

Holger Diessel

Where does language come from? Some reflections on the role of deictic gesture and demonstratives in the evolution of language

Abstract: This paper considers Arbib's hypothesis that (oral) language has its roots in gesture in light of recent research on demonstratives, joint attention, and deictic pointing (Michael Arbib. 2012. *How the brain got language: The Mirror System Hypothesis*. Oxford: Oxford University Press). It is argued that demonstratives provide an important link between gesture, discourse, and grammar that rests on their communicative function to coordinate the interlocutors' focus of attention. Combining evidence from linguistic typology and historical linguistics with evidence from research on social cognition, the paper argues that demonstratives constitute a universal class of linguistic expressions that are commonly used in combination with a deictic pointing gesture to establish joint attention, a cognitive phenomenon that is closely related to Arbib's notion of "complex imitation". No other class of linguistic expressions is so closely tied to the speaker's body and gesture than demonstratives. However, demonstratives are not only used to focus the language users' attention on concrete entities in the surrounding situation, they are also used to organize the information flow in discourse, which in turn underlies their frequent development into a wide range of grammatical markers, e.g. definite articles, third person pronouns, relative markers, complementizers, subordinate conjunctions, copulas, and focus markers. In this way, demonstratives provide an explicit link between gesture, imitation, and grammar that is consistent with Arbib's theory of the evolution of language.

Keywords: demonstrative, deixis, joint attention, grammaticalization, evolution of language

Holger Diessel: Friedrich-Schiller-Universität Jena, Institut für Anglistik, Ernst-Abbe-Platz 8, 07743 Jena, Germany. E-mail: holger.diessel@uni-jena.de

1 Introduction

What are the origins of human language? One hypothesis that has been proposed by many scholars is that oral language has its roots in gesture. Arbib's book *How*

the brain got language: The Mirror System Hypothesis casts new light on this view. Combining evidence from biology, psychology, and linguistics, it presents a comprehensive theory of the evolution of language that emphasizes the importance of gestural communication and embodiment for the rise of (oral) language(s).

The heart of Arbib's theory is the Mirror System Hypothesis, which can be seen as an alternative to the standard view of linguistic innateness. Challenging the assumption that language (evolution) rests on an innate language faculty, Arbib argues that modern languages evolved from gestural communication through cultural evolution; but this had a specific biological prerequisite – the brain had to be “ready” (biologically) to acquire language. It is the central hypothesis of the book that the “mirror neuron system for grasping”, which originally had nothing to do with communication let alone language, provided the biological basis for the emergence of language in *Homo sapiens* some 100,000 years ago. Specifically, Arbib argues that the mirror mechanism, which is involved in both the production and perception of manual action, notably grasping, gave rise to “complex imitation”, i.e. the ability to understand and imitate intentional behaviors, which in turn led to the development of gestural communication and finally to the emergence of oral language(s).

Like other theories of the evolution of language, the Mirror System Hypothesis is difficult to verify. There are no data to prove or falsify the proposed connections between mirror neurons, complex imitation, gestural communication, and the emergence of oral language(s). However, Arbib's theory is intriguing because it combines new insights from neurobiology with recent research on animal communication, social cognition, and grammaticalization, leading to a complex scenario of language evolution that is more consistent with what is known about the brain and language than many other theories on the origin of human language, notably the nativist theory of generative grammar.

The Mirror System Hypothesis stresses the importance of biology for the evolution of language but without assuming the existence of an innate language faculty. In Arbib's theory linguistic categories are not genetically prespecified but emergent from communication and information processing (e.g. “fractionation”). This is in accordance with recent usage-based research on grammar and language development (cf. Bybee 2010; Tomasello 2003). In the usage-based approach, language is seen as a dynamic system of fluid categories and emergent constraints that are constantly restructured and reorganized under the influence of “domain-general cognitive processes” (Bybee 2010), which do not only affect language but also other cognitive activities such as vision and thought (see Diessel 2011 for a review).

