novel word learning: The contribution of form similarity and participant characteristics. *Bilingualism: Language and Cognition*, 25(4), 548–565. <https://doi.org/10.1017/S1366728921000857>
- Frances, C., Navarra-Barindelli, E., & Martin, C.D. (2021). Inhibitory and facilitatory effects of phonological and orthographic similarity on L2 word recognition across modalities in bilinguals. *Scientific Report*, 11, 12812. <https://doi.org/10.1038/s41598-021-92259-z>
- Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1545–1559. <https://doi.org/10.1037/a0012481>
- Fritz, I., & Lahiri, A. (2025). Native prosodic structures constrain L2 word recognition: Evidence from Bengali-English bilinguals. *Brain and Language* 264: 105553. <https://doi.org/10.1016/j.bandl.2025.105553>
- Fritz, I., Lahiri, A., & Kotzor, S. (2023). Shared Loanword Recognition in German–English Bilinguals: The Role of Metrical Phonology. *Journal of Cognitive Neuroscience*, 35(12), 2028–2048. https://doi.org/10.1162/jocn_a_02063
- Guediche, S., Baart, M., & Samuel, A. G. (2020). Semantic priming effects can be modulated by crosslinguistic interactions during second-language auditory word recognition. *Bilingualism: Language and Cognition*, 23(5), 1082–1092. <https://doi.org/10.1017/S1366728920000164>
- de Groot, A. M., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30(1): 90–123. [https://doi.org/10.1016/0749-596X\(91\)90012-9](https://doi.org/10.1016/0749-596X(91)90012-9)

Kennard, H. J., & Lahiri, A. (2020). Nonesuch phonemes in loanwords. *Linguistics*, 58(1), 83-108.

<https://doi.org/10.1515/ling-2019-0033>

Kondrak, G. (2002). *Algorithms for language reconstruction*. Unpublished PhD thesis. University of

Toronto. <http://hdl.handle.net/1807/121117>

Kondrak, G. (2003). Phonetic Alignment and Similarity. *Computers and the Humanities*, 37(3), 273-

291. <https://doi.org/10.1023/A:1025071200644>

Ladefoged, Peter. 1975. *A Course in Phonetics*. 2nd edn. New York: Harcourt Brace Jovanovich.

Ladefoged, Peter & Keith Johnson. (2015). *A Course in Phonetics*. 7th edn. Stamford (CT): Cengage Learning.

Lahiri, A., & Marslen-Wilson, W. D. (1992). Lexical processing and phonological representation. In:

R. D. Ladd & G. Docherty (Eds.), *Papers in Laboratory Phonology II* (pp. 229–260). Cambridge University Press. <https://doi.org/10.1017/CBO9780511519918.010>

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, & N. Warner

(Eds.), *Laboratory Phonology 7* (pp. 637–676). De Gruyter Mouton.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation

and processing. *Journal of Phonetics*, 38(1), 44-59. <https://doi.org/10.1016/j.wocn.2010.01.002>

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals.

Soviet Physics – Doklady, 10, 707–710.

Lijewska, A. (2020). Cognate Processing Effects in Bilingual Lexical Access. In *Bilingual Lexical*

Ambiguity Resolution (pp. 71-95): Cambridge University Press.

<https://doi.org/10.1017/9781316535967.005>

Marslen-Wilson, William D. (1987). Functional parallelism in spoken word-recognition. *Cognition*,

25, 71–102. [https://doi.org/10.1016/0010-0277\(87\)90005-9](https://doi.org/10.1016/0010-0277(87)90005-9)

