Language as a multimodal phenomenon: implications for language learning, processing and evolution

Gabriella Vigliocco, Pamela Perniss and David Vinson

References	This article cites 49 articles, 5 of which can be accessed free http://rstb.royalsocietypublishing.org/content/369/1651/20130292.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections behaviour (542 articles) cognition (365 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Introduction

Author for correspondence:
Gabriella Vigliocco
e-mail: g.vigliocco@ucl.ac.uk

†Present address: Cognitive, Perceptual and Brain Sciences Department, 26 Bedford Way, London, WC1H 0AP.

Language as a multimodal phenomenon: implications for language learning, processing and evolution

Gabriella Vigliocco[1,2,†], Pamela Perniss[1,2] and David Vinson[1]

[1]Cognitive, Perceptual & Brain Sciences Department, 26 Bedford Way, London WC1H 0AP
[2]Deafness, Cognition & Language Research Centre, 49 Gordon Square, London WC1H 0PD

Our understanding of the cognitive and neural underpinnings of language has traditionally been firmly based on spoken Indo-European languages and on language studied as speech or text. However, in face-to-face communication, language is multimodal: speech signals are invariably accompanied by visual information on the face and in manual gestures, and sign languages deploy multiple channels (hands, face and body) in utterance construction. Moreover, the narrow focus on spoken Indo-European languages has entrenched the assumption that language is comprised wholly by an arbitrary system of symbols and rules. However, iconicity (i.e. resemblance between aspects of communicative form and meaning) is also present: speakers use iconic gestures when they speak; many non-Indo-European spoken languages exhibit a substantial amount of iconicity in word forms and, finally, iconicity is the norm, rather than the exception in sign languages. This introduction provides the motivation for taking a multimodal approach to the study of language learning, processing and evolution, and discusses the broad implications of shifting our current dominant approaches and assumptions to encompass multimodal expression in both signed and spoken languages.

1. Language studies: the current focus, approaches and assumptions

Current theories in linguistics, psychology and cognitive neuroscience have been developed largely from the investigation of spoken languages. More precisely, theories tend to be based on only a few spoken languages, primarily English and other Indo-European languages. It is therefore critical to ask whether those properties that have been described and assumed to be foundational properties of language—and which have thus defined our theories of language—might, rather, be linked to linguistic properties especially salient in only this handful of the world's languages.

This narrow lens on language has led to the widely accepted approaches and assumptions about *what* language is and *how* language is structured that we challenge here. First, we challenge the assumption that language is sufficiently investigated as speech or text, and second, we challenge the assumption that language is a wholly arbitrary system. A major consequence of the first assumption is the deeply entrenched distinction between what we may call *language proper*, i.e. language as a structured system amenable to linguistic analysis, and *communication*, i.e. the broader context of language use, which includes the use of other channels of information (e.g. co-speech gesture, prosody). The majority of language studies have been firmly focused on language proper, to the exclusion of context and multimodal expression that contribute to utterance and meaning construction. A major consequence of the second assumption is our conception of language processing and language development as an abstract and symbolic system, where linking between linguistic and conceptual levels is a process of transduction from linguistic symbols to cognitive representations only arbitrarily linked [1]. Below, we discuss the consequences of these two assumptions, which we see as arising from an excessively narrow lens on language as the object of study, in more detail (§1a,b). We then, in §2, introduce a thought experiment: what if the

study of language had started with the study of signed language rather than spoken language? In general, language studies and our theories of language, defined and moulded by structures salient in Indo-European spoken languages, have largely ignored components of language that are immediately obvious and highly salient in sign languages, namely the multimodal nature of language and the iconicity of language. If these features of language had been instrumental in determining the course of language research from the beginning, then our dominant ideas about language processing, language development and language evolution, and the relationship between language and cognition more generally, might be very different. All contributions to this issue, as we outline in §3, reflect work that addresses either the multimodal nature of language (§3*a*) or the iconicity of language (§3*b*), thus in effect approaching language from the perspective of features of language that are salient in sign languages and escaping the constraints and consequences of a narrow-lens view of language. We provide concluding remarks in §4.

