Wilcox, Sherman 2007. Routes from gesture to language. In: Elena Pizzuto, Paola Pientrandrea and Raffaele Simone (eds.), *Verbal and Signed Languages: Comparing Structures, Constructs and Methodologies*, 107–131. Berlin: Walter de Gruyter.

Wundt, Wilhelm 1921. *Völkerpyschologie: Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte.* Erster Band. Die Sprache. Leipzig, Germany: Engelmann.

Cornelia Müller, Frankfurt (Oder), Germany
Jana Bressem, Chemnitz, Germany
Silva H. Ladewig, Frankfurt (Oder), Germany

46. Towards a unified grammar of gesture and speech: A multimodal approach

1. Introduction
2. Prerequisites for a multimodal approach to grammar
3. Grammar and the concept of multimodality
4. Kinesthemes as syntactic units: Processes of typification and semantization
5. Syntactic structures as code manifestation of the language faculty: Constituency and recursion in co-speech gestures
6. Syntactic functions as a device for code integration within single languages: Multimodal attribution in German noun phrases
7. Conclusion: Why we need a multimodal approach to grammar
8. References

Abstract

This chapter argues for a multimodal approach to grammar (Bressem 2012; Fricke 2008, 2012; Harrison 2008, 2009; Ladewig 2012) and offers a sketch of the theoretical foundations according to Fricke (2012). Two main research traditions in linguistics are considered: generative grammar and linguistic structuralism and functionalism. The enterprise of a multimodal grammar is substantiated by the analyses of typification and semantization of gestures as potential syntactic constituents, by giving the rules of a generative phrase structure grammar of co-speech gestures which displays recursion and self-embedding, and by the grammatical analysis of multimodal attribution in German noun phrases. If we conceive of multimodality as a global dimension of linguistic and semiotic analysis which is generally applicable to language and other systems of signs, then we have to broaden our perspective by also including grammars of single languages and the human language faculty.

1. Introduction

Until recently, the idea that a multimodal approach to grammar is necessary was by no means evident. Most grammarians so far focus their grammatical analyses on written and spoken language without considering co-speech gestures. Yet the progress in

Müller, Cienki, Fricke, Ladewig, McNeill, Teßendorf (eds.) 2013, Body – Language – Communication (HSK 38.1), de Gruyter, 733–754

gesture studies offers a new perspective on the grammatical capacity of gestures accompanying speech (Bressem 2012; Fricke 2008, 2012; Harrison 2008, 2009; Ladewig 2012). Human speech is not only composed of articulations of the mouth, primarily perceived by ear, but also of visible articulations of other body parts affecting the eye (e.g., Kendon 2004; McNeill 1992, 2005). In this regard, the movements of the hands play a special role: the sign languages of the deaf show that movements of the hands alone can function as articulators of fully established languages (e.g., Wundt [1900] 1904, 1973). If it is the case that movements of the hand inherently have the potential for establishing a grammar, what are the grammatical implications of all those hands movements that accompany the speech of hearing people? Are single languages like German or English partially multimodal? How far is the faculty of language (Hauser, Chomsky, and Fitch 2002) bound to a particular mode of manifestation?

These are basic questions that link the enterprise of a multimodal approach to grammar to a long intermittent linguistic tradition of research affiliated above all with the names Wilhelm Wundt, Karl Bühler, Louis Hjelmslev, and Kenneth Pike. Karl Bühler ([1934] 2011) justifies a necessary combination of gestural pointing and verbal deictics with the argument that only in this way are speakers able to successfully refer to entities in certain utterance situations. Wundt ([1900] 1904, 1973), using the example of sign languages, demonstrates for the first time that the faculty of language is manifest in the visual-gestural mode. Hjelmslev ([1943] 1969), as well as Pike (1967), argues for a perspective of mode neutrality with respect to utterances as a whole because gestures can potentially instantiate structures and functions of language and speech. Therefore, according to this view, defining linguistic categories has to occur independently of any linguistic "substance" (Hjelmslev 1969). In this research tradition, two basic presumptions coexist in an unconnected way: first, the presumption that speech can manifest itself in various media and second, the presumption that subcodes, which may differ medially from the vocal language, can be integrated into a vocal matrix code. At first glance, both principles, manifestation and integration, seem to be incompatible. However, a deeper, more systematic approach facilitates a productive conceptual perspective: The phenomenon of the multimodality of language is to be tackled on different levels and from different linguistic perspectives. With respect to generative as well as structural-functional conceptions of grammar and language, the basic argumentation in the following sections pursues the goal of substantiating that co-speech gestures belong at least partially to the subject area of grammar.