In what follows I will consider some recent (usage-based) research on demonstratives and grammaticalization in light of Arbib's hypotheses about the

evolution of language. I will argue that demonstratives provide an important link between gesture, communication, and grammar that rests on their communicative function to establish joint attention – a cognitive mechanism closely related to Arbib’s notion of complex imitation (for a more comprehensive treatment of the hypotheses outlined in this commentary see Diessel 1999a, 2003, 2006, 2012a, 2012b).

2 Demonstratives: syntactic function and meaning

What are demonstratives? In the linguistic literature, demonstratives are commonly defined as deictic expressions functioning as pronouns and determiners; but this definition is not without problems. To begin with, the morphosyntactic properties of demonstratives exhibit a great deal of cross-linguistic variation. There are languages like English and French in which demonstratives serve particular syntactic functions. In English, *this* and *that* are either used as pronouns, substituting for a nominal (cf. example 1a), or they are used as determiners, occurring in a particular syntactic slot of the noun phrase (cf. example 1b).

- (1) a. *Could you repeat **this**?*
 b. *Do you see **that** yellow bike over there?*

In French, the two uses are also morphologically distinguished: *celui-ci* ‘this (one)’ and *celui-là* ‘that (one)’ are independent pronouns (cf. example 2a) and *ce* ‘this/that.MASC’ and *cette* ‘this/that.FEM’ are determiners (cf. example 2b).

- (2) a. *J’aime **celui-ci** mieux que **celui-là**.*
 b. ***Ce** garçon et **cette** fille se connaissent très bien.*

Unlike English and French, Tuscaroro has demonstratives that do not pertain to a particular grammatical class (cf. Mithun 1987). The Tuscaroro demonstratives *hè:ní:kā* ‘this/these’ and *kyè:ní:kā* ‘that/those’ lack any nominal morphology and are syntactically much less constrained than the demonstratives of English and French. As can be seen in examples (3a–c), the order of demonstrative and noun is flexible in Tuscaroro and the demonstrative may even be separated from the noun by intonation and/or a pause.

- (3) a. ***hè:ní:kā**: áha:θ*
 that horse
 ‘that horse’

- b. *u[?]né:wa:k hè:ní:kǎ:*
ghost that
‘that ghost’
- c. *wá[?]thahá:hi:θ hè:ní:kǎ:, . . . ruya[?]kwáhehr*
it.met.it that he.body.carries
‘It met that dinosaur.’
(Mithun 1987: 184, 184, 186)

Mithun (1987) describes the Tuscararo demonstratives as free nominals that may occur in apposition to a noun, but strictly speaking they are particles that do not pertain to a particular grammatical class. In other languages, demonstratives are also used as adverbs (e.g. English *there*), presentatives (e.g. French *violâ*), or even as verbs (cf. Dixon 2003), indicating that from a crosslinguistic perspective demonstratives do not form a coherent grammatical category. In fact, a number of scholars have argued that genuine demonstratives are particles with no particular morphosyntactic features (cf. Brugmann 1904; Hopper 1991; Koenig 2012).

Semantically, demonstratives are commonly defined as (spatial) deictics, suggesting that they are part of a much larger semantic class of deictic expressions subsuming demonstrative pronouns and determiners (e.g. *this/that*), demonstrative/spatial adverbs (e.g. *here/there*), first and second person pronouns (e.g. *I/you*), temporal adverbs (e.g. *now/then/today/ago*), motion verbs (e.g. *come/go*), tense markers (e.g. *will/be going to*), and manner adverbs (e.g. *so, thus*). In addition, conjunctive adverbs, interjections, discourse markers, vocatives, and imperatives have been analyzed as deictics (cf. Diessel 2012a).

The term deixis has a long history in linguistics and there is no doubt that some of the above mentioned expressions have important properties in common; but they do not form a homogenous semantic class. According to Levinson (2004), deictic expressions are linguistic elements “with built-in contextual parameters” that need to be specified by aspects of the situational and/or discourse context; but this is a very general property of (many) linguistic expressions and not just a particular aspect of deictics.

What most deictic expressions have in common is that they presuppose a particular point of reference, i.e. the deictic center, also called the “origo” (cf. Bühler 1934). But the deictic center has very different properties with different types of expressions. The deictic center of demonstratives is determined by the speaker’s body (or the location of the speaker’s body) at the time of the utterance; but for most other deictic expressions the deictic center has a more abstract, temporal or textual interpretation that is completely independent of the speaker’s body.