(a) Narrow-lens view of language 1: language as speech/text

Language research has focused predominantly on speech and/or text, thus ignoring the wealth of additional information available in face-to-face communication, leading to the (explicit or implicit) assumption that the object of investigation—language—can be properly and sufficiently addressed by ignoring other characteristics of face-to-face interactions: the communicative context in which language has evolved, in which it is learnt by children, and in which it is most often used. Language proper is thereby defined as a rule-governed system characterized by the concatenation of morphological/lexical units (as is evident in speech or text) that can be isolated from other aspects of communicative behaviour that are present in face-to-face contexts. Those other aspects of behaviour are often explicitly labelled 'non-verbal communication' and are typically pursued independently from the study of language (see [2]).

Decades of psycholinguistic research have used acoustic presentation of spoken words, or visual presentation of written words, to study language and to develop theories about language processing and acquisition. Even where the contribution of a secondary source of information, such as visual information from the face, has been widely recognized (as in the McGurk effect [3]), the import of such additional information has been traditionally considered to be limited to supporting the acoustical analysis. However, recent evidence suggests that the integration of information from so-called secondary sources, including face movements and gestures, may be an integral part of language processing and play a critical role in the language acquisition trajectory. In development, gestures predict learning stages, both in vocabulary and conceptual development. For example, gestures used at an early age (14 months) predict vocabulary size at a later age [4], and the production of supplementary speech–gesture combinations (e.g. 'eat' + point at cookie) predicts the productions of two-word utterances ('eat cookie') [5]. In addition, the nature of co-speech gesture and speech combinations has been shown to index changes in conceptual knowledge, for example, in children's understanding of conservation tasks [6] or arithmetic [7]. Regarding processing, there is evidence from both production and comprehension that co-speech gestures are tightly integrated with speech, such that the form of gestures is influenced by the typological structure

of the accompanying speech [8] and that information conveyed in gestures is automatically integrated with information conveyed in speech in comprehension [9]. In addition, gestures without speech have been shown to evoke N400 effects, an EEG component which has been linked to semantic and especially integration processes during word and sentence comprehension [10] suggesting that gesture comprehension invokes semantic processes similar to those engaged in the processing of words [11]. This evidence is important here because it indicates that speech and gesture are part and parcel of the same system and together constitute a tightly integrated processing unit, thus underscoring the need for a multimodal approach to the study of language.

Furthermore, in our literate societies, the distinction between language proper and communication is often based on the difference between language as it is written down (using well-formed, grammatical sentences) and language as it is used in actual interaction (including other channels of information like gesture and prosody). The basis for this distinction, however, is often not given, as, for example, in spoken languages that do not have an associated written form or crucially in signed languages. Signed language necessarily occurs in contexts of face-to-face communication and involves the use of the hands as major articulators in addition to multiple other non-manual channels of expression (face, mouth, eyebrows and body) [12–15].

However, even recognizing the inherent use of different channels of expression in sign language structure, the main concern in the linguistic description of sign languages has often remained focused on being able to describe sign languages in terms of the same linguistic and grammatical structures and constraints familiar from spoken languages. In this context, the use of different channels—even for grammatical purposes—is described as a modality effect. Thus, this analysis of sign languages implicitly preserves the assumption that language can be distinguished from other aspects of communication.

(b) Narrow-lens view of language 2: language as arbitrary

The second major consequence of focusing on Indo-European spoken languages has been to characterize the link between linguistic form and meaning as solely arbitrary. The idea of an arbitrary connection between form and meaning (commonly associated with Saussure [16]) was already argued for by Locke, in his *Essay concerning human understanding* [17]. His argument was that the existence of different (spoken) languages (with very different words for the same objects) is evidence against the idea of there being any natural connection between linguistic form and meaning. Because everyone perceives the world in the same way, there should be only one human language, if properties of objects could determine the names given them by means of natural connections.

Current approaches to the study of language development, processing and its neural underpinnings are based on the assumption that convention alone determines the relationship between form and meaning. Indeed, if we look at the lexicon of English (or that of other Indo-European languages), the idea that the relationship between a given word and its referent is defined by an arbitrary connection alone seems entirely reasonable. For example, there is nothing in the sequence of sounds in the English word *house* that indicates its meaning of 'a building for human habitation'. Moreover, the assumption

3

that arbitrary form–meaning mappings define language is consistent with, and we would argue the source of, the idea that language is a wholly symbolic system, the elements of which are manipulated on an abstract level of representation [18].