- Muntendam, A., van Rijswijk, R., Severijnen, G., & Dijkstra, T. (2022). The role of stress position in bilingual auditory word recognition: Cognate processing in Turkish and Dutch. *Bilingualism: Language and Cognition*, 25(4), 679-690. <https://doi.org/10.1017/S1366728922000037>
- Muegge, J. B., Kim, H., & McMurray, B. (2025). Decoupling speech processing from time. *Trends in Cognitive Sciences*, 30, 66-79. <https://doi.org/10.1016/j.tics.2025.05.017>
- Muylle, M., Van Assche, E., & Hartsuiker, R. J. (2022). Comparing the cognate effect in spoken and written second language word production. *Bilingualism: Language and Cognition*, 25(1), 93–107. <https://doi.org/10.1017/S1366728921000444>
- Nakayama, M., Sears, C. R., Hino, Y., & Lupker, S. J. (2012). Cross-script phonological priming for Japanese-English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes*, 27(10), 1563-1583. <https://doi.org/10.1080/01690965.2011.606669>
- Pylkkänen, L., Brennan, J., & Bemis, D. K. (2011). Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes*, 26(9), 1317-1337. <https://doi.org/10.1080/01690965.2010.527490>
- Roembke, T. C., Koch, I., & Philipp, A. M. (2024). What makes a cognate? Implications for research on bilingualism. *Bilingualism: Language and Cognition*, 27(5), 1023-1028. <https://doi.org/10.1017/S1366728924000233>
- Sánchez-Casas, R., & García-Albea, J. E. (2005). The Representation of Cognate and Noncognate Words in Bilingual Memory: Can Cognate Status be Characterized as a Special Kind of Morphological Relation? In J. F. Kroll & A. M. B. D. Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 226-250): Oxford University Press.
- Spivey, M. J., & Marian, V. (1999) Cross Talk Between Native and Second Languages: Partial Activation of an Irrelevant Lexicon. *Psychological Science*, 10(3), 281–284. <https://doi.org/10.1111/1467-9280.00151>

- Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, 22(3), 259-282. [https://doi.org/10.1016/0010-0277\(86\)90017-X](https://doi.org/10.1016/0010-0277(86)90017-X)
- Tiffin-Richards, S. P. (2024). Cognate facilitation in bilingual reading: The influence of orthographic and phonological similarity on lexical decisions and eye-movements. *Bilingualism: Language and Cognition*, 27(5), 964-981. <https://doi.org/10.1017/S1366728923000949>
- Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, 64(1), 88-107. <https://doi.org/10.1016/j.jml.2010.08.006>
- van der Loo, M. P. J. (2014). The stringdist package for approximate string matching. *The R Journal*, 6(1), 111–122. <https://doi.org/10.32614/rj-2014-011>
- Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, 15(3), 181–198. <https://doi.org/10.3758/BF03197716>
- Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition*, 35(5), 938-952. <https://doi.org/10.3758/BF03193467>
- Zhang, N., Wu, Z., & Wang, M. (2025). Cross-language phonological activation in bilingual visual word recognition: A meta-analysis. *Psychonomic Bulletin & Review*, 32, 2059-2089. <https://doi.org/10.3758/s13423-025-02692-8>

¹ Throughout this paper, unless otherwise specified, Standard Southern British English transcriptions are used (rather than US English or any other variety).

² In the field of phonology, multiple feature systems have been proposed. For the purposes of the present paper, we can remain agnostic on the subtleties of these differences. For an overview, see Lahiri & Reetz (2010).

³ An alternative perspective on word recognition, which places less emphasis on the linearity of the acoustic signal has also recently been proposed. For an overview of such accounts, see, for example, Muegge et al. (2025).

⁴ However, more recent loans with initial [p] in German will *also* correspond to English /p/ (initial /p/ only occurs in loans in German), e.g., *Palast* [pa'last] – *palace* ['pæ.ləs] (note also the differing stress placement in this case). Importantly, the relationship is therefore asymmetric with respect to language, as the two German sounds both correspond to English [p] but English initial [p] can correspond to either [p] or [pf] in German (i.e., German to English correspondences are totally predictable, whereas English to German are not predictable for naïve speakers without etymological knowledge).