COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 48 (2024) e13418

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13418

This article is part of the "Progress & Puzzles of Cognitive Science" letter series.

Cognitive Science From the Perspective of Linguistic Diversity

Yoolim Kim,[a,#] Annika Tjuka[b,#]

[a]*Department of Psychology, Cognitive & Linguistic Sciences Program, Wellesley College*
[b]*Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology*

Received 11 October 2023; received in revised form 27 December 2023; accepted 11 February 2024

Abstract

This letter addresses two issues in language research that are important to cognitive science: the comparability of word meanings across languages and the neglect of an integrated approach to writing systems. The first issue challenges generativist claims by emphasizing the importance of comparability of data, drawing on typologists' findings about different languages. The second issue addresses the exclusion of diverse writing systems from linguistic investigation and argues for a more extensive study of their effects on language and cognition. We argue for a refocusing of cognitive science research on linguistic diversity in all modalities to develop the most robust understanding of language and its role in human cognition more broadly.

Keywords: Lexicon; Writing systems; Comparability; Modality

[#]Shared authorship.
Correspondence should be sent to Yoolim Kim, Cognitive & Linguistic Sciences Program, Wellesley College, 106 Central Street, Wellesley, MA, 02481, United States; and Annika Tjuka, Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. E-mail: annika_tjuka@eva.mpg.de & ykim6@wellesley.edu

1. Introduction

In this letter, we bring to light two, often-overlooked, issues in language research that are consequential for cognitive science as a field: (1) a lack of comparability of word meanings across languages, and (2) a lack of an integrated approach to examining language and its writing systems. First, comparability of data and modality are crucial as increasing evidence from typologists documenting diverse languages bears exception to generativist claims on language universals. Second, writing systems and scripts are almost entirely excluded from linguistic inquiry, or often studied from a cultural perspective focusing on its evolution as constrained by cognitive biases, but with limited reference to the languages they represent.

These two issues are related. Many of the theoretical constructs, such as "word," are artifacts of research constrained by comparing languages with writing systems. It is important to acknowledge that our understanding and perception of linguistic systems have been shaped and influenced by the structures brought on by writing systems. As a result, we frequently debate terms like "word" for their lack of robustness (Haspelmath, 2023; Mansfield, 2021; Tallman, 2020).

Comparability of word meanings and variation in writing systems are compounded by the predominant focus on English-speaking populations, also limiting generalizability to the larger population. Though recent research has shown increasing awareness of these shortcomings (Blasi, Henrich, Adamou, Kemmerer, & Majid, 2022; Henrich, Heine, & Norenzayan, 2010), we believe more can be done. We aim to recenter focus on linguistic diversity across all modalities in cognitive science so that our understanding of language and its status in broader human cognition is maximally robust.

2. Comparability of word meanings

The investigation of word meanings across languages spans various research fields, including anthropology and linguistics, and more recently, psychology and neuroscience, which highlights its interdisciplinary relevance to cognitive science. Language comparison has a long history in linguistics, with the earliest studies dating to Wilhelm von Humboldt who examined grammatical and lexical structures across languages (Humboldt, 1836). While Humboldt emphasized the individuality of languages, with the rise of generativism in linguistics, studies in language typology aimed to establish linguistic universals (Greenberg, 1963). Today, lexical typologists examine the vocabularies of the world's languages to find recurring patterns in semantic domains, such as kinship, color, or human body parts (Evans, 2010; Koptjevskaja-Tamm, 2008; Koptjevskaja-Tamm, Rakhilina, & Vanhove, 2015). There has been a growing emphasis on integrating the rich body of knowledge on language diversity, especially coming from anthropology (Levinson, 2012), into cognitive science, and as a result, the assumption of language universals has been challenged (Evans & Levinson, 2009). Research into domains previously considered well-structured has revealed a wealth of diverse possibilities, for example, how languages name human body parts (Majid, Enfield, & van

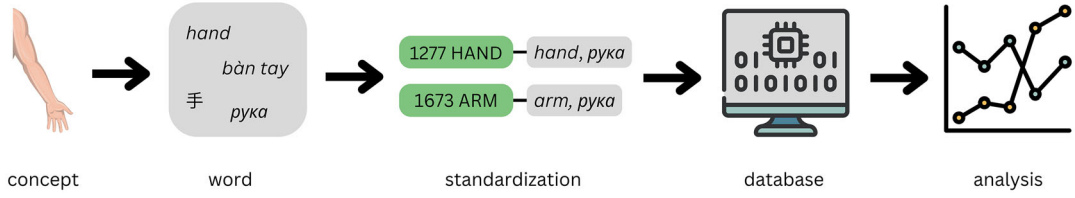


Fig. 1. Schematic data standardization process and outcome from the concept to a collection of words, the standardization of the data in the Cross-Linguistic Data Formats (Forkel et al., 2018), the creation of a database (Concepticon, List et al., 2016; Tjuka et al., 2023) and finally, the analysis of the data.