What demonstratives typically encode is the relative distance between the deictic center, i.e. the speaker's body, and an object or location in the surrounding speech situation. For instance, English *here* and *this* indicate the location of an element near the deictic center and the demonstratives *there* and *that* refer to elements outside of this domain. The encoding of distance is characteristic of demonstratives but not sufficient to identify the referent (cf. Diessel 1999a, 2005; see also Kemmerer 1999). What is needed in addition to the indication of distance is information about the direction or angle between the deictic center and the intended referent. In what follows I argue that this information is commonly provided by non-verbal means of communication, notably by deictic gesture.

3 Demonstratives, deictic gesture, and joint attention

Like several other spatial expressions (e.g. *left/right*, *up/down*), demonstratives are interpreted in the context of a spatial frame of reference that is usually anchored by the speaker's bodily coordinates at the time of the utterance (Diessel 2012a). In face-to-face conversation, demonstratives are commonly accompanied by eye gaze and deictic pointing gestures that indicate the location of the referent relative to the speaker's body. The frequent combination of demonstratives and deictic pointing has been observed by many scholars (e.g. Brugmann 1904; Bühler 1934; Clark 1996; Erikson 2008; Levinson 2004). But a deictic pointing gesture is not just a guidepost for spatial orientation, it also serves to create what psychologists call a joint focus of attention (cf. Butterworth 1998; Eilan et al. 2005; Tomasello 1999).

Joint attention is a complex social and cognitive phenomenon, which is related to Arbib's notion of "complex imitation". Joint attention involves at least two intentional agents, i.e. actor and addressee (or speaker and hearer), who have to coordinate their attention in order to communicate. A cognitive prerequisite for the creation of joint attention (and complex imitation) is that actor and addressee are able to understand the communicative partner as an intentional and mental being who looks at the surrounding situation from his or her subjective perspective. In order to create a joint focus of attention, speaker and hearer must have at least a basic understanding of mental states and intentional behaviours, which in turn is a prerequisite for communication, social cognition, and language (cf. Butterworth 1998; Eilan et al. 2005; Tomasello 1999).

Deictic pointing is the most basic communication device that people of all cultures use to establish or manipulate joint attention (cf. Kita 2003). Like human

beings, chimpanzees may learn how to point when they have frequent interacts with humans; but, as Arbib (2012: 81–82) notes, the pointing activities of non-human primates are different from those of humans.

Bates et al. (1976, 1979) distinguish two basic types of pointing gestures: proto-imperatives and proto-declaratives (see also Butterworth 1998; Camaioni et al. 2004; Tomasello 1999). Proto-imperatives resemble reaching gestures – they are produced with the intention of obtaining an object; whereas proto-declaratives are used with the sole intention of sharing attention. The pointing gestures that some nonhuman primates produce in interaction with humans are proto-imperatives. When chimpanzees point they want to get something, usually food, and they have learned that pointing triggers a particular reaction in the human addressee providing them with food (cf. Tomasello and Call 1997). At the surface, proto-imperatives and proto-declaratives are similar; but proto-imperatives do not presuppose an understanding of mental states and intentions – they are goal-directed activities at the brink of true communication. In the context of Arbib's theory, proto-imperatives could be seen as semi-gestures at the transition between "grasping" and "protosign", i.e. between manual action and sign language.

Like chimpanzees, young children make common use of proto-imperatives; but in contrast to non-human primates infants also produce declarative pointing gestures. Some researchers observed that proto-imperatives precede proto-declaratives in child development (cf. Camaioni et al. 2004); but other researchers have found that they emerge approximately at the same time (cf. Lizskowski et al. 2004). The earliest pointing gestures children produce appear at around the first birthday. A few months earlier, infants begin to follow eye gaze and head movement, which has been interpreted as an early form of joint attention (cf. Carpenter et al. 1998). But the emergence of deictic pointing at the age of 12 months is the first strategy infants use to *create* a joint focus of attention, and a few months later they begin to talk. As Iverson and Goldin-Meadow (2005) have shown, there is a close connection between the appearance of gesture in infants and the onset of (oral) language.

