

Compositionality in Different Modalities: A View from Usage-Based Linguistics

Michael Pleyer, Ryan Lepic & Stefan Hartmann

Article - Version of Record

Suggested Citation:

Pleyer, M., Lepic, R., & Hartmann, S. (2022). Compositionality in Different Modalities: A View from Usage-Based Linguistics. *International Journal of Primatology / the International Primatological Society*, 45 (3), 670–702. <https://doi.org/10.1007/s10764-022-00330-x>

Wissen, wo das Wissen ist.



UNIVERSITÄTS- UND
LANDESBIBLIOTHEK
DÜSSELDORF

This version is available at:

URN: <https://nbn-resolving.org/urn:nbn:de:hbz:061-20260721-144605-3>

Terms of Use:

This work is licensed under the Creative Commons Attribution 4.0 International License.

For more information see: <https://creativecommons.org/licenses/by/4.0>



Compositionality in Different Modalities: A View from Usage-Based Linguistics

Michael Pleyer¹ · Ryan Lepic² · Stefan Hartmann³ 

Received: 10 August 2022 / Accepted: 26 August 2022 / Published online: 26 September 2022
© The Author(s) 2022, corrected publication 2022

Abstract

The field of linguistics concerns itself with understanding the human capacity for language. Compositionality is a key notion in this research tradition. Compositionality refers to the notion that the meaning of a complex linguistic unit is a function of the meanings of its constituent parts. However, the question as to whether compositionality is a defining feature of human language is a matter of debate: usage-based and constructionist approaches emphasize the pervasive role of idiomaticity in language, and argue that strict compositionality is the exception rather than the rule. We review the major discussion points on compositionality from a usage-based point of view, taking both spoken and signed languages into account. In addition, we discuss theories that aim at accounting for the emergence of compositional language through processes of cultural transmission as well as the debate of whether animal communication systems exhibit compositionality. We argue for a view that emphasizes the analyzability of complex linguistic units, providing a template for accounting for the multimodal nature of human language.

Keywords Usage-based linguistics · Philosophy of language · Sign languages · Cultural evolution · Animal communication systems

Handling Editor: Joanna M. Setchell

✉ Stefan Hartmann
hartmast@hhu.de

Michael Pleyer
pleyer@umk.pl

Ryan Lepic
ryan.lepic@gallaudet.edu

¹ Center for Language Evolution Studies, Nicolaus Copernicus University in Toruń, Toruń, Poland

² Linguistics Department, Gallaudet University, Washington, DC, USA

³ German Department, Faculty of Arts and Humanities, Heinrich Heine University of Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

Introduction

The field of linguistics concerns itself with understanding the human capacity for language. Linguists study the characteristics of the human mind that allow any human to learn any language. Linguists also research the properties of languages that are widely, perhaps even universally shared across linguistic communities, even as languages differ considerably from one to the next. Compositionality is a key concept in linguistics, and it has become increasingly important in recent comparative research on communication systems in non-human animals (animals hereafter) as well. However, the notion of compositionality is multi-faceted, and has spurred many debates among linguists and philosophers of language, as well as adjacent disciplines. The aim of the present paper is to give an overview of how the concept of compositionality is operationalized in linguistics and in comparative research. In doing so, it aims to identify commonalities and differences between research on human language and animal communication systems.

One of the reasons why compositionality is such a multi-faceted and highly debated concept is its history. It is hard to understand current debates about compositionality without taking the origins of the concept in late-nineteenth and early-twentieth century philosophy of language into account. For this reason, we first provide an overview of compositionality in linguistics and neighboring disciplines (*What is compositionality?*). We then zoom in on *Perspectives on compositionality in functional approaches to language and other communication systems*. The next section discusses compositionality from the point of view of one particular framework, namely usage-based construction grammar, thus introducing *A constructionist view of compositionality in spoken and signed languages*. Last, we give an overview of *Compositionality from the point of language evolution and animal communication*, indicating important points of contact between research on animal communication, language evolution, and usage-based approaches.

As this paper is geared towards a broad, interdisciplinary audience, our discussion is often slightly simplified and necessarily incomplete. Despite these simplifications, our goal is to provide a fairly comprehensive overview that helps establish a common ground in discussions about compositionality across disciplines and frameworks. In line with the scope of the present special issue, we will focus particularly on the question “to what extent concepts of compositionality that have been developed for describing linguistic communication are also relevant to the study of other communication systems.”

What is Compositionality?

In this section, we briefly discuss the history of the concept of compositionality, and explain how it has been theorized in contemporary linguistics. The linguistic understanding of compositionality has been strongly shaped by the philosophy of language (see, for example, Szabó, 2020). In the second half of the twentieth

century, compositionality has played a major role in some influential schools of linguistics, namely American structuralism and generative grammar. But its central status has also been challenged by other frameworks within linguistics, especially by so-called usage-based approaches, for reasons that will be spelled out in more detail below.

Various definitions for “compositionality” have been suggested in the literature. Szabó (2020) suggests the following principle as a common denominator: “The meaning of a complex expression is determined by its structure and the meanings of its constituents.” For example, the meaning of a sentence such as *the cat is on the mat* can be easily decoded if we know the meanings of the individual words, even if we have never encountered this specific sentence before. The meaning of the whole is determined by its parts.

This principle of compositionality, according to which the meaning of a complex linguistic unit arises from the meanings of its parts, is canonically attributed to the philosopher Gottlob Frege, but this is not entirely correct (cf. Janssen, 2012; Pelletier, 2001). Both Janssen (2012) and Pelletier (2001) posit that it was Rudolf Carnap who attributed the principle to Frege, though some argue that the principle of compositionality is implicit in Frege’s writing (e.g., Krifka, 1999).

To a certain extent, compositionality can be seen as “a self-evident fact” (Taylor, 2002: p. 97). As Sweetser (1999: p. 132) points out, no linguist would see meaning as intrinsically non-compositional. However, the scope of compositionality has been subject to considerable debate. Some linguistic approaches have tried to formalize meaning, and to make linguistic meaning susceptible to quasi-mathematical modeling. In these formalist approaches, the principle of compositionality engenders a “building-block” approach to language (a metaphor used by Taylor, 2012, among others): we use a relatively small inventory of linguistic units to form a potentially infinite range of sentences. This can only work if language users can, metaphorically speaking, easily put the pieces together, such that the meaning of an utterance can be derived from the meanings of its constituent parts. The principle of compositionality has therefore figured prominently in formal approaches to grammar in general, and to semantics in particular. Taking a generative approach, Katz and Fodor (1963) put compositionality center stage in their endeavor to describe the set of rules that allow language users to understand a potentially infinite amount of novel utterances.

In contrast, others have focused more on the communicative and interactional function of language. This is especially true for functional and cognitive linguistics. These frameworks put the communicative function of language center stage, and argue that language is not an isolated module of the mind, but rather closely intertwined with other aspects of human cognition. Most of those approaches also take a usage-based point of view, i.e., they argue that the meanings of linguistic units, as well as the conventions of how they are used, arise from language users’ experience with them. Such functional, cognitive, and usage-based approaches have argued that purely formalist, structuralist, and generative approaches do not do justice to the full complexity of linguistic meaning, suggesting alternatives to a strict view of compositionality. On the one hand, there are advocates of a more holistic approach in which “there are some items whose makeup at least partially consists in things that are not in its parts” (Pelletier, 2017: p. 33). On the other hand, there are scaffolding

models, which assume an important contribution of contextual, cognitive, world-knowledge, and other factors in ‘fleshing out’ the semantic skeleton created by a sequence of combined items (Cruse, 2000: pp. 79–80). Both accounts argue that there is much more to meaning than compositionality allows. Accordingly, the principle of compositionality is not explicitly framed as the defining feature of human language.

Langacker (1987: p. 449), a primary architect of the cognitive approach, argues that the validity of the principle of (semantic) compositionality “is less an empirical issue than a matter of defining the scope of grammar and linguistic semantics”. He argues that violations of compositionality are often relegated to other domains: in many linguistic frameworks, non-compositional phrasal structures (idioms such as *kick the bucket*, meaning ‘to die’) are excluded from grammatical description by assigning them to the so-called mental lexicon, where individual words and non-compositional phrases are assumed to be stored. Also, the distinction between semantics — the field of linguistic research that is concerned with “context-free” aspects of meaning — and pragmatics — which deals with aspects of meaning that arise in context — is often invoked if the interpretation of a novel expression is too specific to be invoked by general rules. Consider, for instance, the very specific usage contexts in which a phrase like *ok boomer* occurs; it is a dismissive comment made by members of younger generations, linking interpersonal and intergenerational values and conflicts. To understand it, we need to understand much more than the two words involved: we need a fair amount of contextual knowledge. So, linguistic constructs such as the lexicon or pragmatics are invoked to save compositionality as an essential principle in formal approaches to meaning.

Both the clear distinction between lexicon and grammar and the strict divide between semantics and pragmatics are questioned by Langacker and by other cognitive linguists. On this view, a linguistic unit cannot encode a single, invariable meaning, but rather prompts a set of possible context-dependent meanings. Consider the difference in meaning for the word *cat* in *There’s a cat on the mat* (‘one animal’) vs *After the cat got in the way of the SUV, there was cat all over the driveway* (‘a grisly mass’; Langacker, 2008: 144). In Langacker’s terms, the same linguistic units allow for multiple construals depending on the context, which entails a blurring of the seemingly clear line between context-independent and context-dependent facets of meaning, i.e., between semantics and pragmatics. In such a framework, the units do not determine the meaning of the whole, but rather prompt possible interpretations of imagined scenarios.

At the intersection of cognitive linguistics, cognitive science, and philosophy, George Lakoff and Mark Johnson (Lakoff & Johnson, 1999, [1980] 2003; Lakoff, 1987; Johnson, 1987) have also argued against strict views of compositionality while emphasizing the importance of *partial* compositionality. In particular, they argue for an experientialist semantics which views meaning as embodied, i.e., arising from pre-conceptual bodily experiences (Lakoff, 1987: 267). They propose that our pre-conceptual experience is structured in terms of what they call basic-level categories. These include perceptual categories such as *red* and *blue*, or categories rooted in gestalt perception such as *cat* and *person*. Our experiences are also structured in terms of image schemas, which are physical concepts such as *up–down* and

enter–exit that recur in our everyday experiences in the world. Abstract meanings can then arise by metaphorical projection from concrete to abstract domains, or by the projection from basic-level to more complex superordinate and subordinate categories (Lakoff, 1987: 268). Importantly, basic-level and image-schematic concepts provide a starting point for partial semantic compositionality (Lakoff, 1987: p. 280), not conceptual primitives (or “semantic atoms” Cruse (2000: p. 380)) as assumed by formalist approaches that adhere to a strict “building-block” conception of meaning.

The idea that meaning is embodied, experience-based, dynamic, and not fully compositional has become the standard assumption in cognitive linguistics, and it has given rise to what is arguably its most successful variant in present-day linguistics, namely construction grammar. Construction grammar, often abbreviated CxG, is now a fairly heterogeneous field comprising a variety of different approaches (see Goldberg, 2013; Hoffmann & Trousdale, 2013), but they largely share the assumption that human language can be exhaustively described as an inventory of learned pairings of form and meaning, or *constructions*. We will return to usage-based construction grammar in the section “A constructionist view on compositionality in spoken and signed languages” below. Functionally oriented approaches, such as cognitive–functional, emergentist, constructionist, and usage-based approaches, tend to stress a) the cognitive abilities involved in compositionality and b) that compositionality is just one layer of the multifaceted mechanisms of combinatoriality found in human language and cognition. Approaches in this vein often assume that compositionality is not a unitary concept which explains all the different ways of meaningful combination in a communication system. Instead, communication systems such as human language also contain non-compositional mechanisms of combination.

