Reimagining language

Neil Cohn and Joost Schilperoord

Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Netherlands

Corresponding Author:
Neil Cohn
Tilburg University
Tilburg School of Humanities and Digital Sciences
Department of Communication and Cognition
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
Email: neilcohn@visuallanguagelab.com
ORCID: 0000-0003-0074-1995
Twitter: @visual_linguist

Joost Schilperoord
Email: j.schilperoord@tilburguniversity.edu
ORCID: 0000-0002-1800-3600

This document is a preprint. For referencing, please consult the final version:

Cohn, Neil and Joost Schilperoord. 2022. Reimagining language.
Cognitive Science. 46(7): 1-6, e13164. Doi: 10.1111/cogs.13174

Since its inception, the study of language has been a central pillar to Cognitive Science. Despite an "amodal view" where language is thought to "flow into" modalities indiscriminately (i.e., a modality-agnostic "phonology" or "sensorimotor system"), speech has always been considered the prototypical form of the linguistic system. Due to this predominant view, expressive systems like gesture or drawing often remain considered as mere auxiliary systems to speech. Within this view, only language is assumed to use both arbitrary symbols (e.g., Hockett 1960) and recursion (e.g., Hauser, Chomsky, and Fitch 2002), which contrast with the non-arbitrary expressions in the bodily and graphic modalities. In addition, different expressive behaviors, like speech, gesture, and drawing have been considered as indivisible in relation to each other, leading to multimodality being approached as the combination of these different systems: speech *plus* gesture, or writing *plus* images. However, this modality-agnostic view does not hold up to the evidence about language and expressive modalities, and therefore we argue for the necessity of a reconsideration to capture the full richness of human communication.

It may seem that "language" is modality-agnostic because speech also "manifests" in writing, or, secondarily, in forms like braille, semaphore, morse code, etc. Yet, such forms recycle the properties of other modalities (Hervais-Adelman et al. 2019), foremost those of the graphic modality naturally afforded to create pictures. Amodal transference also persists in beliefs that the bodily modality and speech simply provide alternate and indistinguishable sensorimotor outputs, despite the body having affordances that are unavailable in speech. Indeed, forms like Manually Coded English do not engage such bodily affordances (Supalla 1991), but they are present in natural sign languages (Liddell 2003). This view also manifests in theories that "language" evolved in one modality before transferring to another (whether body-to-vocal or vocal-to-body), a view that is unable to account for the fact that we retain both productive modalities, also after acquiring a full language (McNeill 2012).

Despite the modality-agnostic positioning of language, speech has retained a status as more primary than other modalities. This view yields questions about whether gestures belong as "part of" the language system (Ladewig 2020), and also relegates other modalities (like graphics or sign) as optimal only when the "primary" speech system is impaired. Yet, this auxiliary nature of both the graphic and bodily modalities has been called into question by research on co-speech gestures (Ladewig 2020, McNeill 1992) and on combinations of writing and images such as found in forms like text-emoji combinations, comics, memes, advertisements, and many other contexts (Bateman 2014). In addition, consistent neural responses are recruited in the semantic processing of speech, writing, gestures, drawings, pictures, sounds, and other meaning-making domains (Kutas and Federmeier 2011, Lambon Ralph et al. 2016), while manipulating the grammars of musical and visual narrative sequencing evokes neural responses similar to those evoked by the syntax of written/spoken languages (Cohn 2020, Koelsch 2011). Thus, based on both those behavioral and cognitive data, the primacy of speech and the auxiliary nature of other modalities appears to be unwarranted.

In addition, our expressive systems of speech, gestures, drawings and others are not indivisible, nor do they use unique componential structures. Indeed, language is composed of different, independent subcomponents (phonology, (morpho-)syntax, semantics), but those structures persist in the bodily modality (Liddell 2003) and in non-verbal drawings as well (Cohn 2020). In addition, arbitrariness is not a unique design feature of language. Not only do iconicity and indexicality occur alongside symbolicity within speech (Dingemanse et al. 2015), but this full range of signification also persists in gestures and sign language (Liddell 2003, McNeill 1992) and in graphic representation. Moreover, recursion persists across meaningful expressions in the bodily

and graphic modalities (Cohn 2020, Willats 1997), in non-meaningful expressions in music (Jackendoff and Lerdahl 2006), and in other cognitive behaviors like perception and action (Jackendoff 2011, Zacks and Tversky 2001).