Among the first words children produce in close connection with (deictic) pointing gestures are demonstratives. Clark (1978) reports that the demonstratives *this*, *that*, *here*, and *there* are among the few non-content words that English-speaking children use during the one-word stage, and Diessel (2006) observed that between the ages of 1;0 and 2;0 the demonstrative *that* (pronounced [dæt]) is often the most frequent word in corpora of spontaneous child language. In combination with a deictic pointing gesture, demonstratives allow the child to talk about any element in the surrounding situation without using particular lexical expressions. The gestural use of demonstratives provides a powerful mechanism

for the child to engage in verbal activities with a limited vocabulary. As children get older, they acquire a larger inventory of referential expressions that can occur in lieu of a demonstrative plus pointing gesture; that is, with age, language becomes more independent from gesture and situational cues. But demonstratives continue to play an important role in adult language and are crucially involved in the diachronic evolution of grammatical morphemes and constructions.

4 Demonstratives and the emergence of grammar

The traditional analysis of demonstratives as pronouns and determiners obscures their particular function and status in language. In the linguistic literature, demonstratives are commonly analyzed as grammatical markers on a par with auxiliaries, adpositions, and third person pronouns; but in Diessel (2006) I have argued that demonstratives constitute a unique class of items that have to be kept separate from both lexical expressions (i.e. content words) and closed-class grammatical morphemes (i.e. function words). In that paper I present four arguments why demonstratives are distinct from other function morphemes and should be regarded as a particular class:

- First, as pointed out above, in face-to-face conversation demonstratives need the support of deictic pointing gestures and/or eye gaze and body posture. No other class of linguistic expressions is so closely tied to the human body and associated with a particular type of gesture than demonstratives.
- Second, although young children tend to omit grammatical function morphemes, they begin to use demonstratives very early. As we have seen, the demonstratives *this* and *that* and *here* and *there* are among the first words English-speaking children learn and they are extremely frequent in early child language.
- Third, recent research in linguistic typology has emphasized the enormous amount of cross-linguistic variation. According to Evans and Levinson (2009) there are very few (non-trivial) aspects of language that are truly universal; but demonstratives exist in all languages across the world and are surprisingly similar in terms of their semantic features (i.e. the indication of distance) and their pragmatic functions (e.g. the creation of joint attention).
- Fourth, although grammaticalization researchers have argued that all closed-class function morphemes are ultimately derived from content words, notably from nouns and verbs (see below), there is no evidence that demonstratives are based on lexical expressions. In contrast to genuine grammatical markers, the deictic roots of demonstratives cannot be traced back to content words.

Together these four features characterize demonstratives as a unique class of linguistic expressions that are crucially distinct from both content words and other closed-class function morphemes. The particular status of demonstratives is, of course, a consequence of their communicative function to establish joint attention. Since demonstratives serve one of the most fundamental functions in communication, cognition, and language, they are included in the basic vocabulary of every language (cf. Diessel 2006).

What is more, demonstratives are also crucially involved in the diachronic development of grammatical morphemes and constructions (cf. Diessel 1999a, 1999b, 2006, 2012a). In their basic use, demonstratives refer to concrete objects or events in the physical world, but they are also commonly used with reference to linguistic elements in discourse. In fact, the discourse use of demonstratives is one of the most frequent strategies speakers use to make a sequence of sentences more coherent. In this use, the deictic center is transferred from the physical world, i.e. the speaker's body, to a particular position in the unfolding speech stream. Demonstratives that are used with text-internal reference express a referential link between the sentence (or noun phrase) in which they are embedded and a linguistic element of the preceding or subsequent discourse (cf. example 5a–b).

- (5) a. [*unintelligible speech*] *I couldn't hear you. Could you repeat **this**?*
 b. *It was raining. **That's** why we left early.*

The discourse use of demonstratives is based on their communicative function to establish joint attention. Like exophoric demonstratives (i.e. demonstratives referring to concrete elements in the surrounding situation), endophoric demonstratives (i.e. demonstratives referring to linguistic elements in discourse) function to manipulate the interlocutors' focus of attention; but the endophoric use does not involve the speaker's body, eye gaze, or gesture (cf. Diessel 2006, 2012b).