As the distinction between “strict” and “partial” compositionality implies, cognitive linguists argue that compositionality should be viewed as encompassing different types (e.g., Engesser & Townsend, 2019; Pleyer & Hartmann, 2020; Townsend *et al.*, 2018), or should be conceived of as a spectrum (Wulff, 2013). For example, Fauconnier and Turner (2002: pp. 162–165) distinguish between truth-conditional compositionality and compositionality in a more general sense. Discussions about compositionality in formal linguistics are often concerned with truth-conditional compositionality: For instance, “the conditions under which an expression such as ‘brown cow’ will be true of something can be computed from the conditions under which ‘brown’ is true of something and ‘cow’ is true of something” (Fauconnier & Turner, 2002: 163). On a more general view, however, the concept of compositionality can be broadened to aspects of semantics that arguably cannot be modeled in truth-conditional terms, such as figurative language. On this view, then, contextual factors such as interactional cues (e.g., certain intonation patterns indicating that something is not to be understood in a literal sense) or the interlocutors’ knowledge about a particular entity can enter the interpretation of linguistic units, such that their full meaning is “composed” of both the meanings of the units themselves on the one hand and non-linguistic contextual aspects on the other. Some misunderstandings about compositionality in linguistic debates can potentially be attributed to a lack of differentiation between these broader and narrower senses of the concept.

There are also other distinctions between types of compositionality. For example, Townsend *et al.* (2018) distinguish between simple and complex compositionality,

Gil and Shen (2019) distinguish between symmetric and asymmetric compositionality, and Engesser *et al.* (2020) discuss the “openness” of compositionality, that is, the question to what extent the structure of a meaningful combination can be modified. In an attempt to operationalize compositionality for corpus linguistics, i.e., work with authentic language data, Wulff (2008, 2013) treats compositionality as a scalar, quantifiable concept. She proposes that how much of the meaning of a linguistic pattern is accounted for by its individual component words can be expressed by a numeric value. In this way, different constructions can be assigned different compositionality values. Such an approach does not distinguish between compositional and non-compositional constructions in a binary fashion. Instead, idioms such as *get ONE's act together* have a low compositionality score, whereas very transparent collocates such as *write a letter* or *tell a story* rank high in compositionality. Other constructions, such as the quasi-metaphorical expressions *hold ONE's breath* or *scratch ONE's head*, have intermediate compositionality scores.

Overall, then, we have seen that with the emergence of new frameworks and paradigms within linguistics, perspectives on compositionality have diversified over time, and that the multitude of definitions can lead to confusion when the concept is discussed in contemporary linguistics. Our necessarily incomplete overview of the history of compositionality in linguistics shows that when discussing compositionality, we have to distinguish between a) strict compositionality versus partial compositionality, and between b) compositionality on the level of concepts versus compositionality as brought about by the structure of utterances. These distinctions also have reflexes in the literature on animal communication systems.

Perspectives on Compositionality in Functional Approaches to Language and other Communication Systems

As we have seen, compositionality has been construed in different ways. In linguistics, we can roughly distinguish between more formally-oriented approaches that prioritize the principle of compositionality and more functionally-oriented ones in which compositionality plays a lesser role. Another frequent distinction that is found in research on animal communication and language evolution as well as in linguistics is between *compositionality* and *combinatoriality*.

Combinatoriality refers to any system in which parts are combined to yield a new structure, and in contrast to this, compositionality refers to the combination of meaning-bearing units into a new structure where the meaning of the structure is a function of its constituent parts and the way they are combined (e.g., Zuidema & de Boer, 2018; Pelletier, 2017; Hurford, 2012, Zuberbühler, 2020). For example, a word like *cat* is composed of three sounds, written as /k/, /æ/, and /t/ in the International Phonetic Alphabet. These sounds do not carry meaning themselves. Hence, this can be considered an example of combinatoriality. If we combine two or more words, as in *wildcat*, we combine multiple meaning-bearing units (in this case, the words *wild* and *cat*). This would instead be considered an example of compositionality.

Importantly, in animal communication research, the term ‘combinatoriality’ traditionally only refers to the combination of meaningless elements into a larger structure (such as in the phonological syntax of birdsong, Marler, 1977). However, more recent research suggests that some combinatorial systems might also combine meaningful elements into a larger structure without the meaning of that structure being related to the meaning of its parts (Engesser & Townsend, 2019). One example of this are black-fronted Titi monkey (*Callicebus nigrifrons*) alarm calls, which have an “A” call for aerial predators and a “B” call for terrestrial predators. However, callers produced mixed “A” and “B” sequences if a predator is in a ‘non-standard location’, for example if a terrestrial predator is in the canopy, or if an aerial predator is on the ground (Berthet *et al.*, 2018; Zuberbühler, 2020). Such calls, then, are combinatorial in that they combine meaningful elements, but are not compositional as their meaning is not directly predictable from the meaning of the items and their combination (see Schlenker *et al.*, 2016 and Narbona Sabaté *et al.*, 2022 for explanations of this combinatorial structure). In the following, when talking about the ‘combinatorial structure’ of a system, we follow Hurford (2012, p. 5), in using it as a “fancier term for ‘putting things together.’” That is, all systems that combine elements are combinatorial in nature, and the key question is what kind of combinatoriality they exhibit, with compositionality being one kind of combinatorial structure among others (see also Engesser & Townsend, 2019; Zuidema & De Boer, 2018).

For researchers working on animal communication systems, the theoretical proliferation around the concept of compositionality brings with it the challenge of which model(s) of compositionality/combinatoriality to refer to when investigating the way a communicative structure is composed. A second challenge when looking at theories of compositionality in different disciplines pertains to the modalities in which it is investigated. With regard to language, for example, compositionality can be investigated as a property of spoken or signed utterances, and as we have shown in our historical–conceptual overview above, it has been theorized as a defining property of human language.

Moreover, much recent research has emphasized that human communication is fundamentally multimodal in nature. The term *multimodality* refers to two different but interrelated facts: on the one hand, language makes use of multiple resources (or “modes”) simultaneously. Spoken language, for example, is intricately integrated with manual gestures and other bodily forms of expression such as posture and facial expressions (Vigliocco *et al.*, 2014; Levinson & Holler, 2014; Rasenberg *et al.*, 2020; Waciewicz *et al.*, 2020). On the other hand, multimodality also means that different resources can be primary in different forms of language: for signed languages, the gestural–visual modality is central, while in spoken languages, the lion’s share of the communicative interaction takes place in the vocal–auditory modality, even though the importance of non-verbal signals, e.g., co-speech gesture, should not be underestimated.

Human language, whether signed or spoken, then, is fundamentally a multimodal, polysemiotic system (Perlman, 2017; Zlatev, 2019). This view intersects with the recent emphasis in primatology that communication in non-human animals should also be seen as fundamentally multimodal (Liebal *et al.*, 2014; Slocombe *et al.*, 2011). In principle, research on multimodality, meaning, and

compositionality in the language sciences and animal communication research and primatology could be said to move into highly compatible directions. However, an increasing dialogue between these disciplines is hindered by two factors: First, there is a lack of agreement on the concept of meaning. Even within research on primate communication (meaning in research on vocal communication) is operationalized quite differently from the way meaning is operationalized in gestural communication (Liebal & Oña, 2018). Second, there presently still is a “lack of connection between theories of human semantics, and theories of animal communication” (Kershenbaum *et al.*, 2016: 16). Our discussion here has attempted to bridge some of this theoretical disconnect.

Summing up, it seems reasonable to say that the open questions that revolve around the concept of compositionality fall into two broad categories: (i) those concerning the definition of compositionality, which is in turn strongly dependent on theoretical frameworks and their conceptualization of language (see Waciewicz *et al.*, 2020 for discussion), and (ii) those concerning the scope of compositionality (or, more specifically, the compositionality principle) in various communication systems.

A Constructionist View of Compositionality in Spoken and Signed Languages

We have mentioned above that functional, usage-based approaches to compositionality question the conceptualization of compositionality in formal linguistics: cognitive linguistics frames compositionality as a gradient concept. In weighing the relative importance of compositionality vs idiomacity in language, cognitive linguists argue that idiomacity plays a more central role than previously assumed. In this section, we want to zoom in on this approach to compositionality in more detail. More specifically, we discuss the implications of one particular account in the broader framework of usage-based linguistics, namely construction grammar (CxG).

Apart from the fact that much of our own work is situated in this framework, there are two main reasons for focusing on a constructionist approach. First, it has become increasingly popular over the last decades, as a growing amount of handbooks and introductory texts show (e.g., Hilpert, 2019; Hoffmann, 2022; Hoffmann & Trousdale, 2013; Ungerer & Hartmann, *forthc*). Second, its dynamic approach to linguistic signs arguably allows for reconciling some of the seemingly conflicting definitions of compositionality reviewed above to a certain extent. Importantly, CxG has been applied to both spoken and sign languages, and arguably, principles of CxG can be broadly applied to non-human communication systems (see Pleyer & Hartmann, 2020). Thus, our discussion of linguistic compositionality in usage-based CxG can also serve as a link to the discussion on compositionality in animal communication systems in the respective section below, as well as in the other papers in this special issue. In the remainder of this section, we will first discuss CxG analysis of spoken languages and then move

on to signed languages, and we conclude with a discussion of multimodal spoken language.

Spoken Languages

Usage-based construction grammar is one of multiple approaches within a broader family of approaches to language that are connected by the key concept of *constructions*. Constructions are conceived of as conventionalized pairings of form and meaning. They exist at various sizes and levels of abstraction, from individual words like *avocado* to highly schematic constructions such as the so-called ditransitive construction instantiated in utterances like *She gave him the book* or *He baked me a cake* (see, e.g., Goldberg, 1995, 2006; Croft, 2001; Hilpert, 2019). The early development of CxG from influential papers like Fillmore (1988) and Fillmore *et al.* (1988) to Goldberg's (1995) seminal monograph was motivated by the observation that language is much more idiomatic, i.e., tied to fixed, often non-compositional expressions, than previously assumed. For instance, when we hear the above-mentioned sentence *He baked me a cake*, we immediately understand that the speaker is the recipient of the cake. But this 'transfer' meaning is not inherent to any of the words involved. Instead, CxG argues that it is due to the syntactic construction, as ditransitive constructions typically express transfer events (*She gave me a book, she sent me a letter*, etc.), rather than some constituent in the structure.

Note that such an account of meaning construction still involves a fair amount of (gradient) compositionality, but with different "building blocks", as it were. The meaning of a sentence such as *He baked me a cake* is computed from the meanings of the individual words in combination with the conventional meaning of the syntactic construction. As Diessel (2015: p. 300) points out, constructions above the word level can generally be divided into smaller components that contribute to their meanings. But importantly, they are also associated with holistic properties that are linked to the entire pattern, not the individual components. For instance, the 'command' meaning of imperative sentences like *Go home!* is seen as a construction-particular property of the imperative construction that cannot be explained by rules that are independent of this particular construction (see Diessel, 2015: p. 300). These holistic properties are, according to Diessel (2015: p. 300), comparable to the idiosyncratic properties of fully idiomatic expressions such as *kick the bucket*. In such cases, the relationship between the form and meaning of a whole expression and the form and meaning of its parts may be partially analyzable, even if the relationship is not strictly compositional in the traditional sense.

