

Cognate sound correspondences and cross-linguistic coactivation

Joshua Booth, Isabella Fritz & Aditi Lahiri
Language and Brain Laboratory, University of Oxford



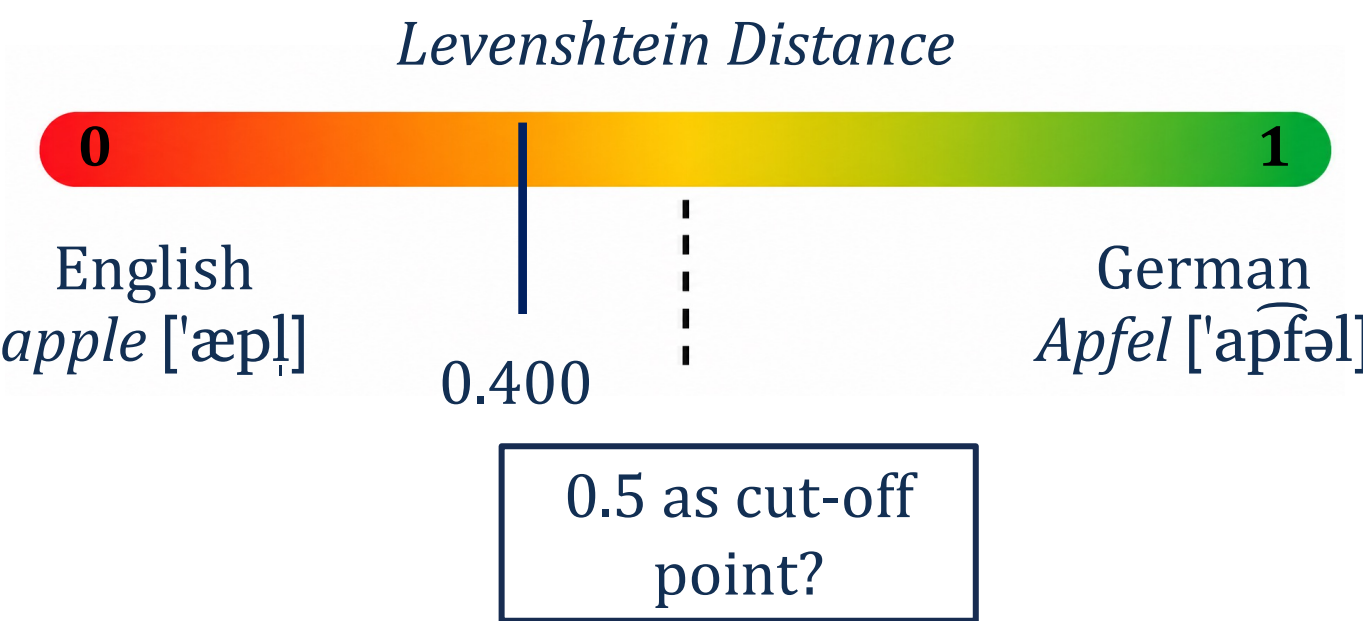
Defining cross-linguistic similarity

In psycholinguistics, **COGNATES** are defined as translation equivalents that overlap in meaning, phonology and orthography (for same script languages).

How can we define cognate similarity?

1. Algorithms

- Levenshtein Distance (Levenshtein, 1966)
- PHOR-in-One (Costa et al., 2023)
- ALINE (Kondrak, 2003)



2. Bilingual ratings

Focusing on bilinguals' intuition about phonological similarity

- These approaches conflate all phonological similarities into a **single score**.
- With different **algorithms** and **cut-off points**, the same word can potentially be categorised differently across studies, i.e. cognate vs. non-cognate.

Cross-linguistic correspondences

Cognates in linguistics: inherited words from a common ancestor. Certain classes differ systematically in their initial consonants across related languages.

/s/-/h/		/w/-/v/		/f/-/p/	
Latin	Ancient Greek	English	German	English	Italian
<i>sāl</i> /sa:l/	ἅλς /hals/	<i>wind</i> /wind/	<i>Wind</i> /vɪnd/	<i>father</i> /fɑ: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/

Shared loans (Fritz et al., 2023): Loans across languages coming from the same source. Examples are Romance loans in Germanic languages with differing stress pattern.