Starting from this disembodied use, demonstratives develop into a wide range of grammatical markers. Across languages, demonstratives provide a common historical source for definite articles, third person pronouns, relative pronouns, complementizers, conjunctions, copulas, and focus markers (cf. Diessel 1999b). Some of these markers can also be derived from lexical expressions (e.g. copulas, complementizers), but others are almost exclusively based on demonstratives (e.g. definite articles, third person pronouns).

That demonstratives provide a frequent diachronic source for grammatical markers has been well-known for a long time (cf. Brugmann 1904); but since current research on grammaticalization concentrates on the development of gram-

matical morphemes from lexical expressions, notably from nouns and verbs, it underestimates the importance of demonstratives for the diachronic evolution of grammar. In fact, some studies have argued that the grammaticalization of demonstratives can be subsumed under the grammaticalization of lexical expressions, assuming that demonstratives are ultimately based on content words (cf. Heine and Kuteva 2007: chapter 2; see also Figure 13-2, page 336 in Arbib 2012). But, as pointed out above, there is no evidence for this hypothesis. The deictic roots of demonstratives are generally so old that they cannot be linked to content morphemes, and the communicative function of demonstratives suggests that they are likely to have emerged very early in the evolution of language.

If this is correct, grammatical morphemes are derived from two major sources, demonstratives and lexical expressions, i.e. nouns and verbs. Interestingly, the two types of expressions have given rise to different types of grammatical markers. Demonstratives provide a common historical source for (third person) pronouns, determiners, and conjunctions indicating links across clause and intonation boundaries, whereas lexical expressions are commonly reanalyzed as auxiliaries, adpositions, and modal markers elaborating the meanings of adjacent content words. Of course, some grammatical markers can arise from both types of expressions (e.g. copulas, complementizers); but generalizing across the many developments that have been subsumed under the notion of grammaticalization, it seems fair to say that the grammaticalization of demonstratives gives rise to grammatical markers that are primarily used for the encoding of inter-clausal relationships (i.e. clause combining and reference tracking), whereas the grammaticalization of lexical expressions leads to grammatical markers for the encoding of intra-clausal relationships (i.e. case markers, adpositions, mood and epistemic markers, reflexive pronouns, and auxiliaries).

5 Conclusion

To conclude, the discovery of the mirror neuron system for grasping provided new (biological) evidence for the longstanding hypothesis that some central aspects of human cognition are grounded in bodily activities. If grasping and imitation are (biologically) related, it seems plausible that social cognition and communication have (some of) their roots in manual gesture. Arbib elaborates these ideas in a complex theory of language evolution that leads us all the way from grasping to grammar. In this commentary, I have reviewed some recent research on demonstratives that support some aspects of Arbib's theory. Specifically, I have argued that demonstratives constitute a unique class of expressions that speakers of all languages use in combination with pointing gestures to establish

joint attention, a cognitive phenomenon that underlies Arbib's notion of complex imitation. No other linguistic device is so closely associated with the body and gesture than demonstratives; but demonstratives are not only used to direct the interlocutors' attention to concrete entities in the outside world, they are also used to organize the information flow in discourse, which in turn leads to their development into grammatical markers. In this way, demonstratives provide an explicit link between gesture, imitation, and grammar that is consistent with Arbib's theory of language evolution.