Intriguingly, this view has much in common with Hockett's (1960) concept of productivity, which he sees as one of the most important among his thirteen proposed "design features" of human language. According to Hockett (1960: p. 6), "[I] language is open, or 'productive', in the sense that one can coin new utterances by putting together pieces familiar from old utterances, assembling them by patterns of arrangement also familiar in old utterances." Note that the idea of productivity in the Hockettian sense presupposes compositionality. But what arguably distinguishes CxG from Hockett's account and especially from formalist approaches is that it sees

the construction of meaning as highly dynamic and interactive (see, for example, Pleyer, 2017; Hartmann & Pleyer, 2021). Hence, contextual and social-interactional factors also play a major role in our understanding of linguistic utterances. Arbib (2012: 38) makes this point particularly forcefully:

[T]o the extent that some form of compositionality is involved in the meaning of an utterance, it may be necessary to incorporate into the utterance the bodily cues that accompany it as well as relevant properties of the immediate physical and mental worlds of the participants in a conversation. Even when we read a text, where no bodily or intonational cues are present, our interpretation of a sentence may depend on the “mental world” created by the preceding sentences of the text, as well as expectations about the author’s intentions in creating the text.

In a similar vein, Goldberg (2016: p. 428) remarks that many discussions of compositionality seem to presume that meaning resides in the expressions itself. She argues for a more strongly interactional view instead, making a comparison to pointing: “Pointing is understood, but the meaning does not reside in the point” – instead, common ground, context, and background knowledge are needed to make sense of the gesture (see Goldberg, 2016: 429). From a usage-based CxG point of view, then, meaning is constructed dynamically in interaction, which means that the whole “meaning” of a multimodal utterance will always be, to some extent, more than the sum of its parts.

Importantly, many debates regarding how the concept of “constructions” should be adequately defined revolve around (non-)compositionality. It has often been pointed out that Goldberg (2006) has significantly revised her influential definition of constructions originally proposed in Goldberg (1995): According to her original definition, non-compositionality is a *sine qua non* for identifying a linguistic unit as a construction. Thus, highly frequent but fully compositional units such as *I love you* would not qualify as constructions. Her 2006 definition, by contrast, allows for fully compositional constructions as she assumes that even fully compositional units are cognitively entrenched, i.e., recorded mentally as coherent wholes, if they occur with sufficient frequency. In a usage-based perspective, sequences of units that frequently occur together are “chunked” and stored as units. These can then be accessed directly as chunks. Such chunks also become associated with a meaning that is related to the entire chunk, or construction, and is not accessed and formed compositionally. Importantly, these can either be non-transparent idiomatic constructions such as *kick the bucket* or *spill the beans*, but also compositionally analyzable sequences stored as meaningful units, such as *all of a sudden*, *here and there*, or *in a manner of speaking* (Bybee, 2013).

By affording convention and experience a stronger role in determining linguistic structure than compositionality, the usage-based CxG view prevents non-compositional structures from being shunted off to other domains, such as the lexicon or pragmatics, as discussed above. Instead, it is recognized that linguistic units of varying sizes can be conventionalized as pairings of meaning and form within a language community.

Signed Languages

As mentioned in the preceding sections, the symbolic notion of compositionality has played a large role in linguistics: compositionality is central to many theories of human language, and to definitions of human language itself. As a result, when linguists began to study signed languages in earnest in the 1950s and 1960s, compositionality became a pivotal research topic. The prevailing and damaging folk misconception of “sign language” at the time was that signed languages are not true, human languages, but rather unstructured (“nonlinguistic”) or invented pantomime. Against this background, linguists were compelled to demonstrate that signed languages indeed exhibit the agreed-upon unique and defining characteristics of human language (such as the design features identified by Hockett, 1960), thus qualifying as fully-fledged languages.

The earliest of these linguistic projects demonstrated that combinatoriality (also called “duality of patterning” in these studies, following Hockett, 1960) is found in national sign languages such as American Sign Language (ASL; Battison, 1978; Frishberg, 1975; Stokoe, 1960). As described above, combinatoriality refers to the idea that a relatively small amount of meaningless units, such as phonemes, can be combined to yield a virtually unlimited amount of meaningful units — for example, the words *cat*, *act*, and *tack* have vastly different meanings but share the same three phonemes /k/, /æ/, and /t/ (Hockett, 1960). After decades of this research into the componential structure of signs, it is generally agreed among linguists that the signs of sign languages are composed of recurring sub-word features or parameters that include the position of the body and arms and the handshape and location of each hand, on analogy to the phonemes of spoken languages. It is also generally agreed that well-formed signs also usually involve some sort of internal change such as the hands moving from one position to another. Examples of such sign-internal changes include i) the hand changing from one configuration to another related configuration (as in the ASL sign meaning ‘understand’), ii) the hand moving from one location to another within a region of signing space (as in the ASL sign ‘develop’), or iii) a combination of the two, with the hands changing both their configuration and their location (as in the ASL sign ‘inform’) (Fig. 1). These features and dynamics are often compared to units below the word level such as phonemes and syllables in phonological research (and see Morgan, 2017 for a much fuller discussion of the componential structure of signs).

Subsequent research into the linguistic structure of sign languages also addressed specifically compositional sign-internal structure. Klima and Bellugi (1979), for example, emphasized at length (pp. 9–34) that ASL signs exhibit “two faces”, one face which can be characterized as a global or holistic resemblance between the form of a sign and its meaning, and another face that is compositional, and therefore more comfortably “linguistic” (pp. 10, 28), in nature. Klima & Bellugi’s detailed analysis of ASL aspectual inflectional paradigms shows how predicate adjective signs such as those meaning ‘sick’ and ‘careful’ may be formed with varying movement patterns that systematically distinguish aspectual interpretations such as the frequentative (‘often being X’; repeated straight movements) or continuative (‘X for a long time’; repeated elliptical movements) (pp. 257, 259), (Fig. 2).

Fig. 1 The ASL signs meaning (a) ‘understand’, (b) ‘develop’, and (c) ‘inform’ are formed through changes of the configuration of the hand(s) and/or the position of the arm(s). Sign images from <https://aslsignbank.haskins.yale.edu/>. a) ‘understand’ <https://aslsignbank.haskins.yale.edu/dictionary/gloss/311.html>. b) ‘develop’ <https://aslsignbank.haskins.yale.edu/dictionary/gloss/1251.html>. c) ‘inform’ <https://aslsignbank.haskins.yale.edu/dictionary/gloss/1733/>.

Such aspectually inflected signs superimpose a movement pattern coding a particular aspectual value onto the overall shape and configuration of a conventional adjectival sign. This semantic and formal compositionality was later likened to compositional yet nonlinear root-and-pattern morphological systems found in Arabic and other Semitic languages (Sandler, 1989; Aronoff *et al.*, 2005). Though there is no separate segmental affix that can be isolated in ASL aspectual forms, they can nevertheless be analyzed as exhibiting compositional, morphological contrasts in which sign-internal movement patterns co-vary with the aspectual values being expressed.

As a result of this tradition of research, contemporary descriptions of sign language structure often adopt a broadly “building block” framing, where signs are described as being made up of meaningless phonemes as well as meaningful morphemes, and signs are arranged into sentences following general syntactic rules (accessible introductory books adopting this framing include, for example, Sandler & Lillo-Martin, 2006 and Hill *et al.*, 2019; see also Kendon, 2014). An obvious advantage of this framing is that it solidifies the scientific finding that sign languages are indeed human languages, just like any spoken variety. This is important because sign languages have been and continue to be marginalized not only in the societies in which they are used, but also in linguistics as a science. Analyzing sign languages using the same techniques and terminology that are standardly used in spoken language research allows for direct comparisons of languages across the spoken and signed modalities.

However, just as there has been a usage-based, construction-theoretic critique of the principle of compositionality in spoken language research, there has been a parallel reaction in sign language linguistics. These approaches have also risen alongside growing interest in multimodal linguistic analysis: as sign languages are articulated using visible articulators, but not the vocal channel, sign languages provide a natural point of comparison for studies of multimodal spoken language as well.

In the domain of sign language research, the cognitive, usage-based view has proceeded on a few fronts. First, linguists researching iconicity in sign languages have built from the observation that signs often exhibit a puzzling “dual” character: sign forms often evoke their meanings in some way. This is true even for highly conventionalized signs. Under a strict view of compositionality, such forms pose a challenge. Is the sign itself a simple symbol whose meaning is defined by convention alone, or is its meaning a function of its parts? An example of a conventional, iconic sign is the ASL sign meaning ‘hour’. The form of the sign depicts the minute hand of an analog clock traveling around the face of the clock (Fig. 3). Though this sign can be broken down into components that seem to relate to the meaning of the whole, it is not possible to derive the meaning of the ASL sign ‘hour’ from its parts, and the ASL sign ‘hour’ can always be used to refer to the concept ‘hour’, independently of the workings of analog clocks, in particular. Rather than semantically compositional, this sign’s form is analyzable in terms of its conventional meaning (Lepic & Padden, 2017).

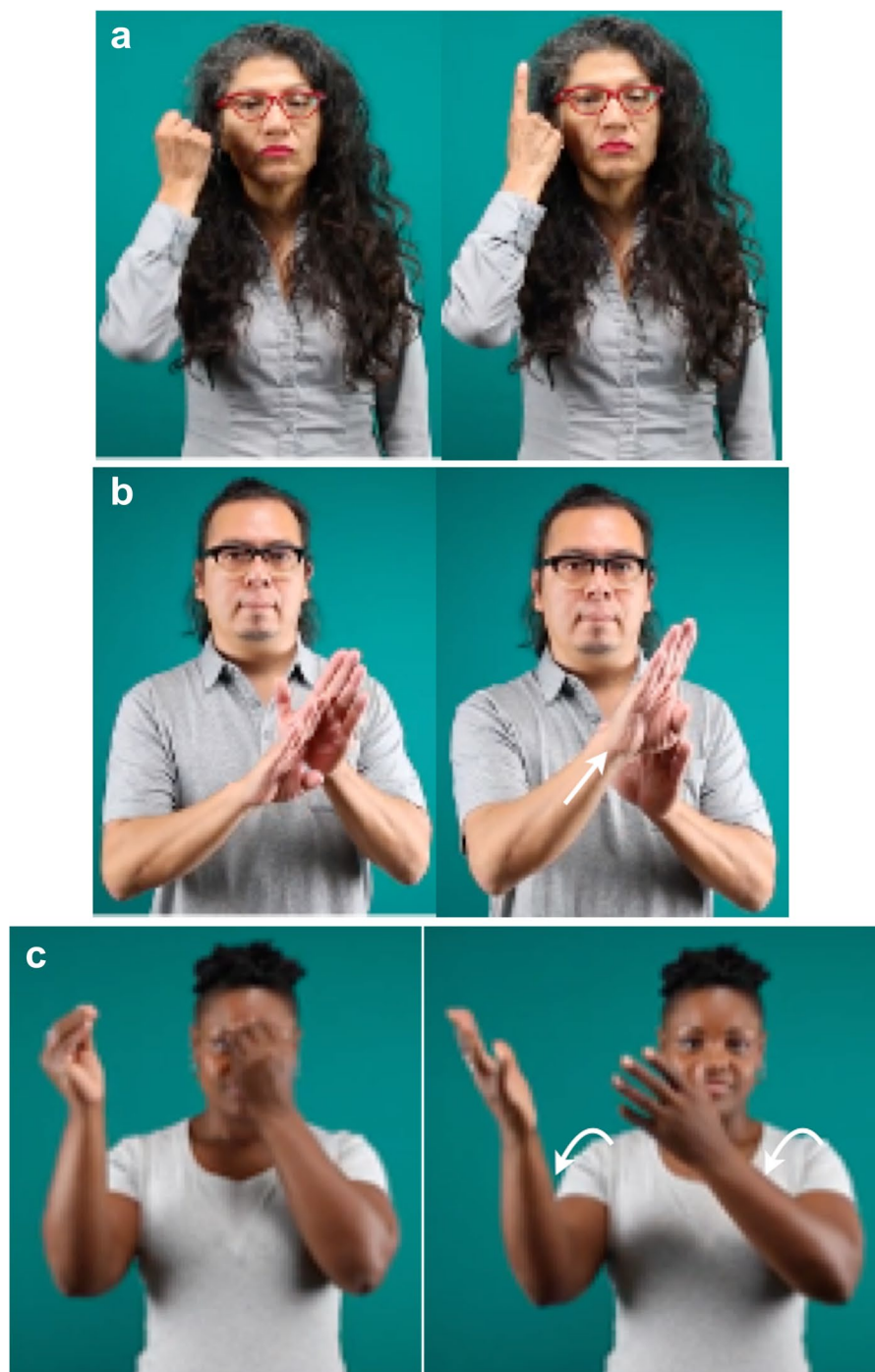




Fig. 2 The same hand configuration used to form (a) the ASL sign meaning 'sick', is also used to form (b) the ASL sign meaning 'often being sick', in which the hands repeatedly move toward the signer's body with a straight movement. a) 'sick', signed with both hands touching the body at the head and abdomen <https://aslsignbank.haskins.yale.edu/dictionary/gloss/255/>. b) 'often being sick', signed with several repeated movements toward the body <https://youtu.be/7hQw3XWLv2U?t=50>.