2. Prerequisites for a multimodal approach to grammar

How do gestures and words interact on the level of grammar? In order to prove that the project of a multimodal grammar is at all feasible, it is necessary to show that co-speech gestures can be structurally and/or functionally integrated into grammars of single vocal languages or, alternatively, that co-speech gestures can be manifestations of the "faculty of language in the narrow sense" according to Hauser, Chomsky, and Fitch (2002) (Fricke 2008, 2012). This can only be proven if it is possible to show that gestures are capable of being typified and semantisized independently of the simultaneously accompanying vocal utterance. An important argument which is commonly held against co-speech gestures as potential units of the language system is their lack of conventionalization. Without conventionalization and segmentation there are no well-defined

linguistic units with stable form-meaning relations which might be capable of entering into syntactic constituent structures as constituents. How can we address this objection? The claim that movements of the hand are categorically not capable of instantiating linguistic structure and functions and, in a narrower sense, not capable of building morphemes is easily invalidated. Consider the sign languages of the deaf, which are fully developed linguistic systems possessing both a manual syntax and a manual lexicon. Gestures of the hearing may also be meaningful, comparable to morphemes in vocal utterances. These gestures with stable form-meaning relations are the so-called "emblematic gestures" (Efron [1941] 1972; Ekman and Friesen 1969). In section 4 we introduce the concept of kinesthemes as submorphemic units, which allows for modeling semiotic processes of typification and semantization and thereby provides terminal constituents for gestural constituent structures in section 5 (see also volume 2). This concept supports the assumption of a "rudimentary morphology" (Müller 2004: 3) as well as substantiating the category of "recurrent gestures" located between idiosyncratic and emblematic gestures in Kendon's continuum (e.g., Ladewig 2010, 2011, 2012; Müller 2008, 2010, submitted; Fricke, Bressem, and Müller volume 2). With regard to co-speech gestures, gesture scholars have so far neglected grammar and the syntactic dimension of analysis and linguists have largely considered gesture as "non-verbal" and as a phenomenon of language use only, excluding it from their subject. The crucial questions of the following sections are:

- How can linguistic multimodality be defined and distinguished from multimediality? (Section 3)
- To what extent can the faculty of language and the grammars of single vocal languages be considered as multimodal? (Section 4 to 6)
- Are co-speech gestures capable of being typified and semantisized independently of verbal utterances? (Section 4)
- Is it possible to analyze co-speech gestures independently of speech in terms of constituency? Do gestural constituent structures display recursion? If so, what would be the implications of gestural recursion for language theory? (Section 5)
- To what extent can gestures be integrated into verbal syntax, for example using syntactic functions? (Section 6)

Proving the possibility of typification for gestures is the prerequisite for the assumption of syntactic constituents that enter syntactic constituent structures. Proving the possibility of their semantization is the prerequisite for assigning the syntactic relation of modification in multimodal attribution in verbal noun phrases. On the basis of Eisenberg's grammar (1999), we show that co-speech gestures can fulfill syntactic as well as semantic attributive functions in German as a single language. This implies that they must be seen as part of the subject area of German grammar. With regard to the faculty of language, co-speech gestures can be assigned syntactic constituent structures that are recursive. According to the hypothesis that recursion is the defining criterion for the language faculty in the narrow sense (Hauser, Chomsky, and Fitch 2002) recursive co-speech gestures then have to be considered as an integral part of human language (see volume 2). Consequently, analyzing the grammar of single vocal languages as well as modeling the human language faculty require a multimodal approach (section 7).

3. Grammar and the concept of multimodality

3.1. Medium and modality

Every articulation necessarily requires a medium. However, media are not neutral with respect to the mediated subject but leave traces within it (Krämer 1998; Stetter 2005). Thus, the definition of the concept of "medium" is part of an indispensable basis of an epistemically reflected linguistics (Stetter 2005: 266). It is therefore advisable to begin with several conceptual distinctions with respect to the following basic questions: What is to be understood by the terms "multimediality" and "multimodality" in the field of language? To what extent can vocal language be considered as multimodal? What is the difference between multimediality and multimodality? How can these terms be defined sufficiently with respect to the objectives pursued in this article?

The basic idea can be characterized as follows: The concept of linguistic multimodality is not primarily based on the criterion of a simultaneous occurrence of different sensory impressions. Rather, with respect to a single language such as German, multimodality only exists if different media can adopt the same linguistic structures and/or functions. Examples can be found in the spoken and the gestural instantiation of the syntactic function of an attribute in a noun phrase (see section 6), and also in the substitution of a spoken constituent with a gestural constituent in the same syntactic position, which means that two different media can effect common linguistic structures. Furthermore, with respect to language in general, multimodality exists when the same structural principles manifest themselves simultaneously in different media. Examples of structural principles may be constituency and recursion (see section 5) or basic semiotic principles of sign constitution, for example, in processes of typification and semantization (see the parallels between phonesthemes and kinesthemes in section 4).