References

- Bates, E., L. Camaioni & V. Voltera. 1976. The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly* 21. 205–226.
- Bates, E., I. Bretherton & L. Snyder. 1979. *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.
- Brugmann, K. 1904. *Demonstrativpronomina der Indogermanischen Sprachen*. Leipzig: Teubner.
- Bühler, K. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.
- Butterworth, G. 1998. What is special about pointing in babies? In F. Simion & G. Butterworth (eds.), *The development of sensory, motor and cognitive capacities in early infancy. From perception to cognition*, 171–190. Hove: Psychology Press.
- Bybee, J. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Camaioni, L., P. Perucchini & P. Bellagamba. 2004. The role of declarative pointing in developing a theory of mind. *Infancy* 5. 291–308.
- Carpenter, M., K. Nagell & M. Tomasello. 1998. *Social cognition, joint attention, and communicative competence from 9 to 15 months of age*. Oxford: Blackwell.
- Clark, E.V. 1978. From gesture to word: On the natural history of deixis in language acquisition. In J. S. Bruner & A. F. Garton (eds.), *Human growth and development*, 85–120. Oxford: Oxford University Press.
- Clark, H. H. 1996. *Using language*. Cambridge: Cambridge University Press.
- Diessel, H. 1999a. *Demonstratives. Form, function, and grammaticalization*. Amsterdam: John Benjamins.
- Diessel, H. 1999b. The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology* 3. 1–49.
- Diessel, H. 2003. The relationship between demonstratives and interrogatives. *Studies in Language* 27. 581–602.
- Diessel, H. 2005. Distance contrasts in demonstratives. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (eds.), *World atlas of linguistic structures*, 170–173. Oxford: Oxford University Press.
- Diessel, H. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17. 463–489.
- Diessel, H. 2011. Review article of *Language, usage and cognition* by J. Bybee. *Language* 87. 830–844.

- Diessel, H. 2012a. Deixis and demonstratives. In C. Maienborn, K. v. Heusinger & P. Portner (eds.), *An international handbook of natural language meaning*, 1–25. Berlin: Mouton de Gruyter.
- Diessel, H. 2012b. Buehler's two-field theory of pointing and naming and the deictic origins of grammatical morphemes. In T. Breban, L. Brems, K. Davidse & T. Mortelmans (eds.), *New perspectives on grammaticalization: Theoretical understanding and empirical description*, 35–48. Amsterdam: John Benjamins.
- Dixon, R. M. W. 2003. Demonstratives. A cross-linguistic typology. *Studies in Language* 27. 61–122.
- Eilan, N., C. Hoerl, T. McCormack & J. Roessler. 2005. (eds.), *Joint attention: Communication and other minds issues in philosophy and psychology*. Oxford: Oxford University Press.
- Eriksson, M. 2008. Referring as interaction: On the interplay between linguistic and bodily practices. *Journal of Pragmatics* 41. 240–262.
- Evans, N. & S. C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32. 429–48.
- Heine, B. & T. Kuteva. 2007. *The genesis of grammar. A reconstruction*. Oxford: Oxford University Press.
- Hopper, P. 1991. On some principles of grammaticalization. In E. C. Traugott & B. Heine (eds.), *Approaches to grammaticalization*. Vol. I, 17–35. Amsterdam: John Benjamins.
- Iverson, J. M. & S. Goldin-Meadow. 2005. Gesture paves the way for language development. *Psychological Science* 16. 367–371.
- Kemmerer, D. 1999. “Near” and “far” in language and perception. *Cognition* 73. 35–63.
- Kita, S. 2003. Pointing. A foundational building block of human communication. In S. Kita (ed.), *Pointing: Where language, culture, and cognition meet*, 1–8. Mahwah: Lawrence Erlbaum.
- Koenig, E. 2012. Le rôle des déictiques de manière dans le cadre d'une typologie de la deixis. *Bulletin de la Société de Linguistique de Paris* 107. 11–42.
- Levinson, S. C. 2004. Deixis and pragmatic. In L. Horn & G. Ward (eds.), *The handbook of pragmatics*, 97–121. Oxford: Blackwell.
- Lizskowski, U., M. Carpenter, A. Henning, T. Striano & M. Tomasello. 2004. 12-month-olds point to share attention and interest. *Developmental Science* 7. 297–307.
- Mithun, M. 1987. The grammatical nature and discourse power of demonstratives. *Berkeley Linguistics Society* 13. 184–194.
- Tomasello, M. 1999. *The cultural origins of human cognition*. Cambridge: Harvard University Press.
- Tomasello, M. 2003. *Constructing a language. A usage-based approach*. Cambridge: Harvard University Press.
- Tomasello, M. & J. Call. 1997. *Primate cognition*. Oxford: Oxford University Press.