Staden, 2006). Insights from lexical typology have also gained importance in neuroscience (Kemmerer, 2019).

With the increase in lexical data from different languages and the advance of computer-assisted methods, one issue is often taken for granted: the comparability of words and their meanings across languages. English *tree* can be translated into *Baum* (German), *árbol* (Spanish), and *mù* 木 (Mandarin Chinese), and the meaning of a large wooden plant is considered cross-linguistically stable. In other cases, however, the translation is not as straightforward. Russian *рука* translates into two separate words in English: *hand* and *arm*, while Vietnamese uses a compound to distinguish between the two terms: *bàn tay* ("hand") and *tay* ("arm"). The words *arm* in Dutch and *ude* 腕 in Japanese add further complexities, as they yield different conceptual representations across speakers (Majid & van Staden, 2015). This variation is a challenge for every study using lexical data. Thus, standardization efforts for language comparison are essential.

The necessity for making cross-linguistic data comparable led to the development of the Cross-Linguistic Data Formats (Forkel et al., 2018) for linguistic research and beyond. The universally applicable format enables the construction of databases, which in turn facilitates the analysis of cross-linguistic patterns on a large scale. The use of standardized data is important for building cross-linguistic databases such as Concepticon (List, Cysouw, & Forkel, 2016) or CLICS (Rzymski et al., 2020) that researchers can use to construct models of language cognition (Brochhagen, Boleda, Gualdoni, & Xu, 2023; Gibson et al., 2019; Xu, Duong, Malt, Jiang, & Srinivasan, 2020). In the Concepticon, comparative concepts are established based on numerous typological studies to mitigate an English-centered view (Haspelmath, 2010; List, Cysouw, & Forkel, 2016), and unique identifiers such as 1277 HAND and 1673 ARM serve to represent the variation in the segmentation of the world into linguistic units (Tjuka, Forkel, & List, 2023). Fig. 1 illustrates the process of data standardization and its outcome. The Concepticon also offers the possibility to select comparative concepts for experimental studies. For example, Ambridge et al. (2020) used comparative verb forms for their study on the acquisition of sentence structure across languages.

Extending the standardization of lexical data to other research fields, that is, psychology, has proven successful and led to the creation of the Cross-Linguistic Database of Norms, Ratings, and Relations (Tjuka, Forkel, & List, 2022). In the future, the patterns emerging from

cross-linguistic data need to be tested with experimental methods. A protocol for standardization and integration of data is one way forward.

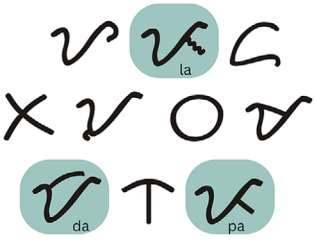
3. Variation in writing systems

Writing systems can vary according to numerous properties, especially in the linguistic information they encode. Previous literature has revealed significant effects of script (Smith, Monaghan, & Huettig, 2021); not only does it affect reading processes, but it also affects how linguistic information is represented in the mind. And yet, much of the literature on the mental lexicon is based on languages written alphabetically, linearly, and read left-to-right. The vast majority of the models from which we base our present understanding derive from research on English (Marslen-Wilson, Tyler, Waksler, & Older, 1994; Rastle, Davis, & New, 2004), German (Schriefers, 1992; Schuster & Lahiri, 2019), Dutch (Ernestus & Baayen, 2007; Zwitserlood, 1989), and other Western European or Indo-European languages.

Although Chinese (Zhou & Marslen-Wilson, 1994, 1995; Nakamura et al., 2012) and Arabic (Boudelaa & Marslen-Wilson, 2001) have received growing attention, languages that are mixed-script, have yet to gain traction. Korean, for example, uses both Hangul, an alpha-syllabary, and Hanja, Chinese logographic characters. However, research on language and writing systems remains divorced from one another. For anthropologists and psychologists, writing has been fruitful for investigating cultural forms as they evolve, giving insights into key cognitive biases and visual preferences (Morin, 2018; Morin, Kelly, & Winters, 2020; Miton & Morin, 2021). But interest lies firmly in the visual properties of the letter shapes themselves, and less so in the linguistic information they encode. Linguists, on the other hand, seldom take an interest in a language's writing system. A combined interest is often subsumed by research on orthographic processing examining issues in reading and spelling.