This means that multimodality, in contrast to multimediality, not only requires the simultaneity of at least two media but also either their structural and/or functional integration into a matrix code (code integration) or the manifestation of one and the same code in different media (code manifestation). It is the main thesis of this article that multimodal code integration and multimodal code manifestation occur not only on the level of language use but also on the level of the language system. In the next section we explicate the concepts of multimodality and multimediality from a linguistic and semiotic point of view. For this we will need a clear definition of "medium".

Within his semiotic approach to culture, Posner offers a definition that distinguishes between physical, technological, sociological, functional, and code-related media concepts (Posner 1986, 2004). Particularly important for our further discussion are the biological and the code-related concepts. According to Posner, the biological media concept relies on the sensory apparatus and characterizes sign processes according to the bodily organs (e.g., the ear or the eye) which are involved in the production and reception of signs. We then speak of auditory or visual media. The code-related media concept on the other hand "characterizes sign systems according to the types of rules by means of which the sign users manage to assign messages to the signs" (Posner 2004: 61). Single languages like English, French, Spanish or German are examples of code-related media.

3.2. Multimodality in gesture studies and language-image studies

The starting point of modern gesture studies is the basic idea that one and the same code (e.g., English) manifests itself in vocal utterances as well as co-speech gestures, with each belonging to a different sensory mode (biological media concept). This idea competes with the concept of nonverbal communication, which considers co-speech gestures as non-linguistic or "nonverbal" (e.g., Ruesch and Kees [1956] 1972; Scherer and Walbott 1984; Watzlawick, Beavin Bavelas, and Jackson 1967). According to Kendon (1980: 208), "speech and movement appear together, as manifestations of the same process of utterance", and McNeill (1985), with his article "So you think gestures are nonverbal?", initiated a debate in *Psychological Review*. He adopts Kendon's viewpoint, transfers it to psychology, and argues: "[...] gestures and speech are parts of the same psychological structure and share a computational stage" (McNeill 1985: 350).

As opposed to his earlier concept of multimodality, which he defined as code manifestation through two different sensory modes, Kendon (2004) later argues for a concept integrating two different codes or semiotic resources with reference to a communicative goal: "In creating an utterance that uses both modes of expression, the speaker creates an ensemble in which gesture and speech are employed together as partners in a single rhetoric enterprise" (Kendon 2004: 127). Obviously, Kendon recognizes this shift of focus because he asks: "Is this because they are expressions of two different forms of thought that originate jointly in a single, 'deeper' process? Or are they integrated as a consequence of how a person, engaged in producing an utterance, adapts two separate modes of expression and conjoins them in a single rhetorical aim?" (Kendon 2004: 3).

Kendon's later concept of multimodality parallels Kress' and Van Leeuven's ([1996] 2006), who conceive multimodality in the context of a general social-semiotic approach: "Mode is a socially shaped and culturally given resource for making meaning. *Image, writing, layout, music, gesture, speech, moving image, soundtrack* are examples of modes used in representation and communication. [...] Modes offer different potentials for making meaning; these have a fundamental effect on choices of mode in specific instances of communication" (Kress [2009] 2011: 54).

Like Kress and Van Leeuven, Stöckl conceives of "mode" as "code" (code-based media concept) (Stöckl 2004: 11), modeling multimodality as a networked system of modes and sub-modes: "Firstly, modes cut across sensory channels, so the nature of a sign is not sufficiently characterized by looking at its path of perception. Secondly, one mode can be realized in different media thus creating medial variants of one mode (e.g., speech and writing as variants of the linguistic mode)" (Stöckl 2004: 11). Stöckl, however, primarily addresses the language-image link in printed media. Accordingly, in his model gestures have the status of a marginal sub-mode. He points out that simple monomodal texts or monomodal vocal utterances have always been the exception, and states that we "seem to know more about the functioning of individual modes than about how they interact and are organized in text and discourse" (Stöckl 2004: 10).

If we do not want to constrain the notion of linguistic multimodality to specific types such as gesture-speech relations or language-image relations, then we have to broaden our perspective: towards single languages, towards language types, towards the faculty of language and towards language in general. Such an extension of the perspective on multimodality leads to the observation that we are dealing with a phenomenon resembling an optical illusion. On the one hand, the code of the language faculty can manifest

itself partially in two different "modalities", for example, in single vocal and single sign languages. On the other hand, gestures can be integrated structurally and/or functionally into the more dominant vocal matrix code of a single language. Therefore, we deal first with processes of code manifestation and second with processes of code integration. At first glance, analyses of language-image relations deal primarily with the principle of code integration while gesture-speech relations may be subject to both principles.