Non-arbitrary mappings, coming from domains such as onomatopoeia, are often dismissed as unimportant because they are considered to be very limited. Yet, numerous non-Indo-European spoken languages include wide repertoires of iconic mappings, variously described as mimetic, ideophonic or sound–symbolic (e.g. sub-Saharan African languages, Australian Aboriginal languages, Japanese, Korean, Southeast Asian languages, indigenous languages of South America, and Balto-Finnic languages; see [19] for references). In these languages, iconicity is achieved by the systematic association of properties of vowels and consonants to properties of experiences. These mappings extend to a wide range of domains, including sensory, motor and affective experiences as well as aspects of the spatio-temporal unfolding of an event.

Recent research on spoken language has shown that iconicity expressed in the speech signal influences language processing and development. Both adults and children have been shown to make reliable associations between properties of consonants and vowels and visual features of referents, e.g. *bouba* and *kiki* judged to correspond to a round, curvy versus jagged, pointy shape, respectively [20–22]. Iconic mappings have also been shown to be facilitatory in studies using indirect measures of online processing, including reaction times [23,24] and EEG waveforms [25,26], as well as to facilitate language acquisition in both children and adults [27,28]. Finally, prosody, or the supra-segmental modulation of the acoustic signal, constitutes another channel of expression in which iconic mappings may be expressed. For example, there is evidence that prosodic variations in pitch and amplitude can reliably convey information related to specific semantic domains (e.g. big/small, hot/cold) [29].

Sign languages, produced with the signers' body and perceived visually, afford a particularly high degree of iconic representation, reflected in the large repertoire of iconic forms in the lexicon as well as at the sentential and discourse level. Historically, iconicity has not been assumed to play any role in the processing and acquisition of sign language, nor in the neural organization of networks supporting sign language processing. Recent work, however, suggests that iconicity affects semantic processing [30,31], facilitates lexical retrieval in production [32] and affects language comprehension [32,33]. Such effects may be limited, however, to tasks where semantic activation is necessary [34,35].

With regards to language development, Orlansky & Bonvillian [36] reported no difference in acquisition between iconic and non-iconic signs by native signers learning American Sign Language (ASL) as their first language. However, this study did not control for other variables that might affect age of acquisition (such as familiarity and motoric/phonological complexity) and questions have been raised about their criteria for considering signs as being iconic [37]. Developmental data on a much larger scale do suggest a role of iconicity in learning British Sign Language (BSL). Using data from the communicative development inventory for BSL, Thompson *et al.* [38] showed that deaf children (aged 11–30 months) acquiring BSL natively produced and comprehended more iconic than non-iconic signs. In contrast to previous studies, this study used normative data for iconicity (operationalized as ratings by native signers [39]) and specifically assessed whether iconicity had a role above and beyond other relevant variables.

The study of language as an essentially arbitrary system represented primarily as speech or text has shaped all our current theories of language development and processing. We have argued above that language (both spoken and signed) should be more appropriately characterized as multimodal and iconic. If we had studied language from the start as a system that embeds iconicity and that conveys meaning in multiple channels of expression—features of language that are immediately obvious in signed language—our dominant theories of language may have developed along quite different trajectories. We explore the implications of this thought experiment in §2.

2. What if the study of language had started from signed language rather than spoken language?

The thought experiment expressed in this question sets to challenge the traditional approach to language that has arisen from the central assumptions about language that we have just outlined. First, if the study of language had started from signed language rather than spoken language, would we have thought of language as a phenomenon that could be suitably and sufficiently represented and investigated as only speech or text? Likely, we would have taken language as inherently multimodal and would have considered speech or text only as atypical or an impoverished representation of language. Second, if the study of language had started from signed language rather than spoken language, would we have thought of language as solely arbitrary, with form–meaning mappings determined by convention alone? Likely, we would have taken language as both an arbitrary and an iconic system, with iconicity contributing to language processing and development. Indeed, as reviewed above, the recent evidence from both signed and spoken languages suggests that iconicity plays a role in language processing and development and that language processing obligatorily integrates information from context and visual channels of expression. These findings, highly controversial in current approaches, would not be at all surprising, rather they would have been taken as foundational if the study of language had started from signed languages rather than from spoken languages.