Given this evidence, we argue for a reconsideration of the fundamental assumptions of the language system, one that confronts this amodal speech-centric view of language. Rather than a view where the language architecture maintains a single amodal expressive system for whichever modality may be used (vocal, bodily, graphic), we argue that multiple expressive systems must persist in parallel for each of our productive capacities (speaking, signing, drawing). There are several ways such a ***multimodal*** linguistic architecture could manifest, but three primary components need to be accounted for: ***modalities***, systems for the vocal-auditory, visual-bodily, *and* visual-graphic channels of expression; ***grammar***, a combinatorial system for organizing information; and ***meaning***, a conceptual structure affording both unimodal and multimodal expressions. Recent work has further proposed that these meanings may not be fully abstract and also involve sensory knowledge (Zwaan 2014, Louwerse 2011). Yet no matter the characteristics of meaning, or the representations involved in grammars, any accounting for the richness of human expression must include multiple expressive systems alongside grammar and meaning, however their relationships may be modeled.

One possible model is offered as an extension of Jackendoff's (2002) Parallel Architecture (Cohn and Schilperoord 2022) whereby each component maintains independent but corresponding generative domains (Figure 1a), with unimodal and multimodal behaviors arising out of emergent interactions between these components (Figure 1b).

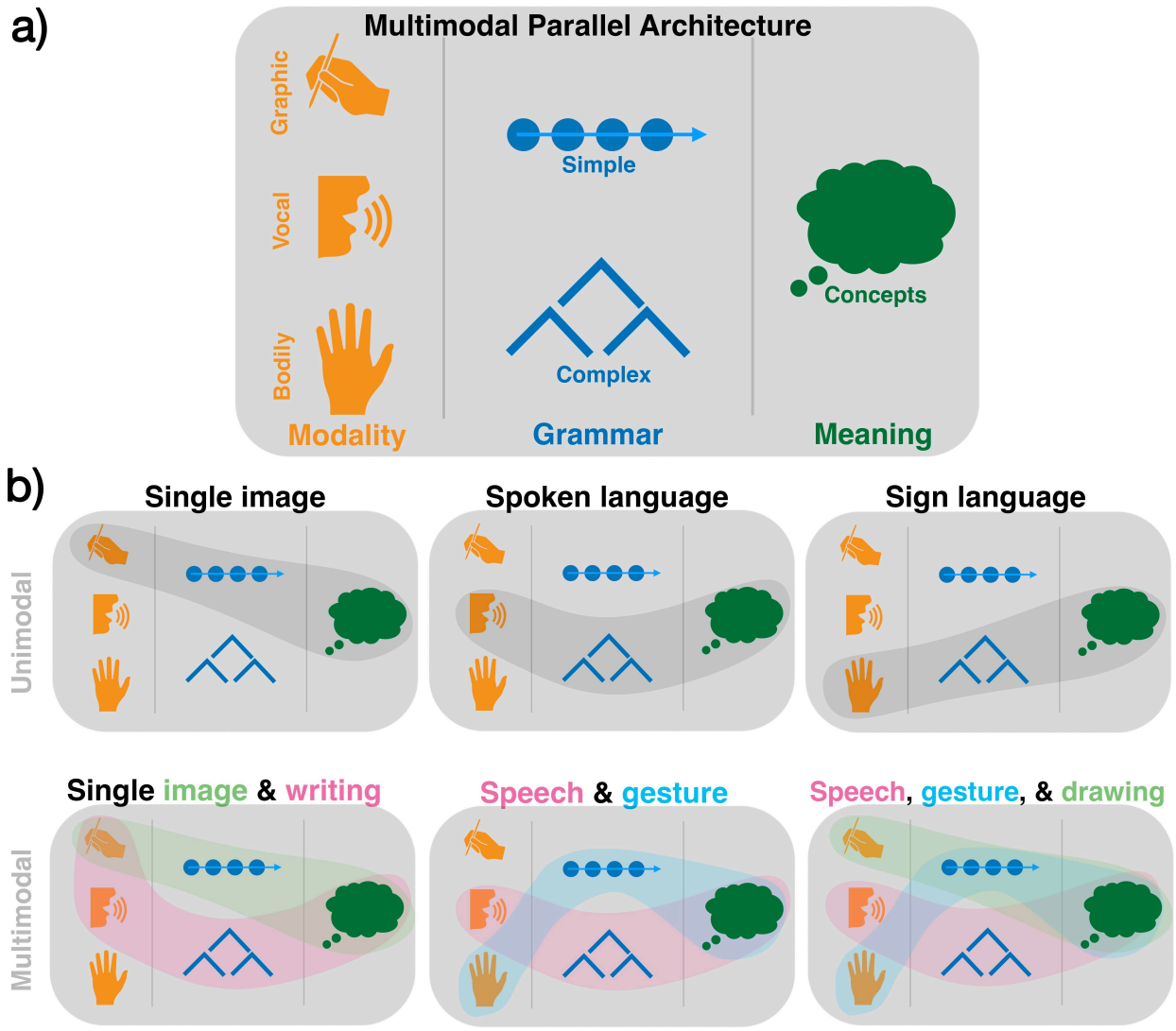

Figure 1. The Parallel Architecture a) including multiple modalities, and b) emergent states within this architecture resulting in unimodal and multimodal expressions.

Within this holistic architecture, all modalities persist in parallel to express meanings, organized by a grammar. In the correspondence of form and meaning, each modality allows symbolicity, indexicality, and iconicity, although modalities may optimally afford certain properties of signification. Modalities may thereby complement each other to create a richer whole communication, without the core versus auxiliary distinction. We further distinguish between *complex grammars*, which use hierarchy and recursion, and *simple grammars*, which use single units or linear sequences (Jackendoff and Wittenberg 2014), that may equally interface with all modalities. This classification is in line with psycholinguistic models differentiating processing of linearity and hierarchy, and models of the neurocognition of sequencing (Dehaene et al. 2015, Uddén et al. 2020). This distinction yields differences in the complexity of the utterances (Figure 1b), such as between gestures which are composed of simple grammars versus sign languages composed of complex grammars.