3.3. The basic principles of linguistic multimodality: Code manifestation and code integration

If we start from a code-based media concept (see Posner's definition in section 3.1.) in Fig. 46.1, then language in general and therefore every single language, German for example, is a medium. According to Merten (1999: 134), language itself is the first communicative medium, and its features mold the basic criteria of all further media (in the sense of a technological media concept). These criteria include "quantization" into the smallest syntactic entities (high resolution capacity), "non-consumability", "relationality" (reference to something other than itself), "perception of distance", "fungibility of the processed contents", "linking of psychic systems", and "multiplier functioning" if language can be adopted by various recipients at the same time (Merten 1999: 134ff.). For language to evolve as a basic code-related medium, however, it is a requirement that high-resolution capacity of a physical medium is present (Merten 1999: 141) as in the acoustic channel with spoken languages or the optical channel with sign languages. Physical media interconnect things and thereby effect our perception. They are media for our perception (Merten 1999: 144).

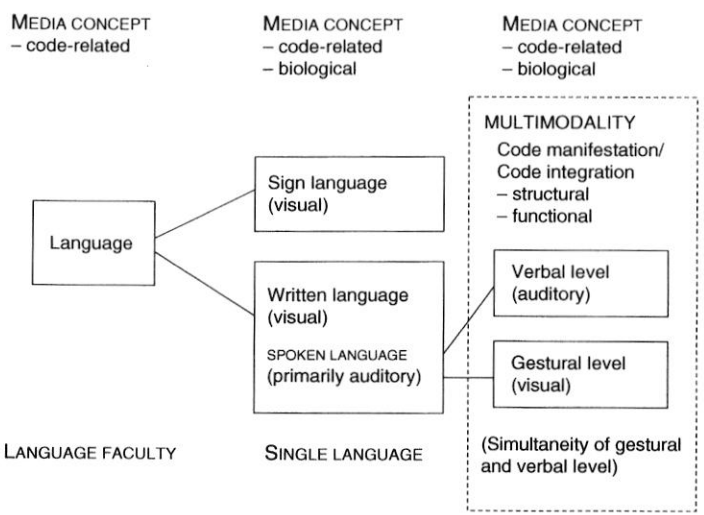

Fig. 46.1: Mediality and linguistic multimodality

When, in a further step, the code-related media concept and the biological media concept are combined, different media within the linguistic sphere can be differentiated: sign language, written language, and spoken language. A single sign language like American Sign Language (ASL) or German Sign Language (DGS) on this level is merely monomedial and monomodal because only one sensory modality is affected: the visual. Barring hypertextual applications on the internet, English or German as a written language is primarily visual and therefore primarily monomedial and monomodal as well. Visual and auditory material integrated into a text presupposes the written language as code but the reverse is not true. Written communication functions independently from other potentially integrable codes which also seems to be the case with spoken language. When on the telephone, for example, people are restricted to an auditive modality and can nevertheless communicate without the visual modality. In such communication situations, spoken language is monomedial and monomodal as well. Visible gestures accompanying speech presuppose an audible, vocal language, not the other way around. Is it therefore sufficient to likewise classify the vocal language as monomedial and monomodal? There are strong arguments for not doing so. When comparing communication by phone with face-to-face communication, it becomes clear that the latter is the ontogenetically and phylogenetically primary form of communication. The telephone is a rather young technological innovation, and its handling is learned relatively late in a child's development. With respect to a biological media concept, face-to-face communication is primarily audiovisual and, according to Lyons (1977: 637–638), the "canonical situation of utterance", serving as point of origin for all other communication situations with its specific contextual constraints.

Although the term "spoken language" suggests that only the auditory sensory modality is affected, the visual sensory modality plays a role via gestures and other body movements in face-to-face communications. Consequently two media (biological media concept) are involved, which fulfill the necessary condition both of multimodality and of multimediality. Is spoken language then to be classified as multimedial or as multimodal? What could be a reasonable differentiation between the two notions with respect to our research intentions of developing a multimodal approach to grammar? When considering characterizations of co-speech gestures (e.g., Kendon 2004; McNeill 1992, 2005), they have to be seen as body movements observed when someone is speaking. Concerning their timing, they are closely related to the uttered speech with which they also share semantic and pragmatic functions.

This points to a workable criterion for differentiating multimodality and multimediality, that is to say, the structural and functional integration into one and the same matrix code or, alternatively, the manifestation of one code in two different media. If two linguistic media are structurally and/or functionally integrated into the same code at the same time, or if, conversely, one code manifests itself simultaneously in two different media, then we can speak of multimodality. If two or more media are neither structurally nor functionally integrated into one code and if there is no manifestation of the same code in different media, then the phenomenon is defined as multimedial. This differentiation underlies our definitions of linguistic multimediality as well as linguistic multimodality in the broad and narrow sense. The following table summarizes the individual terms and their defining features:

Tab. 46.1: The differentiation of multimediality and multimodality

DEFINING FEATURE	TERM		
	Multimodality in the narrow sense	Multimodality in the broad sense	Multimediality
More than one medium	+	+	+
More than one code	+/–	+	+
More than one sense modality	+	–	+/–
Use of a technological medium	–	–	+
Code integration or code manifestation	+	+	–

What all of the terms listed in Tab. 46.1 share is that multimodality and multimediality are only present when more than one medium is involved. This is the defining characteristic as opposed to monomodality or monomediality. Multimediality can be distinguished from multimodality when there is no structural and/or functional integration of the same primary code or the same code is not manifested in different media. Within multimodality we distinguish between the broad and narrow case. Multimodality in the narrow sense occurs when the media involved in an expression belong to different sense modalities in terms of a biological media concept and are structurally and/or functionally integrated in the same code or, alternatively, manifest the same code in terms of a code-based media concept. Multimodality in the broad sense differs from the narrow sense in that the media involved belong to different codes in terms of a code-based media concept and the same sense modality in terms of a biological media concept. For the concepts of code integration and code manifestation, there are precedents in the speech theories of Pike (code integration) and Hjelmslev (code manifestation). In our current analysis, we are fundamentally open to both forms of multimodality.

4. Kinesthemes as syntactic units: Processes of typification and semantization

The widespread assumption that gestures are primarily expressive neglects the observation that even co-speech gestures, which are considered to be idiosyncratic (e.g., McNeill 1992), are based on particular gestural codes. The crucial question from a linguistic point of view is: Are co-speech gestures only part of the concrete utterance (parole) or can they also be considered to be part of the abstract language system (langue)? Scholars who argue against abstract linguistic properties of co-speech gestures emphasize that co-speech gestures lack conventionalization, which is the basis for the segmentation of linguistic units with a stable form-meaning-relationship. According to this assumption, only conventionalized types called "morphemes" and "words" can be combined into higher complex units in syntax.

Some linguists question such rigid concepts of morphemes and assume systematic processes of typification and semantic loading on a sub-morphemic level of the spoken

language. The delimitable, meaningful segments resulting from such processes are called "phonesthemes" or "sub-morphemic" units (Bolinger [1968] 1975; Firth [1935] 1957; Zelinsky-Wibbelt 1983). They are defined as intersubjective sound/meaning correlations based on diagrammatic iconicity according to Charles S. Peirce (1931–58). Bolinger (1975) characterizes them as words clustering in groups, for example, the words ending in *-ump*: *bump, chump, clump, crump, flump, glump, grump, hump*. Semantically, most of them suggest "heaviness". Bolinger's crucial observation is that of an "underlying iconic drive to make sound conform to sense" (Bolinger 1975: 218).

The integration of such concepts like that of the phonestheme into grammars of the spoken language is hindered in particular by the sharp separation between language use and language system. This separation is valid for structural linguistics in the tradition of Saussure as well as for generative linguistics in the Chomskyan tradition. In his book *System und Performanz* (Stetter 2005), Christian Stetter makes an interesting proposal to bridge the gap between language use and language system. In his approach, based on Nelson Goodman's work, linguistic types (language system) are understood as sets of tokens (language use) which – rather than being identical – are only similar to each other and do not share a common original as basis. This concept allows for intermediate stages of conventionalization like phonesthemes – and kinesthemes.

Fricke defines kinesthemes analogously to phonesthemes as gestural tokens with intersubjective semantic loading based on diagrammatic iconicity (Fricke 2008, 2010, 2012). Similarities of form correlate with similarities of meaning. Analyses of empirical examples from route descriptions at Potsdam Square in Berlin show that kinesthemes can be simple or complex. Complex kinesthemes can be compared to processes of morphological contamination or blending in word formation of spoken languages (e.g., *smog* is a blend of *smoke* and *fog*, cf. Zelinsky-Wibbelt 1983) (Fricke 2008, 2012). The following pointing gestures in figure 46.2 and 46.3 illustrate an analogous process of gesture formation:

Fig. 46.2: Two types of pointing gestures in German: G-Form and PLOH (Fricke 2012: 110)

In German, we can observe two typified forms of pointing gestures: Firstly, the so-called G-Form with an extended index finger and the palm oriented downwards, secondly, the palm-lateral-open-hand gesture (PLOH) (Fricke 2007, 2008; see Kendon and Versante 2003 for Italian gestures). The G-form is semantically loaded with a meaning which can be paraphrased as "pointing to an object", whereas the meaning of the palm-lateral-open-hand gesture is directive ("pointing in a direction"). Fig. 46.3 shows an example of a gestural contamination that blends both types. It can be paraphrased as "pointing to an object in a particular direction".

Fig. 46.3: Blending of G-Form and PLOH (Fricke 2012: 113)

This example shows that processes of formal typification and semantic loading on the verbal and gestural level are both guided by the same principles. Phonesthemes and kinesthemes manifest the same general semiotic code of sign formation and complement other types of meaning construction, for example, metonymies and metaphors (Cienki 2008; Cienki and Müller 2008; Mittelberg 2006, 2008; Müller 2008, 2010).