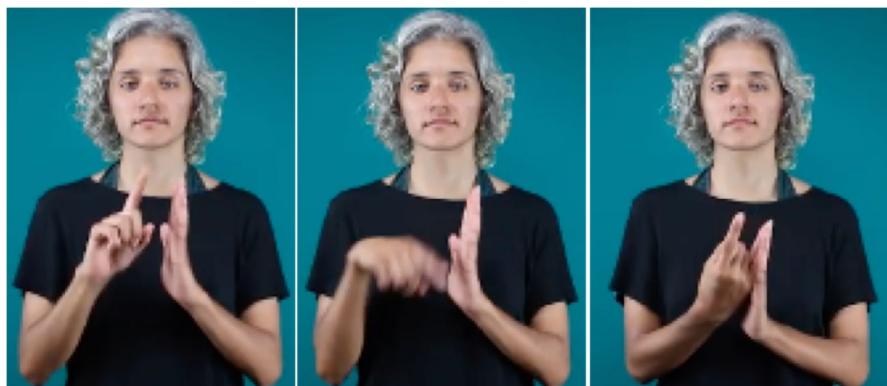


Fig. 3 The conventional ASL sign meaning 'hour' can be analyzed as depicting the movement of the minute hand around the face of an analog clock. <https://aslsignbank.haskins.yale.edu/dictionary/gloss/2745/>.

A large body of literature has examined the various ways that iconicity manifests in ASL and in other sign languages, with iconicity defined as the perception of a motivated link between a sign form and its conventionalized and/or contextual meaning (Wilcox, 2000; Taub, 2001; Wilcox, 2004; Dudis, 2004; Perniss *et al.*, 2010; Emmorey, 2014; Lepic *et al.*, 2016; Occhino *et al.*, 2017). A robust theme in this line of research is that the signs of sign languages are neither entirely arbitrary nor entirely compositional. Instead, iconic signs are motivated

wholes that also exhibit analyzable internal structure, even if that structure is not semantically compositional (Lepic & Padden, 2017).

A related strand of cognitive research in sign language linguistics has demonstrated the ways in which frequency of use interacts with discourse function in processes of grammaticalization, in which lexical units turn into grammatical structures. From this usage-based perspective, the meaning of an individual sign is not determined a priori, but rather signs invite particular interpretations according to the contexts in which they are frequently used. Janzen (1995, 2018), for example, has examined the grammaticalization of content verbs such as the ASL sign forms glossed as FINISH and KNOW into a perfective marker and a discourse marker respectively. Wilkinson (2013, 2016) has similarly identified grammaticalization patterns for the ASL sign forms glossed as SELF and NOT, which are dependent on the contexts and frequency with which particular sign forms are used.

There has not yet been much research on idiomatic phrases in sign languages specifically (cf. Johnston & Ferrara, 2012; Hou, 2022). However, Lepic (2019) argues that there is indeed an expected, gradient relationship between language use and analyzable structure, such that more frequently used signs and sequences of signing have more opportunities to take on a holistic, less analyzable character. From a more broadly cognitive perspective, numerous studies of sign language structure have also shown that signs participate in larger constructions beyond the level of the individual sign (for Irish Sign Language: Leeson & Saeed, 2012; Swedish Sign Language: Nilsson, 2008; Argentinian Sign Language: Wilcox & Martínez, 2020; ASL: Janzen *et al.*, 2001; Auslan: Ferrara & Hodge, 2018). Such a view challenges the strict compositional notion of semantic “building blocks”, and instead supports a construction-theoretic approach to language structure, in which (strict) compositionality plays a secondary role to analyzability of conventional symbolic units of varying sizes and degrees of complexity.

Multimodality

Part of the allure of a structuralist approach to linguistic analysis based on the idea of strict compositionality is its simplicity. Human language is complex, and linguists must make simplifying assumptions in order to proceed toward explaining it. The technology of writing allows for the pinning and preservation of linguistic specimens for analysis. Removing context and abstracting away from particular language users or specific usage events similarly allows for a tightly confined domain of exploration and theorizing. Through the study of frozen, written text, linguists have identified structural units and principles for combining them, within and across languages. However, at the end of the day, language is a social phenomenon, used for communication by actual humans in complex social settings (Enfield, 2009; Kendon, 2014; Müller, 2018). It is not surprising, then, that linguists, psychologists, and philosophers have also expanded their scope beyond isolated written sentences to examine multimodal use of language in more authentic contexts.

Studying *gestures*, movements of the hands and body that accompany speech, McNeill (2005) has argued that language inherently encompasses two dimensions, which he termed the *imagery-language dialectic*. McNeill’s terminology is a bit unfortunate,

as it uses the single word *language* in two contradictory senses. First, *language* is used as a cover term for speech and gesture together. But as part of the imagery–language dialectic, *language* is also used to identify the “linguistic” properties that distinguish speech from gesture. McNeill is quite explicit in conceiving of “linguistic properties” as essentially compositional: gestures are “non-linguistic” because they are “non-morphemic”, are “not realized through a system of phonological form constraints,” and have “no potential for syntactic combination with other gestures” (2005, p. 7).

In an assumed linguistic/non-linguistic dichotomy, “linguistic” but non-compositional units such as idiomatic and formulaic utterances present an analytic challenge, as do “non-linguistic” gestures which are nevertheless used in conventional and structured, hence “language-like” ways (Müller, 2018; Shaw, 2018; Lepic & Occhino 2018). An example of the latter is Zima’s (2017a) construction grammar analysis of circular motion constructions in English, which shows that constructions like [VERB(motion) *in circles*] and verbs like *spin around*, *rotate*, *orbit* are systematically accompanied by circular gestures that are strikingly similar across many different speakers. Zima’s analysis demonstrates that not only do circular motion gestures often co-occur as elements of larger constructional units, but the gestures themselves exhibit systematic analyzable internal structure. In multimodal construction grammar, gestures are therefore seen as components of linguistic constructions, just as words are (see, for example, Feyaerts *et al.*, 2017). From this perspective, multimodal language is inherently “linguistic”, because of its conventionality and analyzability, rather than its possible compositionality alone.

In contrast to “building block” approaches like McNeill’s imagery–language dialectic, there are other accounts of multimodal language use which maintain the observation that language is intricately structured, without equating linguistic structure with strictly compositional structure. Kendon’s (2004) notion of *visible action as utterance* and Enfield’s (2009) notion of the *composite utterance* both avoid treating compositionality as the primary determinant of “language”. Instead, Enfield’s definition, in particular, highlights the semiotics of language use: a composite utterance has multiple components which differ in their degree of conventionalization and in the function that they accomplish in context. In cognitive linguistics, specifically, the recognition that language is essentially multimodal has also led to a specifically multimodal theory of construction grammar, as put forward by, e.g., Cienki (2017); Mittelberg (2017); Zima (2017b); Bresse and Müller (2017). As suggested above, the multimodal turn in linguistics also opens fruitful new avenues for comparing human and non-human animal communication systems.

Compositionality from the Point of View of Language Evolution and Animal Communication

Compositionality in Animal Communication

Compositionality has also played an important role in discussions of language evolution as well as in research on animal communication. In language evolution research, the emergence of compositional structure as a key feature of human syntax has often been treated as a fundamental step in the evolution of language

(e.g., Arbib, 2012; Smith & Kirby, 2012; Zuberbühler, 2020). The question “Can an ape create a sentence?” (Terrace *et al.*, 1979) has also inspired much of the research on symbol-trained primates (see Lyn, 2012 for a review). With regard to animal communication, Fishbein *et al.* (2020, p. 2) state “[t]he question of whether animal communication exhibits even a primitive form of human syntax has sparked some of the fiercest debates in the field.”

Much research comparing human language and animal communication sees compositionality as one of the defining hallmarks of human language. As Adams and Beighly (2013: 400) point out, “the usual case to be made for why animal signals are not language is based on lack of compositional syntax.” For example, in his review of ‘animal syntax’ Hurford (2012, p. 96) comes to the conclusion that “[n]o non-human has any semantically compositional syntax, where the form of the syntactic combination determines how the meanings of the parts combine to make the meaning of the whole.” However, given that many animal communication systems do exhibit complex forms of combinatoriality, this notion has not gone unchallenged. For example, recent research indicates that many vocal sequences in animals might be more complex than was previously believed (Ker-shenbaum *et al.*, 2014). Moreover, in general, in the vocal domain, combination of vocal units into sequences is a pervasive feature of much of non-human primate communication (e.g., Girard-Buttoz *et al.*, 2022). This poses the question which kinds of combinatoriality are found in these sequences.

One reason why compositionality has occupied such a prominent role in these discussions is that there is an influential strand of research that has focused on formal operations found in human language that are hypothesized to be absent in other animals (e.g., Hauser *et al.*, 2002; Miyagawa & Clarke, 2019). However, as we have argued, modern usage-based and constructionist approaches in linguistics offer an important additional perspective on this question (e.g., Pleyer & Hartmann, 2020). This is also in line with proposals in animal communication research that question the value of only looking for compositionality as an all-or-nothing property (Engesser & Townsend, 2019).

Traditionally, the debate surrounding “animal syntax” has distinguished between compositional communication systems and combinatorial communication systems (e.g., Hurford, 2012; Zuberbühler, 2020; Zuidema & De Boer, 2018; see Section “[Perspectives on compositionality in functional approaches to language and other communication systems](#)” above on the distinction between compositionality and combinatoriality). In combinatorial systems, elements are combined to yield new meanings. Putty-nosed monkeys (*Cercopithecus nictitans*), for example, have two alarm calls: ‘pyow’ calls, which are emitted in response to leopards, and ‘hack’ calls, emitted when spotting an eagle. Interestingly, putty-nosed monkeys have also been found to combine both these signals into ‘pyow-hack’ combinations, which are used to initiate group travel (Arnold & Zuberbühler, 2006). However, although this is clearly a combinatorial system, the meaning of ‘pyow-hack’ combinations is not a combination of the meaning of its constituent parts. Thus, it cannot be considered a semantically compositional system, and has instead been interpreted to be “idiomatic”, or exhibiting

“semantic combinatoriality” (Engesser & Townsend, 2019; Schlenker *et al.*, 2016; Zuberbühler, 2020).