Across disciplines, researchers study a language's script from a visual perspective and its phonological inventory with great linguistic interest, but each in isolation of one another; however, how a language occupies both spaces may help us to better understand how spoken and written language relate to one another. Important questions regarding cognition arise once we begin to integrate currently disparate strands: If a language can be written using more than one script, how does it affect processing and its mental representation? Are the letter shapes themselves optimally constructed? How does a language's writing system make use of the space of possible shapes and phonological space? It is important to note that writing systems negotiate dynamic and often competing constraints of conveying sound and meaning, and that semantic contributions are not to be neglected in the discussion of writing and phonology. Such contributions are particularly notable in Chinese characters or Mayan glyphs as well as in the disambiguation of English homophones, for example, "pear" and "pair." Writing is a fundamental process by which the language is translated into a graphic medium; thus, the relationship between linguistic structure and the structure of writing systems requires a more thorough examination.

As cognitive scientists, we have yet to exploit and probe how exactly both systems interact, as well as the diversity in writing systems attested cross-linguistically. Each dynamically



Script inventory

		Labial	Alveolar	Palatal	Velar	Glottal
Plosive	Voiceless	p	t		k	ʔ
	Voiced	b	d		g	
Fricative		β	s			h
Nasal		m	n		ŋ	
Lateral			l			
Rhotic			r			
Approximant		w		j		

Phonological inventory

Writing

Speaking

Fig. 2. A cognitive theory to account for how information about a language's sound and writing systems will advance our understanding of how language is represented in the mind.

informs our representation of language in the mind and thus an equally dynamic account that pieces together knowledge across disciplines is needed to push our understanding. Fig. 2 visualizes this pipeline that we believe may help to diversify the next steps in research.

4. Conclusion

Our letter invites cognitive scientists to consider a cross-linguistic approach using both computer-assisted and psycholinguistic methods to better discern the robustness of language universals. We believe that linguistic diversity needs to be represented in a standardized format to make data comparable across different research fields. In addition, we invite cognitive scientists to consider the structural diversity of the world's writing systems, whether it is the specific information encapsulated by a single graphic mark, to the visual features that characterize the marks themselves. Each of these considerations presents important and unique nuances that unless taken into account render our theories and models static and one-dimensional.

References

Ambridge, B., Maitreyee, R., Tatsumi, T., Doherty, L., Zicherman, S., Pedro, P. M., Bannard, C., Samanta, S., McCauley, S., Arnon, I., Bekman, D., Efrati, A., Berman, R., Narasimhan, B., Sharma, D. M., Nair,

R. B., Fukumura, K., Campbell, S., Pye, C., Pixabaj, S. F. C., Pelíz, M. M., & Mendoza, M. J. (2020). The crosslinguistic acquisition of sentence structure: Computational modeling and grammaticality judgments from adult and child speakers of English, Japanese, Hindi, Hebrew and K'iche'. *Cognition*, *202*, 104310. https://doi.org/10.1016/j.cognition.2020.104310

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, *81*(1), 65–92. https://doi.org/10.1016/S0010-0277(01)00119-6

Brochhagen, T., Boleda, G., Gualdoni, E., & Xu, Y. (2023). From language development to language evolution: A unified view of human lexical creativity. *Science*, *381*(6656), 431–436. https://doi.org/10.1126/science.ade7981

Ernestus, M., & Baayen, H. (2007). Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes*, *22*(1), 1–24. https://doi.org/10.1080/01690960500268303

Evans, N. (2010). Semantic typology. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp. 504–533). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199281251.001.0001

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525x0999094X

Forkel, R., List, J.-M., Greenhill, S. J., Rzymski, C., Bank, S., Cysouw, M., Hammarström, H., Haspelmath, M., Kaiping, G. A., & Gray, R. D. (2018). Cross-linguistic data formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data*, *5*(1), 1–10. https://doi.org/10.1038/sdata.2018.205

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Greenberg, J. H. (1963). *Universals of language*. MIT Press.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, *86*(3), 663–687. https://doi.org/10.1353/lan.2010.0021

Haspelmath, M. (2023). Defining the word. *WORD*, *69*(3), 283–297. https://doi.org/10.1080/00437956.2023.2237272

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525x0999152X

Humboldt, W. von. (1836). *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwickelung des Menschengeschlechts*. Druckerei der Königlichen Akademie der Wissenschaften.

Kemmerer, D. (2019). *Concepts in the brain: The view from cross-linguistic diversity*. Oxford University Press.

Koptjevskaja-Tamm, M. (2008). Approaching lexical typology. In M. Vanhove (Ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (pp. 3–52). John Benjamins. https://doi.org/10.1075/slcs.106.03kop

Koptjevskaja-Tamm, M., Rakhilina, E., & Vanhove, M. (2015). The semantics of lexical typology. In N. Riemer (Ed.), *The Routledge handbook of semantics* (pp. 434–454). Routledge.