5. Syntactic structures as code manifestation of the language faculty: Constituency and recursion in co-speech gestures

Some linguists claim that recursion is a fundamental characteristic of human language. They argue that recursion is shared neither by animals nor by cognitive capacities other than the language faculty (Hauser, Chomsky, and Fitch 2002). In contrast to this view, Everett (2005) claims that the Amazonian Pirahã lacks evidence of recursion in its syntax. Other researchers consider recursion as the defining feature not only of human language, but of human cognition in general (Corballis 2007).

New data offer a chance to break vicious circles of theoretical argumentation. Fricke (2008, 2012) presents empirical examples from route descriptions in German that give evidence of recursion in co-speech gestures.

The fundamental questions for a multimodal approach to grammar are: Is it possible to analyze co-speech gestures independently of speech in terms of constituency? Do gestural constituent structures display recursion? What would be the implications of gestural recursion for language theory? The concept of recursion entered linguistics from mathematics and computer science. However, as Lobina and García-Albea (2009) pointed out, the adaptation of this notion in lingustics and cognitive science is not very clear. If recursion – in the context of linguistics – applies to "a constituent that contains a constituent of the same kind" in vocal language (Pinker and Jackendoff 2005: 203) as well as to the possibility of producing an infinite number of expressions with finite means, then some gestural structures can be called recursive, for example, gesture units (GU) in German.

In this section we will differentiate recursion from iteration, make a distinction between recursion as a process and structure and characterize the relation between recursion and self-embedding. If we look at recursion as separate from iteration then we see that they are different forms of sequentializing units: with the help of recursion, structures take on an increasing depth of embedding whereas iteration produces flat structures with the same level of embedding (Karlsson 2010: 45). While the units

sequentialized through iteration are completely independent of one another, the same is not true for recursive sequentialization: "Iteration involves repetition of an action or object, where each repetition is entirely independent of those that come before and after. Recursion involves the embedding of an action or object inside another of the same type, each embedding being dependent in some way on the one it is embedded inside" (Kinsella 2010: 180).

Let us look at the following examples of iteration (1) and recursion (2) (Kinsella 2010: 181):

(1) Iteration: *Jack ate* [$_{NP1}$ *the sandwiches* $_{NP1}$] *and* [$_{NP2}$ *the doughnut* $_{NP2}$] *and* [$_{NP3}$ *the apple* $_{NP3}$].
(2) Recursion: [$_{NP1}$ [$_{NP2}$ [$_{NP3}$ *John's* $_{NP3}$] *mother's* $_{NP2}$] *neighbor* $_{NP1}$] *bought the car.*

At first glance, both examples appear on the surface to be chains of noun phrases. Looking closer at the structure of each sentence, however, it becomes apparent that example (1) has a flat structure in which the noun phrases are independent of each other. In example (2), on the other hand, there exists a dependence between the noun phrases that determines the relation of modification (Kinsella 2010: 181). This is also the reason why in example (1) the order of the noun phrases could, in general, be changed whereas in example (2) this is not the case (Kinsella 2010).

Lobina (2011: 155) and Fitch (2010: 78) advise that it is necessary to maintain a strict separation between recursive structures and fundamental recursive algorithms in which iterative structures, such as in example (1) above, can be created:

> [...] many studies focus on the so-called self-embedded sentences (sentences inside other sentences, such as I know that I know etc.) as a way to demonstrate the non-finiteness of language, and given that self-embedding is sometimes used as a synonym for recursive structures (see infra), too close a connection is usually drawn between the presence of these syntactic facts and the underlying algorithm of the language faculty. (Lobina 2011: 155–154)

Lobina's distinction between structure and fundamental process also allows a different perspective on Everett's objections to the article by Hauser, Chomsky, and Fitch (2002) in that the absence of self-embedding as a structure cannot be an argument against accepting recursion as a fundamental algorithm that, as a key element of the faculty of language, might be present in all natural languages. He argues:

> However, even if there were a language that did not exhibit self-embedding but allowed for conjunction, you could run the same sort of argument and the non-finiteness conclusion would still be licensed. These two aspects must be kept separate; one focuses on the sort of expressions that languages manifest (or not), while the other is a point about the algorithm that generates all natural language structures. (Lobina 2011: 156)