As noted above, the study of signed language has, since its beginnings in the 1960s, taken the basic claims developed for spoken languages as basic assumptions. The need to prove the status of sign languages as fully-fledged natural human languages meant proving the existence of structures and categories in sign languages equivalent to those in spoken languages in all respects [40–43]. Modality differences were invoked when these structures and categories seemed to differ ([44–46], but see [47]), but the fundamental theoretical assumptions remained intact. However, it may well be the case that if the study of signed languages did not have to carry this baggage, it could likely have led to different approaches to and assumptions about the study of language (development, processing and evolution). In particular, language might not have been decontextualized from its use in face-to-face communication, where multiple channels converge and contribute to the meaning being conveyed, and the presence of iconic form–meaning mappings, along with arbitrariness, might have been taken as a foundational assumption with respect to vocabulary, language and processing structures.

3. The Theme Issue: rationale and road map

The motivation for this Theme Issue is to engage the community working on language from different disciplines (linguistics, psychology, neuroscience and anthropology) in our thought experiment: 'What if the study of language started from signed rather than spoken languages?'. The contributors provide theoretical arguments as well as evidence for why we should or should not: (i) embrace a multimodal approach to language which does not pose a strong divide between language proper and communication and in which meaning is derived by the integration of the different channels of information, and (ii) include iconicity, along with arbitrariness, among the foundational properties of language, and understand the role each plays in language development and processing. The issue is structured around these two main themes, with papers addressing each theme from the perspective of the acquisition, processing or evolution of language.

The contribution by Kendon [48] sets the stage for the issue by providing an historical overview of the study of language, putting into relief how the divide between language and communication became entrenched and how language research came to be dominated by a formal, structuralist ethos that focused on the analysis of (spoken, Indo-European) language as a self-contained internally structured system. Kendon illustrates the shortcomings of such an approach, in particular, with respect to the analysis of sign languages and the history of sign language research, convincingly arguing that signed language cannot be appropriately described with models borrowed from structural and formal spoken language linguistics. Kendon sets this concept of language against a view that sees language as part of a larger construct of human communication conduct. In this wide-angle lens view, language is something that is engaged in and constructed, comprising contextual cues and visible action resources (which afford a high potential for iconicity) available to both signers and speakers. As such, and as Kendon clearly elucidates, an overarching concept and understanding of language—for linguistic pursuits as well as for the psychology and neurobiology of language and for cognitive science, more generally—that encompasses both signed and spoken language requires a multimodal approach to language that dispenses with the idea that language consists only of linguistic units expressed in speech or sign.

(a) A multimodal approach to language

Perhaps the most compelling argument for pursuing a multimodal approach to language, including multiple concomitant channels of expression (i.e. gesture, prosody, facial expression and body movement), is that to understand language, the object of study needs to be brought into line with its predominant manifestation as a system of communication in face-to-face interaction. It is in this manifestation that language is learnt by children and it is in this form that it has evolved. Two contributions demonstrate the importance of such a wide-angle view from the perspective of language acquisition and emergence. Goldin-Meadow [49] makes a comprehensive plea for widening the lens on language to include the manual modality, showing how gesture plays a role in learning in both spoken and signed languages as well as how gesture can come to assume the forms and functions of fully fledged language when children are not exposed to a language model, as happens with deaf homesigners. The paper first reviews evidence showing that

the use of gestures by hearing children precedes and predicts the acquisition of structures in speech. Goldin-Meadow goes on to show that the use of gesture accompanying speech and signs continues to promote learning in older children, suggesting that the power of gesture in learning lies in its ability to offer another representational format, i.e. an analogue format, to the categorical information encoded in either the speech of spoken languages or the signs of signed languages. The contribution by Liszkowski [50] looks at communication in infants before they start to use spoken language forms and demonstrates the vital role of multimodal information in structuring infant comprehension and production in pre-linguistic communicative contexts. Liszkowski first reviews evidence that infants are sensitive to common ground and use information from preceding action contexts in their communication and then shows that infants are also able to systematically extract meaning from multimodal cues in the communicative act itself, including prosody, posture and gesture, independent of situational information. Together, the contributions by Goldin-Meadow and Liszkowski provide strong evidence for the use of multimodal cues in language acquisition, especially in shaping the language learning trajectory.