Dutch	German	English	Word stress
balláde	Balláde	bállad	D=G≠E
signáal	Signál	sígnal	
talént	Talént	tálent	

Languages are structured systems, phonological cross-linguistic variation is systematic:

- Historical sound change produces systematic correspondences in cognates.
- Phonological adaption of loanwords also results in systematic correspondences.

These patterns can be exploited in experimental studies investigating questions of bilingual word recognition.

Comparison of algorithms

Stimuli taken from Tiffin-Richards (2024) who investigated the role of phonological and orthographic overlap in English-German cognate processing.

Example stimuli (using the transcription provided in the PHOR-in-One)

Identical cognates E: *element* [ɛlɪmənt] – G: *Element* [element]
Close cognates E: *camel* [kæmətʃ] – G: *Kamel* [kame:l]
Non-cognate E: *chair* [tʃeə] – G: *Stuhl* [ʃtu:l]

Condition	Phon NLD ¹	NLD ²	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)

<i>circus</i> [sɜ:kəs]	—	0.286	0.815	0.904
<i>Zirkus</i> [tsɪrəkʊs]				
<i>museum</i> [mju:ziəm]	—	0.500	0.567	0.840
<i>Museum</i> [mu:ze:ʊm]				

1. Phonological Normalised Levenshtein Distance (as in Tiffin-Richards); 2. NLD (based on PHOR-in-One transcription), 0 = maximally distant ; 1 = identical

An example: Booth et al. (2026)

Takes systematic correspondences in **historical cognates** as starting point and defined cognate similarity in **phonological** terms.

Investigated the extent to which German-English bilingual speakers' L1 representations were co-activated in the processing of L2 lexical items.

Cross-modal fragment priming paradigm, comparing **noncognates** (*meat* [mi:t] – *Fleisch* [flaɪʃ]) with **historical cognates of three different types**:

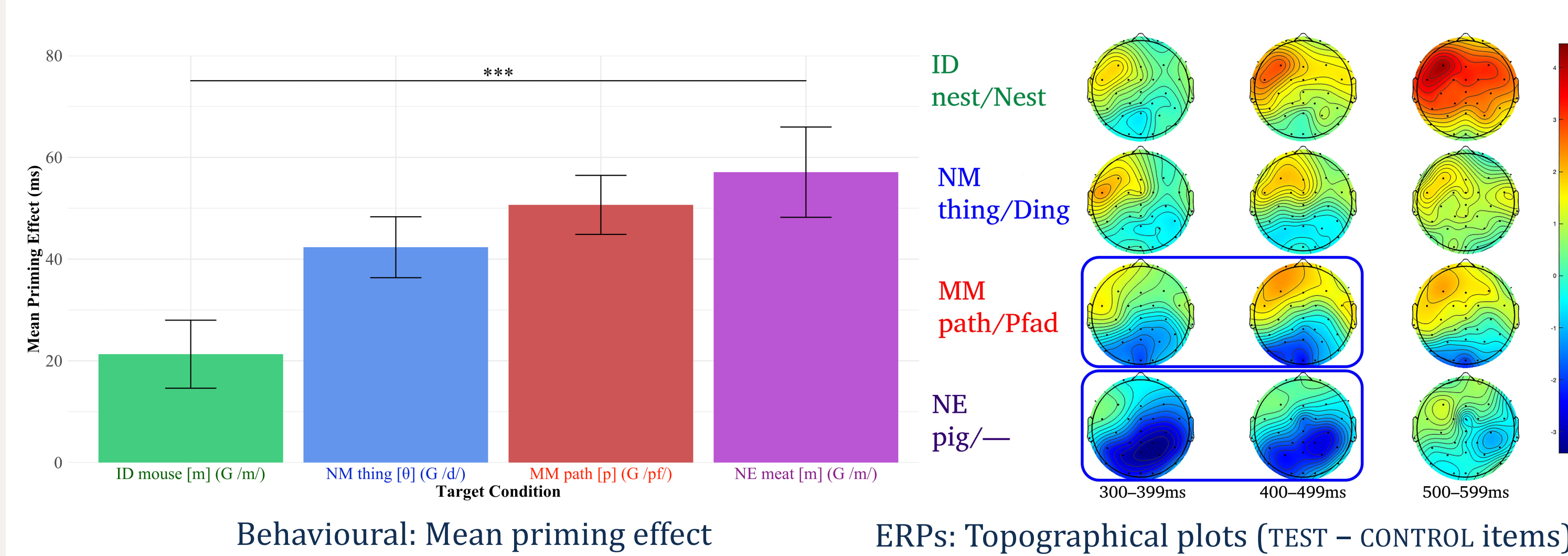
- “identical”** word-initial sequences: *nest* [nɛst] – *Nest* [nɛst]
- initial **fricative** (E) vs. **plosive** (G): *thing* [θɪŋ] – *Ding* [dɪŋ]
- initial **plosive** (E) vs. **affricate** (G): *pan* [pæn] – *Pfanne* [pʰanə]