Engesser and Townsend (2019) have recently argued that the dichotomy of combinatoriality vs compositionality falls short of capturing the actual complexity of human and animal combinatorial systems (see also Townsend *et al.*, 2018). For example, the complex combinatorial systems of Campbell’s monkeys (*Cercocebus campbelli campbelli*) (Ouatarra *et al.*, 2009) and bird species such as Japanese tits (*Parus minor*) (Suzuki *et al.*, 2019) and southern pied babblers (*Turdoides bicolor*) (Engesser *et al.*, 2016) seem to exhibit compositional features that defy a neat categorization into one of the two categories. Campbell’s monkeys, for instance, use affixation to alter the meaning of calls. Whereas ‘hok’ calls are emitted in response to eagles, and ‘krak’ calls in response to leopards, the affixation of an ‘-oo’ suffix to these call stems alters their meaning. ‘Hok-oo’ calls are emitted to a range of disturbances in the canopy, and ‘krak-oo’ calls are emitted in response to unspecified danger (Ouattara *et al.*, 2009). Suffixation in Campbell’s monkeys therefore seems to serve a generalizing function that relaxes and broadens the meaning of the call stem and therefore seems to carry some semantic content (Coye *et al.*, 2015; see also Kuhn *et al.*, 2018).

Both Japanese tits and southern pied babblers have ‘alert’ and ‘recruitment’ calls. ‘Alert’ calls are responses to predators, whereas ‘recruitment’ calls are used to recruit group members to a food source or nest. Both species of bird combine their calls into “alert-recruitment” calls sequences which function as a “mobbing sequence” that recruit others to mob a predator. Interestingly, both species seem to be sensitive to the order of the call sequence, reacting differently to an artificially reverse “recruitment-alert” sequence than to the “alert-recruitment” sequence (Engesser *et al.*, 2016; Suzuki *et al.*, 2019). Both species therefore seem to exhibit a limited or basic form of compositional structure in this call sequence. In Engesser and Townsend’s (2019) categorization, Campbell’s monkeys examples therefore exhibit “meaning-modifying segmental structures (affixation)”, whereas Japanese tits and pied babblers exhibit “meaning-derived call combinations.” These are semantically compositional, but differ from, e.g., human language and its multiple combinatorial and compositional mechanisms in that compositionality is limited to “a rudimentary, two-call, compositional structure” (Engesser & Townsend, 2019, p. 6).

Engesser and Townsend (2019) therefore propose that research on animal communication should not treat combinatoriality and compositionality as all-or-nothing categories. Instead, combinatorial systems should “be decomposed into finer, transitional forms (e.g., affixation)”, and also acknowledge systems with both combinatorial and compositional aspects. Such an approach is highly compatible with the usage-based and constructionist view discussed above which highlights the role of analyzability of component structures instead of assuming a wholesale, top-down notion of compositionality.

Different Modalities in Animal Communication

As mentioned above, another area where usage-based, constructionist approaches and animal communication research meet is in their appreciation of communication

as a polysemiotic phenomenon that can take place in different modalities and integrate different modalities as well. This means that combinatoriality and compositionality can be investigated in a number of different channels, such as the vocal and gestural domain, as well as their integration. For example, modalities such as face, touch, and body, including gesture, as well as different, nonverbal acoustic properties of vocalizations are all informative aspects of behavior that contribute to communicated meaning in both humans and non-humans (Hall *et al.*, 2019; Filippi, 2020; Liebal *et al.*, 2014).

One modality that has received increasing attention in animal communication research is the gesture system of great apes. Great apes have relatively large inventories of gestures, with documented inventories ranging from 64 (in orangutans, Cartmill & Byrne, 2010) to 102 (in gorillas, Genty *et al.*, 2009; also see Byrne *et al.*, 2017). As great apes also combine these gestures into sequences, and the number of the inventory affords a large number of combinations, it has also been investigated whether these systems exhibit compositionality. However, just as for alarm calls, the question whether these sequences follow ‘syntactic’ ordering principles is hotly debated.

Liebal *et al.* (2004), for example, argue that great ape gesture sequences should not be seen as “syntactic” as they are highly repetitive and have a very limited sequence length. In their study of chimpanzee gestures, Liebal *et al.* (2004) found sequences of gestures whose length ranged from two to 39 gestures. Two-thirds of gesture sequences, however, consisted of only two gestures and 40% of gestures were repetitions of the previous gesture. This led the authors to the conclusion that gesture sequences in chimpanzees should not be seen as “planned sequences aimed to increase signal efficacy” (Liebal *et al.*, 2014, p. 162). Similar findings have been made in a study of gorilla gesture sequences, who produced sequences at roughly the same rate as chimpanzees (Genty & Byrne, 2010; cf. Liebal *et al.*, 2014, p. 163). In both cases, sequences seem to be reactions to the contingencies of interaction instead of truly combinatorial or compositional in nature. At the moment, then, “no evidence of syntactic structure has yet been detected” in ape gestures (Byrne *et al.*, 2017).

However, a usage-based approach adds to the existing arguments that looking for strict compositionality in gestural and vocal combinations might not be the most productive approach. From a usage-based perspective, the interpretation of the meaning of constructions is fundamentally based on inferential, pragmatic as well as contextual factors (e.g., Imo, 2015). Compositionality, in this view, is not only a property of the input, but the result of an active process of interpretation and meaning composition by the receiver.

Similarly, research on animal communication suggests that the meaning of calls and gestures emerges from the integration of signal and social context. Both human and non-human animal communication therefore involve a “rich system of pragmatic inference.” (Seyfarth & Cheney, 2017). Chacma Baboons (*Papio hamadryas ursinus*), for example, possess a rich system of inferential processes and knowledge about the social make-up of their groups that they bring to bear on the interpretation of communicative signals (Cheney & Seyfarth, 2007). It is therefore possible that much of the power and expressivity of animal communication systems come from

the receiver's ability to adaptively integrate multiple sources of information to interpret signals (Fischer, 2021). In other words, a receiver's interpretation of sequences emitted by a signaler can be guided by compositional cognitive processing and pragmatic inferencing without the sequence itself being compositional. For example, Schlenker *et al.* (2016) suggest that in Titi monkey alarm call sequences "the composition of a sequence sometimes reflects the way in which the cognitive situation changes as the sequence is uttered." As such, vocal sequences such as AB-combinations in Titi monkeys can be explained by pragmatic principles influencing encoding of information in a call sequence, making their meaning recoverable through pragmatic inference by receivers (see also Narbona Sabaté *et al.*, 2022). This also suggests that processes of meaning composition in interpretation and comprehension are of strong interest both when looking at human and animal communication.

Recent work on chimpanzee gestures has also proposed that these signals exhibit multimodal "componentiality", that is that they form a system of separate recombining elements (Oña *et al.*, 2019). For example, different manual gestures can be combined with different facial expressions. However, these elements do not contribute stable individual meanings, but different combinations elicit different responses. For example, the addition of a bared teeth face to a stretched arm gesture increased the probability of eliciting affiliative behavior from a conspecific. But adding the same facial expression to a bent arm gesture did not bias conspecifics to react with either an affiliative or non-affiliative response, as opposed to a unimodal bent arm gesture with a neutral face, which increased the probability of receiving an affiliative response. Adding the bared teeth face to a manual gesture thus did not influence the meaning of the gesture in a predictable way, unlike in a compositional system. However, the addition of a facial expression did have an effect on conspecifics, although with different effects in different contexts. This suggests that there is an awareness of the different components of the signal which are then interpreted in a context-dependent, inferential manner. As such, this principle of componentiality in these gestures has interesting parallels to the usage-based concept of analyzability that we have introduced here.

Cultural Transmission and Compositionality

There is a wealth of research on the role of cumulative cultural transmission in the emergence of combinatorial and compositional structure in human language. Both computational (e.g., Brighton, 2002; Smith *et al.*, 2003) and experimental (Kirby *et al.*, 2008; Scott-Phillips & Kirby, 2010) approaches have shown that compositional structure can also emerge when participants are asked to learn an "alien language." In the so-called iterated learning paradigm, participants are asked to learn initially random labels for entities, but in a testing phase are then also tested on untrained entities. The output of the first generation of learners then becomes the input for the second generation of learners, whose output becomes the learning input for the next generation and so forth. Over time, these artificial languages evolve to become more structured and can even evolve compositional structure (see Tamariz, 2017 for a review).

These results are not confined to written stimuli. Iterated learning can also lead to the evolution of systematic structure in the vocal (Perlman *et al.*, 2015) and gestural domains (Motamedi *et al.*, 2019), as well as other auditory and visual signal spaces, for example, using artificial whistled languages (Verhoef *et al.*, 2014) or abstract ‘alien’ graphemes (Cuskley, 2019). Moreover, systematic structure, including compositionality, can also evolve without general turnover in communities of interacting participants (Fay *et al.*, 2010; Nölle *et al.*, 2018; Raviv *et al.*, 2019, 2020).

The emergence of compositional structure, and systematic structure more generally, is often explained in terms of a trade-off between pressures for expressivity and learnability, or simplicity (or “compressibility”, as Smith *et al.*, 2013 call it). These two forces have a long history in the language sciences: von der Gabelenz ([1891] 2016: p. 269) mentions a ‘drive towards clarity’ (*Deutlichkeitstrieb*) and a ‘drive towards ease’ (*Bequemlichkeitstrieb*) as competing forces. In Smith *et al.*’s (2013) iterated learning approach, the two forces of learnability and expressivity are argued to correlate with different types of language structure, “holistic” and “compositional”.

In a holistic system, each meaning is encoded with a specific sign. In an experimental setup, this means that each object in the “meaning space” of an artificial miniature language has its own label assigned to it. For instance, a red circle could be called *wuneho*, a red triangle *lemilipo*, and a red square *wupa*, as in Kirby *et al.*’s (2008) much-cited experiment. In a compositional system, by contrast, the signals can be decomposed into smaller units representing different aspects of the meanings (see Cornish, 2010: 117). For instance, the categories ‘shape’ and ‘color’ could be coded systematically by a set of signals. In Kirby *et al.* (2008), one compositional language evolved the “prefixes” *n-*, *l-* and *r-* for black, white and blue (e.g., *neheki*, *lahoki*, and *reiki* for black, blue, and red circles with horizontal motion, and *nepilu*, *lanepuli* and *repilu* for black, blue, and red squares with spiraling motion respectively). Smith *et al.* (2013) assume that it is the trade-off between learnability and expressivity that gives rise to the compositional structure of language. A purely holistic language would be massively inefficient as it would require a vast amount of different signs, hence being virtually or even literally unlearnable. However, a pressure for learnability alone does not give rise to expressive language systems but rather to “degenerate” ones where all objects are associated with a single, highly ambiguous label (Smith *et al.*, 2013: p. 1349), e.g., *nepa* for circles, triangles, and squares of all colors, as in one of the trials of Kirby *et al.*’s (2008) iterated learning experiment.

One important caveat to research on the emergence of compositionality in iterated learning experiments is that most studies use rather abstract measurements. Kirby *et al.*, (2008: 10,682, see also Brighton *et al.*, 2005; Kirby *et al.*, 2015), for example, quantify the degree of systematicity in the mapping of form and meaning by calculating the edit distance for all pairs of signals and correlating them to the corresponding distances between meanings (differences in, e.g., shape, color, and other meaning dimensions relevant in the signal space that was used for the experiment).

As a simplified (made-up) example, let us assume that our meaning space consists of a black square ■, a white square □, and a white circle ○. In one language,

they are called *napa* ■, *mepa* □, and *metu* ○ respectively. In another one, they are called *mubo* ■, *hraz* □, and *kwgl* ○. ■ and □ as well as □ and ○ differ in one meaning dimension (color in the first example and shape in the second example). ■ and ○ differ in two meaning dimensions (color and shape). Hence, if there is a systematic correspondence between form and meaning, the model expects the words for ■ and □ and the words for □ and ○ to be more similar to each other than the words for ■ and ○. In the first language, this is the case, according to the Levenshtein edit distance, which measures the number of edits necessary to get from one word form to the other: Changing *napa* into *mepa* requires two modifications ($n \rightarrow m$, $a \rightarrow e$), as does changing *mepa* into *metu*, but changing *napa* into *metu* requires four. Hence, the first language is considered highly structured and systematic. In the second, by contrast, the three words do not share any material, hence the language is identified as unstructured and unsystematic by the Kirby *et al.* (2008) measure.