The next two contributions in the issue illuminate the inherently multimodal nature of language from the perspective of language processing, providing both behavioural and neurobiological evidence. Özyürek [51] focuses on the semantic and temporal integration of information from speech and iconic gestures in spoken language comprehension. The paper provides clear evidence for the tight integration and interaction between vocal and visual channels in processing, even showing that the brain's neural responses are similar for processing speech and iconic gestures. The review also demonstrates the context sensitivity of the interaction between the two channels, showing that the level of integrated processing can be modulated by pragmatic knowledge, by the communicative context and by the communicative intent of the speakers. The contribution by Skipper [52] takes a novel perspective on demonstrating the multimodal nature of language by showing that hearing itself is deeply multimodal. By looking at activity in the auditory cortex in meaningful linguistic versus non-meaningful auditory contexts as well as in speech-only versus speech and gesture contexts, Skipper shows that the auditory cortex is less active in multimodal and more meaningful contexts, suggesting that our brain constructs meaning primarily predictively, using information from any kind of context—auditory or visual—to generate predictions.

The question of language evolution is addressed in the contribution by Levinson & Holler [53]. The authors argue for a stratified accumulation of human communicative capacities, rooted in the gestural ritualization of action sequences and turn-based dyadic interaction. According to this scenario, complex vocalization would have been a late addition to the communicative repertoire, requiring the development of voluntary breathing control, and would have complemented an existing system of deictic and iconic gestural communication. However, with a coevolution dating back nearly a million years, vocal and gestural modalities are deeply entwined in human communication, as is reflected in the default multimodal nature of modern human communication. We also include in the issue a fundamentally different perspective on the evolution of language. Sereno [54] elaborates a scenario in which the vocal modality, and the capacity for complex vocalization, has primacy in the evolution of language, with the gestural modality

instead coming into the picture as a later addition. The suggestion is that language evolved from complex birdsong-like vocalizations that were in place in the hominid line initially for purposes of sexual selection and were then taken over for symbolic communicative purposes.

(b) The iconicity of language

The second half of the issue is dedicated to papers that focus on the iconicity of language and its role in language development, processing and evolution. If the multimodal nature of language is recognized, then iconicity becomes visible across all languages as expressed in different channels. As discussed above, iconicity may not be as visible in the lexicon of Indo-European spoken languages (e.g. English) as it is in other languages (e.g. Japanese). However, for iconicity to have any weight in accounting for any critical processes in development, processing and evolution and for iconicity to thus be viable as a foundational assumption for language studies, it should be possible to show that it plays a role across languages, even languages with seemingly little lexical iconicity. Monaghan *et al.* [55] explore the possibility that although clear iconic mappings can be found only for onomatopoeia in English, more subtle statistical cues may nonetheless be distributed in the lexicon. In a large-scale analysis of phoneme–meaning correspondences, these authors show that there are small but significant correlations and that these correlations are stronger for words acquired earlier. This finding underscores the plausibility of iconicity playing a role in language development, and suggests that the smaller vocabularies at early stages of language acquisition may be more tolerant of non-arbitrary form–meaning mappings that may promote word learning.

The role of iconicity in word learning is explored in more depth in the contribution by Imai & Kita [56]. The authors provide a detailed hypothesis and supporting evidence for why sound–meaning mappings (sound–symbolism) would play a pivotal role in language development. According to the 'sound–symbolism bootstrapping hypothesis', sound symbolism would help the child understand that perceived sounds refer to things in the world and would help them zero in on specific form–meaning mappings. Imai and Kita also explore the more general question of why iconicity is present in language at all, suggesting that sound symbolism is a vestige of protolanguage and thus supporting the idea that iconicity was important in language evolution.

In order to be able to understand the role of iconicity in language structure, processing and development, it is necessary to have a cognitive framework for explaining iconicity effects. Taking sign languages as a starting point, where iconic mappings are readily visible, Emmorey [57] suggests that structure-mapping theory [58,59] provides such a framework.

Here, iconicity is a structured mapping between two mental representations, and the theory provides a general mechanism for iconic mappings, which crucially allows for spelling out constraints and making concrete predictions about how iconicity would be used in processing and development.