Phonologically, for the non-identical cognates (**ii-iii**), the relevant consonants share **place** features and only differ in a single **manner** feature.

Phonological theory suggests the mental representation of plosives is **underspecified** for CONstriction (cf. Lahiri & Reetz, 2010; Lahiri, 2018)

→ This predicts the English **fricative** would partially co-activate the **plosive-initial** German cognate, but the English **plosive** was not expected to coactivate the mismatching German **affricate**.

Results



	English	German	Priming Effect	N400
(i) IDENTITY	<i>fish</i> , <i>mouse</i>	→ <i>Fisch</i> , <i>Maus</i>	21.3ms	—
(ii) NO-MISMATCH	<i>thorn</i> , <i>thing</i>	→ <i>Dorn</i> , <i>Ding</i>	42.3ms	—
(iii) MISMATCH	<i>path</i> , <i>tin</i>	→ <i>Pfad</i> , <i>Zinn</i>	50.7ms	attenuated
(iv) NON-EXISTENT	<i>pig</i> , <i>thug</i>	—	57.1ms	attenuated

Algorithm scores by condition

Although word-initial consonants only differ in a single phonological feature across the non-identical cognate conditions, **phonological representations matter**: only **non-mismatching** resulted in cross-language co-activation.

- BUT**: algorithms don't necessarily differentiate these two types of cognate (or might even indicate that the [pæn] – [pʰanə] condition's cognates show greater form overlap):

	Matching (<i>nest–Nest</i>)	Non-mismatching (<i>thing–Ding</i>)	Mismatching (<i>pan–Pfanne</i>)	Noncognates (<i>meat–Fleisch</i>)
Phonological NLD	0.7	0.36	0.48	0.09
PHOR-in-One	0.83	0.6	0.69	0.47
ALINE	0.93	0.81	0.81	0.51

- Nor would they necessarily qualify as cognates at all, even with a permissive 0.5 (or 0.7) cut-off point.

Discussion

- The N400 was only significantly attenuated for the **mismatching** (*pan–Pfanne*) and **non-existent** (*meat–Fleisch*) conditions. These English words were therefore **easier to access** than the **identity** (*nest–Nest*) and **non-mismatching** (i.e. tolerated *thing–Ding*) conditions, where priming effects were also greatest.
 - **Similarities in form and meaning across languages do not equate to faster processing (quite the reverse!).**
- Cognates are characterised by **systematic sound correspondences**, typically involving differences in a **small number of features**.
- Activation depends on the abstract phonological representation and matching/mismatching of features**:
 - Provided the phonological representations don't mismatch and diverge too significantly, e.g. *twig* [twɪɡ] – *Zweig* [tʃvaɪk], L2 processing may lead to coactivation of L1 items in the cohort, **increasing competition for selection and hindering lexical access**.

Suggestions

Phonological considerations which may be taken into account and controlled for before running an algorithm:

- Stress placement and syllable count in word-pairs (E: *olive* – G: *Olíve*)
- Where does phonological overlap occur? (cf. COHORT model; Marslen-Wilson, 1987)
- Use of phonetic vs. phonemic transcription. Decisions on how detailed the transcription should be (e.g., *paper*: [pʰeɪpə] vs. /peɪpə/, *ball* [bɔ:ɫ] vs. /bɔ:l/)

References

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>
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.
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.
Lahiri, A. (2018). Predicting universal phonological contrasts. In L. M. Hyman, & F. Plank (Eds.), *Phonological Typology* (pp. 229–272). Mouton de Gruyter.
Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38(1), 44–59.
Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics – Doklady*, 10, 707–710.
Kondrak, G. (2003). Phonetic Alignment and Similarity. *Computers and the Humanities*, 37(3), 273–291.
Marslen-Wilson, William D. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71–102.
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.