However, such measures of structure are of course not able to distinguish different types of linguistic structure present in a system. For example, they cannot distinguish between whether a system makes use of compositionality, or of structured, consistent use of homonymous, structured ambiguous items which are systematically used with a number of different meanings (Raviv *et al.*, 2019). In fact, in most iterated learning research, “evidence for compositionality (e.g., re-use of sub-strings) was based solely on individual examples of signal systems with such structures, as analyzed manually by the authors” (Raviv *et al.*, 2019, p. 158). However, more recent work has also tried to integrate segmentation algorithms to lend “statistical support for the systematic re-use for sub-strings in addition to subjective observations” (Raviv *et al.*, 2019, p. 58).

The emergence of compositionality in experimental setups is a prime example of cumulative cultural evolution, which has often been treated as a uniquely human feature. However, as Mesoudi and Thornton (2018) point out, there is tentative evidence that at least some animals may show simple forms of cumulative cultural evolution in domains such as foraging and tool use. To the best of our knowledge, there is no evidence of cumulative cultural evolution in animal communication systems so far. If we assume that complex compositionality as we find it in human languages is typically a result of cumulative cultural evolution, this might help explain why we only find limited compositionality in animal communication systems.

Sign Languages as “Emerging” Languages

Research on signed languages has also informed language evolution research in many ways. However, it must be noted explicitly that, in contrast to the non-human animal communication systems discussed in previous sections, the sign languages of the world are used by human individuals living in human societies. As Braithwaite (2020) has argued, researchers working with so-called emerging sign languages have not always taken enough care to avoid exoticizing signing communities. When discussing this topic in a paper like the present one that also takes a comparative perspective, it is therefore important to stress that all sign languages, including and especially “emerging” ones, must be respected as human languages

(see also Kusters, 2009; Zeshan & De Vos 2012; Kusters *et al.*, 2017). It would be dehumanizing to imply that any person's signing practices could be more like non-human animal communication than "true" human language, regardless of the researcher's assumptions or intentions. So, some care must be taken when discussing sign languages together with animal communication systems, to avoid placing signing communities on a scale from "less linguistic" to "more linguistic", on the basis of structural considerations like compositionality alone (this is somewhat similar to the framing of "Creole exceptionalism" as identified and critiqued by DeGraff, 2001, 2005).

However, as stated above in "[Signed Languages](#)", a key theme in sign language research, especially in the late 1900s, was to determine how the accepted properties of language, which had been determined on the basis of spoken languages (and often the written forms of spoken languages), are borne out in sign languages (e.g., Meier *et al.*, 2002). Accordingly, the lens that sign languages provide for the field of language evolution encompasses several facets:

First, because they appear in a different articulatory modality than spoken languages, sign languages afford theorists an opportunity to distinguish the characteristics of human language from characteristics of human speech. For example, of Hockett's (1960) thirteen design features of *language*, at least two (vocal–auditory channel, broadcast transmission) are specific to *speech*, and therefore require modification (cf. Waciewicz & Żywiczyński, 2015). Similarly, a universal statement such as "all human languages have vowels" is empirically false, as this statement is true only of *spoken* human languages. Thus, the existence of signed and spoken languages provides a point of departure for considering how human languages may look in other articulatory modalities (as discussed above in "[Cultural Transmission and Compositionality](#)").

Second, sign languages are articulated in the same articulatory modality as co-speech gestures. This allows researchers to study how similar manual forms may be used for similar functions across signing and speaking populations, as well as how they may differ across populations (e.g., Cooperrider *et al.*, 2021; Shaw, 2018). One theme in this research is the extent to which the manual signal is integrated with the vocal signal in some way, as is typically the case for co-speech gesture, compared to when the hands are the primary articulators, as is typically the case for sign language use. As suggested above in "[Multimodality](#)", in a multimodal theory of construction grammar, co-speech gestures can be analyzed as elements of larger constructions, which parallels how multi-channel utterances are analyzed in CxG analyses of sign languages (Ferrara & Hodge, 2018; Lepic & Occhino 2018).

Third, sign languages have sociolinguistic profiles that make them highly relevant for understanding the pressures at work in language evolution specifically, and language change more generally. The majority of sign languages are used by relatively small communities, embedded in larger, multilingual societies. Additionally, sign language users experience heterogeneous patterns of language exposure, as, statistically, deaf babies are not typically born to signing caregivers, and some language planning is required to ensure access to a language that is accessible for the deaf child (see Lillo-Martin & Henner, 2020 for a fuller discussion of sign language acquisition). So, it behooves researchers interested in the relationship between

language structure and community structure (e.g., the linguistic niche hypothesis, that the structural properties of a language may correlate with the social structure of the language community, Lupyan & Dale, 2010) to also consider sign language communities.

Finally, as mentioned above, signed languages are also thought to be much younger than spoken languages, and it is possible to assign some sign languages a point of origin in recent history (Fenlon & Wilkinson, 2015; Meir *et al.*, 2010). These young languages provide researchers an opportunity to hypothesize about the forces that shape a language that is close to its identified historical origin. In the context of language evolution, “emerging” sign languages, in particular, have also been used to hypothesize about the organization of linguistic properties in a communication system at different time points, or as a function of the social networks of their users. As Hou and de Vos (2022) explain, the variety and complexity of signing practices around the world can be classified according to numerous criteria: the terms “emerging” or “young” sign language contrast with terms such as “established” or “mature” sign language, to identify the age or relative time depth of a community’s signing practices.

Perhaps most famously, the signing practices of two signing communities have been described from this perspective of “emerging sign languages” in language evolution research: Nicaraguan Sign Language (NSL; Kegl *et al.*, 1999; Senghas & Coppola, 2001) and Al-Sayyid Bedouin Sign Language (ABSL; Padden *et al.*, 2010; Sandler *et al.*, 2005). NSL’s identified point of origin is the late 1970s, coinciding with the development of special education programs in the city of Managua, which afforded deaf people new opportunities to come together and communicate (Senghas & Coppola, 2001). ABSL’s identified point of origin is in the 1930s, marked by the birth of two deaf sons to the founder of the Al-Sayyid Bedouin community in present-day Israel (Sandler *et al.*, 2005). In both contexts, deaf individuals have continued to enter the community over time, and the existing members of the signing community have continued to communicate with one another in the course of their daily lives. Through natural use, the signing practices in these communities have drifted according to the same sorts of pressures that shape any living language, over time (see also Hou, 2020 and references cited therein).

As discussed above in “[Cultural transmission and compositionality](#)”, several studies of artificial language learning in tightly controlled laboratory settings have shown that compositional structure often arises as a result of transmission. Compositionality is therefore theorized to be a response to the competing pressures of expressivity and simplicity, as shaped by learnability; artificial languages are easier to learn and to transmit to new learners when there is a clear compositional relationship among the forms that are being learned by adult learners and the (typically highly limited and structured) meaning space that the forms map onto. The hypothesis that transmission leads to increased linguistic structure is also found in the emerging sign languages literature. Often, this research traces stages of complexification in the domains of phonology, morphology, and syntax, from gesture to “homesign”, to emerging sign language, to mature sign language (Brentari & Coppola, 2012; Goldin-Meadow & Brentari, 2017). As noted above, a potential problem with this line of argumentation is that it is easy to misinterpret these findings as

positioning some linguistic communities as more advanced or superior to others, on the basis of their structural properties alone (and see Müller, 2018 for additional commentary).

In contrast to a progression narrative of language emergence, a usage-based approach does not place gesture and sign language at opposite ends of a linguistic continuum (Enfield, 2009, Lepic & Occhino 2018). Under such a framing, change over time is not progress or decay, it is simply change. All languages shift continually, according to the dynamics of how they are used within a community. In line with this view, research on a wider range of signing communities has shifted to take a more ethnographic or community-based approach, to complement the existing, primarily psychological and elicitation-based literature (Green, 2014; Hou, 2020; Kusters & Sahasrabudhe, 2018; Nyst, 2012). It seems likely that a diversity of approaches to documenting the communication practices of signing communities will yield even richer data for understanding the sociolinguistic profiles that characterize these communities, and the role that community structure plays in the evolution of language.

Conclusion

This paper has offered a broad overview of discussions about compositionality in linguistics and beyond, with a special focus on the relationship between multimodality and compositionality. As our review has shown, compositionality is a key concept in linguistics, philosophy of language, and cognitive science, in that it has been understood as a cornerstone of human language in many different approaches. As such it has also garnered much attention in research on animal communication.

While the presence of compositionality in both signed and spoken languages is undisputed, both its definition and scope as well as its relevance for distinguishing human languages from other communication systems have been subject to controversial discussion. As for its scope, the importance of the principle of compositionality has been subject to re-evaluation in recent decades. For a while, compositionality was seen as such a central feature of human language that it was used as a key argument to demonstrate that sign languages are fully-fledged languages. But while the latter point is undisputed by now, the key role of compositionality as a design feature of language has been questioned. Especially usage-based approaches to language argue that language is rarely, if ever, strictly compositional, and assign a much more central role to idiomaticity and the principal analyzability of linguistic constructions.

As for its role as a hallmark of human language, many recent studies have sought to demonstrate that compositionality also exists in non-human communication systems, if only to a limited extent. Moreover, just as human language exhibits many different types of combinatoriality apart from compositionality, animal communication systems exhibit different combinatorial properties.

Combining insights from research on human language and animal communication systems, compositionality has also been a central topic in language evolution research. Increasingly, this research has also pointed to the importance of

multimodality in language evolution, and the role of cultural transmission in the emergence of structure. Sign languages, including “emerging” sign languages, have also been used as a window into properties of language from a multimodal perspective, as well into the pressures that shape language structure and use.

Multimodality, combinatoriality, and gradient compositionality are of central importance when investigating structure in human language and its evolutionary emergence, as well as in animal communication. Overall, in line with multimodal approaches to communication as found in animal communication research, the approach to compositionality we have outlined here offers the potential to look for convergences between usage-based concepts such as idiomaticity, analyzability, and compositionality as a matter of degree on the one hand, and research on combinatoriality and compositionality in animal communication on the other.

Acknowledgements We would like to thank two anonymous reviewers as well as Nathalie Gontier, Evelina Daniela Rodrigues, and Joanna Setchell for helpful comments on a previous version of this paper. The usual disclaimers apply.

Funding Open Access funding enabled and organized by Projekt DEAL. Michael Pleyer was supported by the Centre of Excellence IMSErt - Interacting Minds, Societies, Environments, Nicolaus Copernicus University in Toruń.

Declarations

Ethical Note As this paper reviews the existing literature on the topic, no ethically relevant experiments were conducted. Images of identifiable persons used in this paper are taken from the ASL Signbank, licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license. The depicted individuals have given their informed consent to the use of their pictures for the ASL Signbank and publications building on it.

Conflict of Interest The authors declare that they have no conflict of interest.

Data Deposition Information No data were analyzed for the present paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, F., & Beighly, S. M. (2013). Information, meaning and animal communication. In U. E. Stegmann (Ed.), *Animal Communication Theory: Information and Influence* (pp. 399–420). Cambridge University Press.
- Arbib, M. A. (2012). Compositionality and beyond: Embodied meaning in language and protolanguage. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The Oxford handbook of compositionality* (pp. 475–492). Oxford University Press.