The final contribution in the issue by Perniss & Vigliocco [60] is similarly concerned with providing a clear definition of the concept of iconicity and offering mechanistic accounts of how iconicity may emerge. Like Emmorey [57], the authors advocate moving away from the treatment of iconicity as a monolithic concept, differentiating between more abstract processes of structural alignment that would be involved in establishing more abstract, indirect iconic relationships and more basic and direct processes of alignment based on imitation and visual overlap. Perniss and Vigliocco take a broad perspective and offer an overarching, unified view of iconicity as playing a fundamental role in language (both spoken and signed), across evolution, development and processing. The paper offers a similar account to the one provided by Imai & Kita [56] for how iconicity could benefit language development, and presents a novel perspective on how iconicity would realize *embodiment* of language for adult language users (as the key to coactivation of linguistic and sensory–motor systems) and how iconicity would have played a role in language evolution as supporting *displacement*, namely the ability of language to refer to what is not immediately present.

4. Conclusion

If the study of language were to have started from signed rather than from spoken languages, then the multimodal and iconic nature of language would have been taken as part of the linguistic phenomena to explain. This Theme Issue provides a range of views on how such a change would have affected our understanding of language development, processing and evolution. Furthermore, it provides the evidence base that underscores the plausibility of such a change and highlights future questions and directions for the study of language.

At a more general level, widening the lens on our object of study to include multimodal communication brings a level of *ecological validity* to the scientific investigation of language that is much needed and currently argued for by scholars from different fields [61,62].

References

1. Barsalou LW, Simmons WK, Barbey AK, Wilson CD. 2003 Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* **7**, 84–91. (doi:10.1016/S1364-6613(02)00029-3)

2. Manusov V, Patterson ML (eds). 2006 *The Sage handbook of nonverbal communication.* Thousand Oaks, CA: Sage Publications.

3. McGurk H, MacDonald J. 1976 Hearing lips and seeing voices. *Nature* **264**, 746–748. (doi:10.1038/264746a0)

4. Rowe M, Goldin-Meadow S. 2009 Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science* **323**, 951–953. (doi:10.1126/science.1167025)

5. Özçalışkan Ş, Goldin-Meadow S. 2005 Do parents lead their children by the hand? *J. Child Lang.* **32**, 481–505. (doi:10.1017/S0305000905007002)

6. Church RB, Goldin-Meadow S. 1986 The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* **23**, 43–71. (doi:10.1016/0010-0277(86)90053-3)

6

rstb.royalsocietypublishing.org   *Phil. Trans. R. Soc. B* **369**: 20130292

7. Goldin-Meadow S, Cook SW, Mitchell ZA. 2009 Gesturing gives children new ideas about math. *Psychol. Sci.* **20**, 267–272. (doi:10.1111/j.1467-9280.2009.02297.x)

8. Kita S, Özyürek A. 2003 What does cross-linguistic variation in semantic co-ordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* **48**, 16–32. (doi:10.1016/S0749-596X(02)00505-3)

9. Kelly SD, Özyürek A, Maris E. 2010 Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* **21**, 260–267. (doi:10.1177/0956797609357327)

10. Kutas M, Federmeier KD. 2011 Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann. Rev. Psychol.* **62**, 621–647. (doi:10.1146/annurev.psych.093008.131123)

11. Wu YC, Coulson S. 2005 Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology* **42**, 654–667. (doi:10.1111/j.1469-8986.2005.00356.x)

12. Sandler W. 2009 Symbiotic symbolization by hand and mouth in sign language. *Semiotica* **174**, 241–275. (doi:10.1515/semi.2009.035)

13. Sandler W, Lillo-Martin D. 2006 *Sign language and linguistic universals*. Cambridge, UK: Cambridge University Press.

14. Zeshan U. 2004 Interrogative constructions in signed languages: crosslinguistic perspectives. *Language* **80**, 7–39. (doi:10.1353/lan.2004.0050)

15. Pfau R. 2008 The grammar of headshake: a typological perspective on German sign language negation. *Linguist. Amsterdam* **1**, 37–74.

16. Saussure F. de. 1916 *Cours de linguistique générale*. Paris, France: Payot.

17. Locke J. 1975 *An essay concerning human understanding*. Oxford, UK: Clarendon Press. Original work published 1689.

18. Levelt WJM, Roelofs A, Meyer AS. 1999 A theory of lexical access in speech production. *Behav. Brain Sci.* **22**, 1–75. (doi:10.1017/S0140525X99001776)

19. Perniss P, Thompson RL, Vigliocco G. 2010 Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* **1**, 227. (doi:10.3389/fpsyg.2010.00227)

20. Köhler W. 1929 *Gestalt psychology*. New York, NY: Liveright.

21. Maurer D, Pathman T, Mondloch CJ. 2006 The shape of boubas: sound–shape correspondences in toddlers and adults. *Dev. Sci.* **9**, 316–322. (doi:10.1111/j.1467-7687.2006.00495.x)

22. Ramachandran VS, Hubbard EM. 2001 Synaesthesia: a window into perception, thought and language. *J. Conscious. Stud.* **8**, 3–34.