Thus, this holistic model proposes that "language" (re: speech) does not transfer to, or combine with, other modalities (gesture, pictures), but instead that unimodal and multimodal expressions arise as interactive states out of shared component structures. That is, drawings, gestures, and spoken and sign languages all emerge as combinations of modalities, expressing meanings, organized by grammars. Under this view, multimodality manifests not as interactions between whole behaviors, but as interactions *both within and between* these component structures. Figure 1b illustrates some of many possible emergent multimodal interactions.

Within this architecture, the level of development determines the complexity of each emergent system of expression. Thus, with the requisite exposure and practice people will develop a full sign language, but without such development, they still retain the ability to express meaning with gestures using simple grammars (Goldin-Meadow 2003). Likewise, people who do not acquire full proficiency in a drawing system will retain the ability to create basic drawings (Cohn 2012). In other words, all modalities can potentially develop into robust systems using a lexicon and complex grammar, but even without such development, modalities persist in parallel with meaning-making capacities.

While acknowledgement of both non-vocal modalities and multimodality has grown over the last 40 years in linguistics and psycholinguistics, this has not yet led to a necessary shift in the mainstream linguistic paradigm. Such a shift requires reconfiguring a view on what the linguistic system is and how it works, to address questions including but not limited to: What consequences does a multimodal view have across Marr's (1982) computational, algorithmic, and implementational levels? What affordances do different expressive systems provide for accessing concepts, both unimodally and multimodally? What are the mental representations and processes of lesser-investigated modalities, such as a psycholinguistics of drawing? If the system inherently maintains multiple parallel expressive systems, what does that mean for language evolution, development, relativity, typology, computation, neurobiology, and the lexicon? Given that research on language permeates nearly all disciplines in Cognitive Science, it necessitates a field sensitive to the full, multimodal richness of human communication.

References

Bateman, John A. 2014. *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. New York: Routledge.

Cohn, Neil. 2012. "Explaining "I can't draw": Parallels between the structure and development of language and drawing." *Human Development* 55 (4):167-192. doi: 10.1159/000341842.

Cohn, Neil. 2020. "Your brain on comics: A cognitive model of visual narrative comprehension." *Topics in Cognitive Science* 12 (1):352-386. doi: 10.1111/tops.12421.

Cohn, Neil, and Joost Schilperoord. 2022. "Remarks on multimodality: Grammatical interactions in the parallel architecture." *Frontiers in Artificial Intelligence* 4:1-21. doi: 10.3389/frai.2021.778060.