- Arnold, K., & Zuberbühler, K. (2006). Language evolution: semantic combinations in primate calls. *Nature*, 441(7091), 303.
- Aronoff, M., Meir, I., Padden, C., & Sandler, W. (2005). Morphological universals and the sign language type. *Yearbook of Morphology*, 2004, 19–39.
- Battison, R. (1978). *Lexical borrowing in American Sign Language*. Linstok Press.
- Berthet, M., Neumann, C., Mesbahi, G., Casar, C., & Zuberbühler, K. (2018). Contextual encoding in titi monkey alarm call sequences. *Behavioral Ecology and Sociobiology*, 72, 8. <https://doi.org/10.1007/s00265-017-2424-z>
- Braithwaite, B. (2020). Ideologies of linguistic research on small sign languages in the global South: A Caribbean perspective. *Language and Communication*, 74, 182–194.
- Brentari, D., & Coppola, M. (2012). What sign language creation teaches us about language. *Wiley Interdisciplinary Review of Cognitive Science*, 4(2), 201–211. <https://doi.org/10.1002/wcs.1212>
- Bressem, J. & Müller, C. (2017). The “Negative–Assessment–Construction — a multimodal pattern based on a recurrent gesture? *Linguistics Vanguard* 3(1). <https://doi.org/10.1515/lingvan-2016-0053>
- Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, 8(1), 25–54.
- Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, 2(3), 177–226.
- Bybee, J. L. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 49–69). Oxford University Press.
- Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: Intentional communication with a rich set of innate signals. *Animal Cognition*, 20(4), 755–769.
- Cartmill, E. A., & Byrne, R. W. (2010). Semantics of primate gestures: Intentional meanings of orangutan gestures. *Animal Cognition*, 13(6), 793–804.
- Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind*. Chicago University Press.
- Cienki, A. 2017. Utterance construction grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard*, 3(1). <https://doi.org/10.1515/lingvan-2016-0048>
- Cooperrider, K., Fenlon, J., Keane, J., Brentari, D., & Goldin-Meadow, S. (2021). How pointing is integrated into language: Evidence from speakers and signers. *Frontiers in Communication*, 6. <https://doi.org/10.3389/fcomm.2021.567774>
- Cornish, H. (2010). Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems. *Interaction Studies*, 11(1), 112–137.
- Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers’ behaviour in non-human primates. *Proceedings of the Royal Society b: Biological Sciences*, 282(1807), 20150265.
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.
- Cruse, A. (2000). *Meaning in language: an introduction to semantics and pragmatics. (Oxford textbooks in linguistics)*. Oxford: Oxford University Press.
- Cuskley, C. (2019). Alien forms for alien language: Investigating novel form spaces in cultural evolution. *Palgrave Communications*, 5(1), 1–15.
- DeGraff, M. (2001). Morphology in Creole genesis: Linguistics and ideology. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 53–121). MIT Press.
- DeGraff, M. (2005). Linguists’ most dangerous myth: The fallacy of Creole exceptionalism. *Language in Society*, 34, 533–591. <https://doi.org/10.1017/S0047404505050207>
- Diessel, H. (2015). Usage-based construction grammar. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 296–322). De Gruyter.
- Dudis, P. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, 15(2), 223–238.
- Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B*, 369. <https://doi.org/10.1098/rstb.2013.0301>
- Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge University Press.
- Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*, 113(21), 5976–5981.

- Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1493. <https://onlinelibrary.wiley.com/doi/full/10.1002/wcs.1493>
- Engesser, S., Ridley, A. R., Stuart K. Watson, Kita, S., & Townsend, S. W. (2020). *Open Compositionality in Pied Babbler Call Combinations*. In A. Ravnani, C. Barbieri, M. Flaherty, Y. Jadou, E. Lattenkamp, H. Little, M. Martins, K. Mudd, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 13th International Conference on the Evolution of Language (EvoLang13)* (pp. 81–84). Nijmegen: The Evolution of Language Conferences
- Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.
- Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, 34(3), 351–386.
- Fenlon, J. & Wilkinson, E. (2015). Sign languages in the world. In A. C. Schembri & C. Lucas (eds.), *Sociolinguistics and deaf communities* (pp. 5–28. 1st edn). Cambridge University Press. <https://doi.org/10.1017/CBO9781107280298.002>
- Ferrara, L., & Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in Psychology*, 9, 716. <https://doi.org/10.3389/fpsyg.2018.00716>
- Feyaerts, K., Brône, G., & Oben, B. (2017). Multimodality in Interaction. In B. Dancygier (Ed.), *The Cambridge handbook of cognitive linguistics* (pp. 135–156). Cambridge University Press. <https://doi.org/10.1017/9781316339732.010>
- Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64(3), 501–538. <https://doi.org/10.2307/414531>
- Fillmore, C. J. (1988). The mechanisms of “construction grammar.” *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35–55
- Filippi, P. (2020). Emotional voice intonation: A communication code at the origins of speech processing and wordmeaning associations? *Journal of Nonverbal Behavior*, 44(4), 395–417.
- Fischer, J. (2021). Primate vocal communication and the evolution of speech. *Current Directions in Psychological Science*, 30(1), 55–60.
- Fishbein, A. R., Fritz, J. B., Idsardi, W. J., & Wilkinson, G. S. (2020). What can animal communication teach us about human language? *Philosophical Transactions of the Royal Society b: Biological Sciences*, 375, 20190042. <https://doi.org/10.1098/rstb.2019.0042>
- Frishberg, N. (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language*, 51(3), 696–719.
- Genty, E., & Byrne, R. W. (2010). Why do gorillas make sequences of gestures? *Animal Cognition*, 13(2), 287–301.
- Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (*Gorilla gorilla*): Repertoire, intentionality and possible origins. *Animal Cognition*, 12(3), 527–546.
- Gil, D., & Shen, Y. (2019). How grammar introduces asymmetry into cognitive structures: Compositional semantics, metaphors, and schematological hybrids. *Frontiers in Psychology: Language Sciences*, 10, 2275. <https://doi.org/10.3389/fpsyg.2019.02275>
- Girard-Buttoz, C., Zaccarella, E., Bortolato, T., Friederici, A. D., Wittig, R. M., & Crockford, C. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. *Communications Biology*, 5(1), 1–15.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. The University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Goldberg, A. E. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 15–31). Oxford University Press.
- Goldberg, A. E. (2016). Compositionality. In N. Riemer (Ed.), *The Routledge handbook of semantics* (pp. 419–433). Routledge.
- Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Brain and Behavioral Sciences*, 40. <https://doi.org/10.1017/S0140525X15001247>
- Green, E. M. (2014). Building the tower of Babel: International Sign, linguistic commensuration, and moral orientation. *Language in Society*, 43(4), 445–465. <https://doi.org/10.1017/S0047404514000396>

- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.
- Hill, J. C., Lillo-Martin, D. C., & Wood, S. K. (2019). *Sign languages: Structures and contexts*. Routledge.
- Hilpert, M. (2019). *Construction grammar and its application to English* (2nd ed.). Edinburgh University Press.
- Hockett, C. F. (1960). The origins of speech. *Scientific American*, 203, 88–96.
- Hoffmann, T., & Trousdale, G. (Eds.). (2013). *The Oxford handbook of construction grammar*. Oxford University Press.
- Hoffmann, T. (2022). *Construction grammar: The structure of English* (Cambridge Textbooks in Linguistics). Cambridge University Press.
- Hou, L. (2020). Who signs? Language ideologies about deaf and hearing child signers in one family in Mexico. *Sign Language Studies*, 20(4), 665–692. <https://doi.org/10.1353/sls.2020.0023>
- Hou, L. (2022). LOOKing for multi-word expressions in American Sign Language. *Cognitive Linguistics*, 33(2), 291–337.
- Hou, L., & Vos, C. (2022). Classifications and typologies: LLabeling sign languages and signing communities. *Journal of Sociolinguistics*, 26(1), 118–125. <https://doi.org/10.1111/josl.12490>
- Hurford, J. R., (2012). *The origins of grammar: language in the light of evolution II* (Vol. 2). Oxford University Press.
- Imo, W. (2015). Interactional Construction Grammar. *Linguistics Vanguard*, 1(1), 69–77.
- Janssen, T. (2012). Compositionality: Its historic context. In M. Werning, W. Hinzen, & E. Machery (Eds.), *The Oxford handbook of compositionality* (pp. 19–46). Oxford University Press.
- Hall, J. A., Horgan, T. G., & Murphy, N. A. (2019). Nonverbal communication. *Annual Review of Psychology*, 70, 271–294.
- Hartmann, S., & Pleyer, M. (2021). Constructing a protolanguage: Reconstructing prehistoric languages in a usage-based construction grammar framework. *Philosophical Transactions of the Royal Society B*, 376(1824), 20200200.
- Janzen, T., O'Dea, B., & Shaffer, B. (2001). The construal of events: Passives in American Sign Language. *Sign Language Studies*, 1(3), 281–310. <https://doi.org/10.1353/sls.2001.0009>
- Janzen, T. (1995). *The polygrammaticalization of FINISH in ASL*. Winnipeg, MB: University of Manitoba master's thesis
- Janzen, T. (2018). KNOW and UNDERSTAND in ASL: A usage-based study of grammaticalized topic constructions. In Smith, K.A. & Nordquist, D. (Eds.), *Functionalist and usage-based approaches to the study of language: In honor of Joan L. Bybee* (pp. 59–87: Studies in Language Companion Series 192). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.192.03jan>
- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.
- Johnston, T., & Ferrara, L. (2012). Lexicalization in signed languages: When is an idiom not an idiom? *Selected Papers from UK-CLA Meetings*, 1, 229–248.
- Katz, J. J., & Fodor, J. A. (1963). The Structure of a Semantic Theory. *Language*, 39(2), 170–210.
- Kegl, J., Senghas, A., & Coppola, M. (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. DeGraff (Ed.), *Language creation and language change: Creolization, diachrony and development* (pp. 179–237). MIT Press.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kendon, A. (2014). Semiotic diversity in utterance production and the concept of “language.” *Philosophical Transactions of the Royal Society B*, 369. <https://doi.org/10.1098/rstb.2013.0293>
- Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: Not the Markov chains we thought they were. *Proceedings of the Royal Society b: Biological Sciences*, 281(1792), 20141370.
- Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., & Cäsar, C. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews*, 91(1), 13–52.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102.