23. Ohtake Y, Haryu E. 2013 Investigation of the process underpinning vowel-size correspondence. *Jpn Psychol. Res.* **55**, 390–399. (doi:10.1111/jpr.12029)

24. Westbury C. 2005 Implicit sound symbolism in lexical access: evidence from an interference task. *Brain Lang.* **93**, 10–19. (doi:10.1016/j.bandl.2004.07.006)

25. Kovic V, Plunkett K, Westermann G. 2010 The shape of words in the brain. *Cognition* **114**, 19–28. (doi:10.1016/j.cognition.2009.08.016)

26. Kanero J, Imai M, Okuda J, Okada H, Matsuda T. 2014 How sound symbolism is processed in the brain: a study on Japanese mimetic words. *PLoS ONE* **9**, 1–8. (doi:10.1371/journal.pone.0097905)

27. Imai M, Kita S, Nagumo M, Okada H. 2008 Sound symbolism facilitates early verb learning. *Cognition* **109**, 54–65. (doi:10.1016/j.cognition.2008.07.015)

28. Kantartzis K, Imai M, Kita S. 2011 Japanese sound-symbolism facilitates word learning in English-speaking children. *Cogn. Sci.* **35**, 575–586. (doi:10.1111/j.1551-6709.2010.01169.x)

29. Nygaard LC, Cook AE, Namy LL. 2009 Sound to meaning correspondences facilitate word learning. *Cognition* **112**, 181–186. (doi:10.1016/j.cognition.2009.04.001)

30. Thompson RL, Vinson DP, Vigliocco G. 2009 The link between form and meaning in American sign language: lexical processing effects. *J. Exp. Psychol. Learn.* **35**, 550–557. (doi:10.1037/a0014547)

31. Vigliocco G, Vinson DP, Woolfe T, Dye MW, Woll B. 2005 Language and imagery: effects of language modality. *Proc. R. Soc. B* **272**, 1859–1863. (doi:10.1098/rspb.2005.3169)

32. Vinson D, Thompson RL, Skinner R, Vigliocco G. Submitted. A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language.

33. Thompson RL, Vinson DP, Vigliocco G. 2010 The link between form and meaning in British sign language: effects of iconicity for phonological decisions. *J. Exp. Psychol. Learn.* **36**, 1017–1027. (doi:10.1037/a0019339)

34. Baus C, Carreiras M, Emmorey K. 2013 When does iconicity in sign language matter? *Lang. Cogn. Proc.* **28**, 261–271. (doi:10.1080/01690965.2011.620374)

35. Bosworth R, Emmorey K. 2010 Effects of iconicity and semantic relatedness on lexical access in American sign language. *J. Exp. Psychol. Learn.* **36**, 1573–1581. (doi:10.1037/a0020934)

36. Orlansky M, Bonvillian JD. 1984 The role of iconicity in early sign language acquisition. *J. Speech Hear. Disord.* **49**, 287–292. (doi:10.1044/jshd.4903.287)

37. Lloyd L, Loeding B, Doherty J. 1985 Role of iconicity in sign acquisition: a response to Orlansky and Bonvillian (1984). *J. Speech Hear. Disord.* **50**, 299–301. (doi:10.1044/jshd.5003.299)

38. Thompson RL, Vinson DP, Woll B, Vigliocco G. 2012 The road to language learning is iconic: evidence from British sign language. *Psychol. Sci.* **23**, 1443–1448. (doi:10.1177/0956797612459763)

39. Vinson D, Cormier K, Denmark T, Schembri A, Vigliocco G. 2008 The British sign language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behav. Res. Methods* **40**, 1079–1087. (doi:10.3758/BRM.40.4.1079)

40. Klima ES, Bellugi U. 1979 *The signs of language*. Cambridge, MA: Harvard University Press.

41. Liddell SK. 1980 *American sign language syntax*. New York, NY: Mouton.

42. Lillo-Martin D, Klima ES. 1990 Pointing out differences: ASL pronouns in syntactic theory. In *Theoretical issues in sign language research* (eds SD Fischer, P Siple), pp. 191–210. Chicago, IL: University of Chicago Press.