With regard to the analysis of speech-accompanying gestures, Fricke (2012) shows that they alone – without reference to vocal utterances – can essentially form arbitrarily long "flat" chains: On the one hand, gestures share the structural characteristics of iteration that can be created through a recursive algorithm. On the other hand, however, gestures also share the structural properties of a "deeper" self-embedding in that gestural constituents can contain other gestural constituents of the same type. Based on

empirical analyses, Fricke proposes (2012: 176) the following phrase structure rules for co-speech gestures (for more details see volume 2):

$$
\begin{aligned}
GU \rightarrow & \left\{ \begin{array}{l} GP\ Retr \\ GU\ GU\ (GU_1 \ldots GU_n)\ Retr \\ GP\ GP\ (GP_1 \ldots GP_n)\ Retr \\ GU\ (GU_1 \ldots GU_n)\ GP\ (GP_1 \ldots GP_n)\ (GU_{n+1} \ldots GU_z)\ Retr \\ GP\ (GP_1 \ldots GP_n)\ GU\ (GU_1 \ldots GU_n)\ (GP_{n+1} \ldots GP_z)\ Retr \end{array} \right\} \\
GP \rightarrow & \ (Prep)\ SP \\
SP \rightarrow & \ S\ (S_1 \ldots S_n) \\
S \rightarrow & \ (Hold)\ s\ (Hold)
\end{aligned}
$$

The starting point for this system of rules is the gesture unit. A primary gesture unit is the highest unit of the constituent hierarchy. This fact is reflected by using the category GU as the starting symbol, comparable to the category sentence (S) in generative grammars of vocal languages. The property of self-embedding is indicated when to the left and right of the arrow the same category symbol is present. There is, to date, no empirical evidence for levels of embedding deeper than one (primary and secondary gesture units). The braces show that the vertical listing of alternative symbol chains could each serve as a "replacement" for gesture units. The individual symbols and the constituents they represent can either be obligatory or optional. If they are optional, this is shown by using parentheses.

According to Kendon (2004, 1972) gesture units (GU) are limited by positions of relaxation and – in contrast to gesture phrases (GP) – obligatorily contain a phase of retraction (Fricke 2012). A primary gesture unit is the highest constituent in the gestural constituent structure, whereas secondary gesture units are dominated by a primary gesture unit (Fricke 2012). Gesture units can be simple (GP + Retr) or complex. In principle, complex gesture units consist of an arbitrary number of gesture units and/or gesture phrases. Analyses of selected video sequences show that the embedding of secondary gesture units is indicated by the degree of relaxation and the location of the respective rest position. Primary gesture units show complete relaxation, whereas in secondary gesture units the relaxation is only partial. The hierarchy level is indicated by "gestural cohesion" (McNeill 2005): All coordinated gesture units show the same degree of relaxation in their retraction and the same location of the rest position (Fricke 2012). Stroke phrases (SP), too, can be either simple or complex. They expand to one or more strokes (S) that are ordered next to each other. Strokes (S) expand then to an obligatory stroke nucleus (Kendon 2004: 112) which can be preceded or followed by an optional hold. Whether it is a so-called pre- or post-stroke hold is not categorically determined, but rather by its position in the constituent structure. The terminal constituents are the gesture phases (e.g., Bressem and Ladewig 2011; Kendon 1980, 2004; Kita, Van Gijn, and Van der Hulst 1998) stroke nucleus (s), hold (Hold), preparation (Prep), and retraction (Retr).

What conclusions can we draw from this for language theory? If we consider the current debate about recursion and language complexity started by Hauser, Chomsky, and Fitch (2002) then the fact that co-speech gestures are recursive carries with it the following consequences:

Based on the assumption that recursion is specific to the language faculty in the narrow sense (FLN), then the recursion of co-speech gestures forces them to be considered as an integral element of language. An indication that the human language faculty of Hauser, Chomsky, and Fitch is not viewed as being fundamentally modality-specific but rather the opposite, that the possibility of a change in modality is a determining feature can be found in the following quote: "[...] only humans can lose one modality (e.g., hearing) and make up for this deficit by communicating with complete competence in a different modality (e.g., signing)" (Hauser, Chomsky, and Fitch 2002: 1575). With this, the authors find themselves not far from Hjelmslev's postulate that the substances do not in and of themselves define language and that one and the same form can be manifest in different substances (Hjelmslev 1969). From the acceptance of a compensatory function of gestures through Hauser, Chomsky, and Fitch it is just a small step to accepting a fundamentally multimodal constitution of language. Should the multimodality of language be denied, it follows that, through the recursivity of co-speech gestures, recursion cannot be unique to the faculty of language in the narrow sense.