Dehaene, Stanislas, Florent Meyniel, Catherine Wacongne, Liping Wang, and Christophe Pallier. 2015. "The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees." *Neuron* 88 (1):2-19. doi: http://dx.doi.org/10.1016/j.neuron.2015.09.019.

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen, and Padraic Monaghan. 2015. "Arbitrariness, Iconicity, and Systematicity in Language." *Trends in Cognitive Sciences* 19 (10):603-615. doi: https://doi.org/10.1016/j.tics.2015.07.013.

Goldin-Meadow, Susan. 2003. *The Resiliance of Language: What Gesture Creation in Deaf Children Can Tell Us About How All Children Learn Language*. New York and Hove: Psychology Press.

Hauser, Marc D, Noam Chomsky, and W. Tecumseh Fitch. 2002. "The faculty of language: What is it, who has it, and how did it evolve?" *Science* 298:1569-1579.

Hervais-Adelman, Alexis, Uttam Kumar, Ramesh K. Mishra, Viveka N. Tripathi, Anupam Guleria, Jay P. Singh, Frank Eisner, and Falk Huettig. 2019. "Learning to read recycles visual cortical networks without destruction." *Science Advances* 5 (9):eaax0262. doi: 10.1126/sciadv.aax0262.

Hockett, Charles F. 1960. "Logical Considerations in the study of animal communication." In *Animal sounds and communication*, edited by W.E. Lanyon and W.N. Tavolga, 392-430. Washington, D.C.: American Institute of Biological Sciences, Symposium Series Number 7. Original edition, 1960.

Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackendoff, Ray. 2011. "What is the human language faculty?: Two views." *Language* 87 (3):586-624.

Jackendoff, Ray, and Fred Lerdahl. 2006. "The capacity for music: What is it, and what's special about it?" *Cognition* 100:33-72.

Jackendoff, Ray, and Eva Wittenberg. 2014. "What You Can Say Without Syntax: A Hierarchy of Grammatical Complexity." In *Measuring Linguistic Complexity*, edited by Frederick Newmeyer and L. Preston, 65-82. Oxford: Oxford University Press.

Koelsch, Stefan. 2011. "Toward a Neural Basis of Music Perception – A Review and Updated Model." *Frontiers in Psychology* 2 (110):1-20.

Kutas, Marta, and Kara D. Federmeier. 2011. "Thirty years and counting: Finding meaning in the N400 component of the Event-Related Brain Potential (ERP)." *Annual Review of Psychology* 62 (1):621-647. doi: 10.1146/annurev.psych.093008.131123.

Ladewig, Silva. 2020. *Integrating Gestures: The Dimension of Multimodality in Cognitive Grammar*. Berlin: Walter de Gruyter GmbH & Co KG.

Lambon Ralph, Matthew A., Elizabeth Jefferies, Karalyn Patterson, and Timothy T. Rogers. 2016. "The neural and computational bases of semantic cognition." *Nature Reviews Neuroscience* 18:42. doi: 10.1038/nrn.2016.150.

Liddell, Scott K. 2003. *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.

Louwerse, Max M. 2011. "Symbol interdependency in symbolic and embodied cognition." *Topics in Cognitive Science* 3 (2):273-302.

Marr, David. 1982. *Vision*. San Francisco, CA: Freeman.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

McNeill, David. 2012. *How language began: Gesture and speech in human evolution*: Cambridge University Press.

Supalla, Samuel. 1991. "Manually coded English: The modality question in signed language development." In *Theoretical issues in sign language research. Volume 2: Psychology*, edited by Susan D. Fischer and Patricia Siple, 85-109. Chicago: University of Chicago Press.

Uddén, Julia, Mauricio de Jesus Dias Martins, Willem Zuidema, and W. Tecumseh Fitch. 2020. "Hierarchical Structure in Sequence Processing: How to Measure It and Determine Its Neural Implementation." *Topics in Cognitive Science* 12 (3):910-924. doi: https://doi.org/10.1111/tops.12442.

Willats, John. 1997. *Art and representation: New principles in the analysis of pictures*. Princeton: Princeton University Press.

Zacks, Jeffrey M., and Barbara Tversky. 2001. "Event structure in perception and conception." *Psychological Bulletin* 127 (1):3-21.

Zwaan, Rolf A. 2014. "Embodiment and language comprehension: reframing the discussion." *Trends in Cognitive Sciences* 18 (5):229-234. doi: https://doi.org/10.1016/j.tics.2014.02.008.