43. Padden CA. 1990 The relation between space and grammar in ASL verb morphology. In *Sign language research: theoretical issues* (ed. C Lucas), pp. 118–132. Washington, DC: Gallaudet University Press.

44. Aronoff M, Meir I, Sandler W. 2005 The paradox of sign language morphology. *Language* **81**, 301–344. (doi:10.1353/lan.2005.0043)

45. Emmorey K. 2002 *Language, cognition and the brain*. Mahwah, NJ: Lawrence Erlbaum Associates.

46. Meier RP. 2002 Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In *Modality and structure in signed and spoken languages* (eds RP Meier, K Cormier, D Quinto-Pozos), pp. 1–25. Cambridge, UK: Cambridge University Press.

47. Liddell SK. 2003 *Grammar, gesture, and meaning in American sign language*. Cambridge, UK: Cambridge University Press.

48. Kendon A. 2014 Semiotic diversity in utterance production and the concept of 'language'. *Phil. Trans. R. Soc. B* **369**, 20130293. (doi:10.1098/rstb.2013.0293)

49. Goldin-Meadow S. 2014 Widening the lens: what the manual modality reveals about language, learning and cognition. *Phil. Trans. R. Soc. B* **369**, 20130295. (doi:10.1098/rstb.2013.0295)

50. Liszkowski U. 2014 Two sources of meaning in infant communication: preceding action contexts and act-accompanying characteristics. *Phil. Trans. R. Soc. B* **369**, 20130294. (doi:10.1098/rstb.2013.0294)

51. Özyürek A. 2014 Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Phil. Trans. R. Soc. B* **369**, 20130296. (doi:10.1098/rstb.2013.0296)

52. Skipper JI. 2014 Echoes of the spoken past: how auditory cortex hears context during speech perception. *Phil. Trans. R. Soc. B* **369**, 20130297. (doi:10.1098/rstb.2013.0297)

53. Levinson SC, Holler J. 2014 The origin of human multi-modal communication. *Phil. Trans. R. Soc. B* **369**, 20130302. (doi:10.1098/rstb.2013.0302)

54. Sereno MI. 2014 Origin of symbol-using systems: speech, but not sign, without the semantic urge. *Phil. Trans. R. Soc. B* **369**, 20130303. (doi:10.1098/rstb.2013.0303)

55. Monaghan P, Shillcock RC, Christiansen MH, Kirby S. 2014 How arbitrary is language? *Phil. Trans.*

R. Soc. B **369**, 20130299. (doi:10.1098/rstb. 2013.0299)

56. Imai M, Kita S. 2014 The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Phil. Trans. R. Soc. B* **369**, 20130298. (doi:10.1098/rstb.2013.0298)

57. Emmorey K. 2014 Iconicity as structure mapping. *Phil. Trans. R. Soc. B* **369**, 20130301. (doi:10.1098/ rstb.2013.0301)

58. Gentner D. 1983 Structure-mapping: a theoretical framework for analogy. *Wiley Interdiscip. Rev. Cogn. Soc.* **7**, 155 – 170. (doi:10.1207/ s15516709cog0702_3)

59. Gentner D, Markman AB. 1997 Structure mapping in analogy and similarity. *Am. Psychol.* **52**, 45 – 56. (doi:10.1037/0003-066X.52.1.45)

60. Perniss P, Vigliocco G. 2014 The bridge of iconicity: from a world of experience to the experience of

language. *Phil. Trans. R. Soc. B* **369**, 20130300. (doi:10.1098/rstb.2013.0300)

61. Engel AK, Maye A, Kurthen M, König P. 2013 Where's the action? The pragmatic turn in cognitive science. *Trends Cogn. Sci.* **17**, 202 – 209. (doi:10. 1016/j.tics.2013.03.006)

62. Zwaan R. 2014 Embodiment and language comprehension: reframing the discussion. *Trends Cogn. Sci.* **18**, 229 – 234. (doi:10.1016/j.tics.2014.02.008)

7

rstb.royalsocietypublishing.org *Phil. Trans. R. Soc. B* **369**: 20130292