- Klima, E. S., & Bellugi, U. (1979). *The signs of language*. Harvard University Press.
- Krifka, M. (1999). Compositionality. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 152–153). MIT Press.
- Kuhn, J., Keenan, S., Arnold, K., & Lemasson, A. (2018). On the-oo suffix of Campbell's monkeys. *Linguistic Inquiry*, 49(1), 169–181.
- Kusters, A. (2009). Deaf utopias? reviewing the sociocultural literature on the world's "Martha's Vineyard situations." *The Journal of Deaf Studies and Deaf Education*, 15(1), 3–16. <https://doi.org/10.1093/deafed/enp026>
- Kusters, A., & Sahasrabudhe, S. (2018). Language ideologies on the difference between gesture and sign. *Language & Communication*, 60, 44–63. <https://doi.org/10.1016/j.langcom.2018.01.008>
- Kusters, A., De Meulder, M. & O'Brien, D. (Eds.). (2017). *Innovations in deaf studies: the role of deaf scholars*. Oxford University Press
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. The University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. Basic Books.
- Lakoff, G., & Johnson, M. ([1980] 2003). *Metaphors we live by* (updated edition). University of Chicago Press
- Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford University Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar* (Vol. 1: theoretical prerequisites). Stanford University Press
- Leeson, L. & Saeed, J.I. (2012). *Irish Sign Language: A Cognitive Linguistic account*. Edinburgh University Press.
- Lepic, R. (2019). A usage-based alternative to "lexicalization" in sign language linguistics. *Glossa*, 4(1), 23. <https://doi.org/10.5334/gjgl.840>
- Lepic, R., Börstell, C., Belsitzman, G., & Sandler, W. (2016). Taking meaning in hand: Iconic motivations in two-handed signs. *Sign Language & Linguistics*, 19(1), 37–81.
- Lepic, R. & Occhino, C. (2018). A construction morphology approach to sign language analysis. In Booij, G. (Ed.), *The construction of words: advances in construction morphology* (pp. 141–172). Springer. https://doi.org/10.1007/978-3-319-74394-3_6
- Lepic, R. & Padden, C. 2017. A-morphous iconicity. In Bowern, C., Horn, L. & Zanuttini, R. (Eds.), *On looking into words (and beyond)* (pp. 489–516). Language Sciences Press. <https://doi.org/10.5281/zenodo.495463>
- Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society b: Biological Sciences*, 369(1651), 20130302. <https://doi.org/10.1098/rstb.2013.0302>
- Liebal, K., & Oña, L. (2018). Different approaches to meaning in primate gestural and vocal communication. *Frontiers in Psychology: Language Sciences*, 9, 478. <https://doi.org/10.3389/fpsyg.2018.00478>
- Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *American Journal of Primatology*, 64(4), 377–396.
- Liebal, K., Waller, B. M., Burrows, A. M., & Slocombe, K. E. (2014). *Primate communication: A multi-modal approach*. Cambridge University Press.
- Lillo-Martin, D., & Henner, J. (2020). Acquisition of sign languages. *Annual Review of Linguistics*, 7, 395–419.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One*, 5(1), e8559. <https://doi.org/10.1371/journal.pone.0008559>
- Lyn, H. (2012). Apes and the evolution of language: Taking stock of 40 years of research. In T. K. Shackelford & J. Vonk (Eds.), *The Oxford handbook of comparative evolutionary psychology* (pp. 356–378). Oxford University Press.
- Marler, P. (1977). *The structure of animal communication sounds. Recognition of complex acoustic signals: Report of Dahlem workshop*. Abakon.
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.
- Meier, R., Cormier, K. & Quinto-Pozos, D. (Eds.). (2002). *Modality and structure in signed and spoken languages*. Cambridge University Press
- Meir, I., Sandler, W., Padden, C., & Aronoff, M. (2010). Emerging sign languages. In M. Marschark & P. E. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education* (pp. 267–280). Oxford University Press.

- Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society b: Biological Sciences*, 285(1880), 20180712. <https://doi.org/10.1098/rspb.2018.0712>
- Mittelberg, I. (2017). Multimodal existential constructions in German: Manual actions of giving as experiential substrate for grammatical and gestural patterns. *Linguistics Vanguard*, 3(1). <https://doi.org/10.1515/lingvan-2016-0047>
- Miyagawa, S., & Clarke, E. (2019). Systems underlying human and old world monkey communication: one, two, or infinite. *Frontiers in Psychology*, 1911.
- Morgan, H. (2017). *The phonology of Kenyan sign language (southwestern dialect)*. Doctoral dissertation, University of California, San Diego. <https://escholarship.org/uc/item/9bp3h8t4>
- Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, 192, 103964.
- Müller, C. (2018). Gesture and sign: Cataclysmic break or dynamic relations? *Frontiers in Psychology*, 9, 1651. <https://doi.org/10.3389/fpsyg.2018.01651>
- Narbona Sabaté, L., Mesbahi, G., Dezechache, G., Cäsar, C., Zuberbühler, K., & Berthet, M. (2022). Animal linguistics in the making: The Urgency Principle and titi monkeys' alarm system. *Ethology Ecology & Evolution*, 34(3), 378–394.
- Nilsson, A.-L. (2008). *Spatial strategies in descriptive discourse: Use of signing space in Swedish sign language*. Centre for Deaf Studies, University of Dublin, Trinity College.
- Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, 181, 93–104.
- Nyst, V. (2012). Shared sign languages. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language* (pp. 552–574). De Gruyter Mouton. <https://doi.org/10.1515/9783110261325.552>
- Occhino, C., Anible, B., Wilkinson, E., & Morford, J. P. (2017). Iconicity is in the eye of the beholder: How language experience affects perceived iconicity. *Gesture*, 16(1), 99–125.
- Oña, L. S., Sandler, W., & Liebal, K. (2019). A stepping stone to compositionality in chimpanzee communication. *PeerJ*, 7, e7623. <https://peerj.com/articles/7623/>
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context specific call sequences. *Proceedings of the National Academy of Sciences*, 106(51), 22026–22031. <https://doi.org/10.1073/pnas.0908118106>
- Padden, C., Meir, I., Aronoff, M., & Sandler, W. (2010). The grammar of space in two new sign languages. In D. Brentari (Ed.), *Sign languages: A Cambridge survey* (pp. 573–595). Cambridge University Press.
- Pelletier, F. J. (2001). Did Frege believe Frege's principle? *Journal of Logic, Language, and Information*, 10, 87–114.
- Pelletier, F. J. (2017). Compositionality and concepts—a perspective from formal semantics and philosophy of language. In J. A. Hampton & Y. Winters (Eds.), *Compositionality and concepts in linguistics and psychology* (pp. 31–94). Springer.
- Perlman, M. (2017). Debunking two myths against vocal origins of language: Language is iconic and multimodal to the core. *Interaction Studies*, 18(3), 376–401. <https://doi.org/10.1075/is.18.3.05per>
- Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, 2(8), 150152.
- Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, 1, 227. <https://doi.org/10.3389/fpsyg.2010.00227>
- Pleyer, M. (2017). Protolanguage and mechanisms of meaning construal in interaction. *Language Sciences*, 63, 69–90.
- Pleyer, M., & Hartmann, S. (2020). Construction grammar for monkeys? Animal communication and its implications for language evolution in the light of usage-based linguistic theory. *Evolutionary Linguistic Theory*, 2(2), 153–194. <https://doi.org/10.1075/elt.00021.ple>
- Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in multimodal interaction: An integrative framework. *Cognitive Science*, 44(11), e12911.
- Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition*, 182, 151–164.
- Raviv, L., Meyer, A., & Lev-Ari, S. (2020). The role of social network structure in the emergence of linguistic structure. *Cognitive Science*, 44(8), e12876.
- Sandler, W. (1989). *Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language*. Dordrecht: Foris.

- Sandler, W., & Lillo-Martin, D. C. (2006). *Sign language and linguistic universals*. Cambridge University Press.
- Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar in a new sign language. *Proceedings of the National Academy of Sciences*, 102(7), 2661–2665.
- Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J. P., Kuhn, J., Veselinović, D., Arnold, K., Căsar, C., Keenan, S., Lemasson, A., Ouattara, K., Ryder, R., & Zuberbühler, K. (2016). Formal monkey linguistics. *Theoretical Linguistics*, 42(1–2), 1–90. <https://doi.org/10.1515/tl-2016-0001>
- Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, 14(9), 411–417. <https://doi.org/10.1016/j.tics.2010.06.006>
- Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, 12(4), 323–328. <https://doi.org/10.1111/1467-9280.00359>
- Seyfarth, R. M., & Cheney, D. L. (2017). The origin of meaning in animal signals. *Animal Behaviour*, 124, 339–346. <https://doi.org/10.1016/j.anbehav.2016.05.020>
- Shaw, E. (2018). *Gesture in multiparty interaction*. Gallaudet University Press
- Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: The need for multimodality in primate communication research. *Animal Behaviour*, 81(5), 919–924.
- Smith, K., & Kirby, S. (2012). Compositionality and Linguistic Evolution. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The Oxford handbook of compositionality* (pp. 493–509). Oxford University Press.
- Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: The cultural emergence of compositional structure. *Advances in Complex Systems*, 6(04), 537–558.
- Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Cooperative minds: Social interaction and group dynamics* (pp. 1348–1353). Cognitive Science Society.
- Stokoe, W. C. (1960). *Sign language structure: An outline of the visual communications systems of the American Deaf*. Linstock Press.
- Suzuki, T. N., Griesser, M., & Wheatcroft, D. (2019). Syntactic rules in avian vocal sequences as a window into the evolution of compositionality. *Animal Behaviour*, 151, 267–274.
- Sweetser, E. (1999). Compositionality and blending: semantic composition in a cognitively realistic framework. In T. Janssen & G. Redeker (Eds.), *Cognitive linguistics: foundations, scope, and methodology* (Vol. 15, pp. 129–162). De Gruyter
- Szabó, Zoltán Gendler. (2020). Compositionality. In Edward N. Zalta (ed.), *The Stanford encyclopedia of philosophy* (Fall 2020). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2020/entries/compositionality/>. Accessed 19 Sept 2022.
- Tamariz, M. (2017). Experimental studies on the cultural evolution of language. *Annual Review of Linguistics*, 3, 389–407.
- Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press.
- Taylor, J. R. (2002). *Cognitive grammar*. Oxford University Press.
- Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford University Press.
- Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? *Science*, 206(4421), 891–902.
- Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. *PLoS Biology*, 16(8), e2006425. <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2006425>. Accessed 19 Sept 2022.
- Ungerer, T. & Hartmann, S. (forthc). *Constructionist approaches. Past, present, and future*. Cambridge University Press
- Verhoef, T., Kirby, S., & De Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43, 57–68.
- Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B*, 369, 20130292. <https://doi.org/10.1098/rstb.2013.0292>
- Von der Gabelentz, G. (2016). *Die Sprachwissenschaft: Ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Language Science Press.
- Waciewicz, S., & Żywiczyński, P. (2015). Language evolution: Why Hockett's design features are a non-starter. *Biosemiotics*, 8(1), 29–46.

- Wacewicz, S., Zywiczynski, P., Hartmann, S., Pleyer, M., & Benítez-Burraco, A. (2020). “Language” in language evolution research: Towards a pluralistic view. *Biolinguistics*, 14, 59–101.
- Wilcox, S. (2004). Cognitive iconicity: Conceptual spaces, meaning, and gesture in signed languages. *Cognitive Linguistics*, 15(2), 119–147.
- Wilcox, S., & Martínez, R. (2020). The conceptualization of space: Places in signed language discourse. *Frontiers in Psychology*, 11, 1406. <https://doi.org/10.3389/fpsyg.2020.01406>
- Wilcox, P. P. (2000). *Metaphor in American Sign Language*. Gallaudet University Press.
- Wilkinson, E. (2013). A functional description of SELF in American Sign Language. *Sign Language Studies*, 13(4), 462–490. <https://doi.org/10.1353/sls.2013.0015>
- Wilkinson, E. (2016). Finding frequency effects in the usage of NOT collocations in American Sign Language. *Sign Language & Linguistics*, 19(1), 82–123. <https://doi.org/10.1075/sll.19.1.03wil>
- Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. Continuum.
- Wulff, S. (2013). Words and idioms. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 274–289). Oxford University Press.
- Zeshan, U. & De Vos, C. (eds). (2012). *Sign languages in village communities: anthropological and linguistic insights*. De Gruyter.
- Zima, E. (2017a). Multimodal constructional resemblance: the case of English circular motion constructions. In R. de Mendoza Ibáñez, F. A. Luzondo Oyón, & P. P. Sobrino (eds), *Constructing families of constructions* (pp. 301–337). <https://doi.org/10.1075/hcp.58.11zim>
- Zima, E. (2017b). On the multimodality of [all the way from X PREP Y]. *Linguistics Vanguard*, 3(1). <https://doi.org/10.1515/lingvan-2016-0055>
- Zlatev J. (2019) ‘Mimesis theory, learning and polysemiotic communication’. In: M. Peters (ed.), *Encyclopedia of educational philosophy and theory*. Springer
- Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B*, 375(1789), 20190062.
- Zuidema, W., & de Boer, B. (2018). The evolution of combinatorial structure in language. *Current Opinion in Behavioral Sciences*, 21, 138–144.