Levinson, S. C. (2012). The original sin of cognitive science. *Topics in Cognitive Science*, *4*(3), 396–403. https://doi.org/10.1111/j.1756-8765.2012.01195.x

List, J.-M., Cysouw, M., & Forkel, R. (2016). Concepticon: A resource for the linking of concept lists. In N. Calzolari, K. Choukri, T. Declerck, M. Grobelnik, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation* (pp. 2393–2400). European Language Resources Association.

Majid, A., Enfield, N. J., & van Staden, M. (2006). Parts of the body: Cross-linguistic categorisation (Special Issue). *Language Sciences*, *28*(2–3), 137–360. https://doi.org/10.1016/S0388-0001(06)00005-2

Majid, A., & van Staden, M. (2015). Can nomenclature for the body be explained by embodiment theories? *Topics in Cognitive Science*, *7*(4), 570–594. https://doi.org/10.1111/tops.12159

Mansfield, J. (2021). The word as a unit of internal predictability. *Linguistics*, *59*(6), 1427–1472. https://doi.org/10.1515/ling-2020-0118

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*(1), 3–33. https://doi.org/10.1037/0033-295X.101.1.3

Miton, H., & Morin, O. (2021). Graphic complexity in writing systems. *Cognition*, *214*, 1–15. https://doi.org/10.1016/j.cognition.2021.104771

Morin, O. (2018). Spontaneous emergence of legibility in writing systems: The case of orientation anisotropy. *Cognitive Science*, *42*(2), 664–677. https://doi.org/10.1111/cogs.12550

Morin, O., Kelly, P., & Winters, J. (2020). Writing, graphic codes, and asynchronous communication. *Topics in Cognitive Science*, *12*(2), 727–743. https://doi.org/10.1111/tops.12386

Nakamura, K., Kuo, W.-J., Pegado, F., Cohen, L., Tzeng, O. J. L., & Dehaene, S. (2012). Universal brain systems for recognizing word shapes and handwriting gestures during reading. *Proceedings of the National Academy of Sciences*, *109*(50), 20762–20767. https://doi.org/10.1073/pnas.1217749109

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098. https://doi.org/10.3758/BF03196742

Rzymski, C., Tresoldi, T., Greenhill, S. J., Wu, M.-S., Schweikhard, N., Koptjevskaja-Tamm, M., Gast, V., Bodt, T. A., Hantgan, A., Kaiping, G. A., Chang, S., Lai, Y., Morozova, N., Arjava, H., Hübler, N., Koile, E., Pepper, S., Proos, M., Van Epps, B., Blanco, I., Hundt, C., Monakhov, S., Pianykh, K., Ramesh, S., Gray, R. D., Forkel, R., … List, J.-M. (2020). The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data*, *7*(1), 1–12. https://doi.org/10.1038/s41597-019-0341-x

Schriefers, H. J. (1992). Lexical access in the production of noun phrases. *Cognition*, *45*(1), 33–54. https://doi.org/10.1016/0010-0277(92)90022-A

Schuster, S., & Lahiri, A. (2019). Lexical gaps and morphological decomposition: Evidence from German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 166–182. https://doi.org/10.1037/xlm0000560

Smith, A. C., Monaghan, P., & Huettig, F. (2021). The effect of orthographic systems on the developing reading system: Typological and computational analyses. *Psychological Review*, *128*(1), 125–159. https://doi.org/10.1037/rev0000257

Tallman, A. J. R. (2020). Beyond grammatical and phonological words. *Language and Linguistics Compass*, *14*(2), 1–14. https://doi.org/10.1111/lnc3.12364

Tjuka, A., Forkel, R., & List, J.-M. (2022). Linking norms, ratings, and relations of words and concepts across multiple language varieties. *Behavior Research Methods*, *54*, 864–884. https://doi.org/10.3758/s13428-021-01650-1

Tjuka, A., Forkel, R., & List, J.-M. (2023). Curating and extending data for language comparison in Concepticon and NoRaRe. *Open Research Europe*, *2*(141), 1–13. https://doi.org/10.12688/openreseurope.15380.3

Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. *Cognition*, *201*, 1–9. https://doi.org/10.1016/j.cognition.2020.104280

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes*, *9*(3), 393–422. https://doi.org/10.1080/01690969408402125

Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes*, *10*(6), 545–600. https://doi.org/10.1080/01690969508407114

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*(1), 25–64. https://doi.org/10.1016/0010-0277(89)90013-9