6. Syntactic functions as a device for code integration within single languages: Multimodal attribution in German noun phrases

6.1. Code integration: Structural and functional integration in grammar and language use

The assumption that co-speech gestures are integrated on the level of language use is widely accepted. For example, co-speech gestures are coordinated with prosodic aspects (e.g., Loehr 2004; McClave 1991) and are assumed to be semantically and pragmatically co-expressive with the verbal utterance (e.g., Kendon 2004; McNeill 1992, 2005). Moreover, the idea that certain syntactic functions or structures can be instantiated by entities of different modalities, e.g., visual or auditory, is not new and can be traced back to linguists and semioticians such as Karl Bühler (2011), Louis Hjelmslev (1969), and Kenneth Pike (1967). Nevertheless, most theoretical and descriptive studies in linguistics so far are based on vocal speech and its auditory channel alone, whereas gesture scholars have tended to neglect the dimension of grammar. Therefore, we aim at the core of grammar and focus on the syntactic relationship between spoken language and accompanying gestures. Our basic questions are: Can gestures take over grammatical functions in spoken language? Are there points of structural integration into vocal syntax? And if so, of what kind? The first goal of section 6 is to give proof that co-speech gestures can be structurally integrated as constituents of noun phrases in German spoken language. The second goal is to show that these syntactically integrated gestures can function as attributes to the verbal nucleus of noun phrases. The current debate on the German deictics *so* 'like this' and *son* 'such a' shows that the adverb *so* and the article *son* constitute a point of multimodal integration (e.g., Ehlich 1987, Fricke 2007, 2012; Streeck 2002, 2009; Stukenbrock 2010). According to Hole and Klumpp (2000) analyses of German noun phrases show that *son* has to be considered as an article which is governed by the nuclear noun of the noun phrase. As a qualitative deictic, denoting a quality, *son* requires a qualitative description, which can be instantiated either verbally or gesturally. If the qualitative description takes place through an iconic

gesture, it results in a categorial selection of the gestural modes of representation according to Müller (Fricke 2012).

6.2. Attributes in noun phrases

As the term is commonly defined, attributes are a word or group of words that qualify another word. Attributes in the narrow sense are conceived as syntactically constrained to expansions of the nuclear noun in noun phrases (Eisenberg 1999). They constitute the core area of attribution and are covered by all definitions of attribute. The following example is a typical case of an adjectival attribute:

(3) *the circular table*

On the syntactic level, the adjectival attribute *circular* is an expansion of the nuclear noun and a constituent of the respective noun phrase *the circular table*. On the semantic level, attributes are modifications of the noun, which is the nucleus of the noun phrase (Eisenberg 1999). In this case modification can most easily and simply be understood as the intersection of sets between the semantic extension of the adjective *circular* (all circular entities) and the semantic extension of the noun *table* (all tables). The resulting set is a set of tables with the characteristic "being circular".

Now regard the following examples, which are designed on the basis of an empirical route description. They address the quality of shape of a given office tower, the so-called "Sony Center" in Berlin, which is built in the shape of a semicircle. The speaker localizes this tower, which resembles a bisected cylinder, on the right side of her gesture space. Her mode of modeling evokes the impression of a vertical image with a sense of depth.

In each of the following four examples in figure 46.4 we deal with noun phrases initiated by a finite article and the noun *tower* as its nucleus. In (5) and (6) the noun phrase on its verbal level is expanded with the attribute *semicircular*, which modifies the nuclear noun semantically, whereas in (4) and (7) there are no attributive expansions on the verbal level. The verbal utterances in the examples (6) and (7) are also accompanied by a speaker's gesture of modeling a semicircular shape.

(4) *the tower*
(5) *the semicircular tower*
(6) *the semicircular tower* (+ gesture)
(7) *the tower* (+ gesture)

Fig. 46.4: The gesture modeling a semicircular shape in the examples (4) to (7)

What difference occurs between the examples (5) and (7)? Both utterances inform the addressee about the shape of the object referred to. The difference merely consists of the fact that in example (5) the speaker refers to the shape of the object exclusively verbally, while in example (7) this happens solely gesturally. This shows that the

attributive function of modifying the nuclear noun in a noun phrase can also be instantiated solely by gesture. The resulting intersection of semantic extensions is the same in both cases: A set of towers with the characteristic "being semicircular". So certain occurrences of co-speech gestures fall within the scope of Eisenberg's concept of an attribute mentioned above.

6.3. Types of multimodal code integration: Substitution, temporal overlap, and cataphoric integration

With regard to noun phrases we can distinguish between two main types of multimodal code integration: First, positional integration (substitution of a syntactic gap and temporal overlap) and second, cataphoric integration by way of using the German article *son* (Fricke 2008, 2012). A further subordinated type is the categorial selection of iconic gestures with respect to specific modes of representation according to Müller (1998) (see section 6.4.). Ladewig (2012) has shown that primarily co-speech gestures in linearly-constructed multimodal utterances substitute for syntactic constituents (positional integration) and not emblematic gestures as, until recently, was widely assumed. The following example (fig. 46.5 and 46.6) focuses on temporal overlap and cataphoric integration (Fricke 2012: 251).

Fig. 46.5: Modeling a right angle (Stroke 1)

Fig. 46.6: Modeling a right angle (Stroke 2)

The German speaker on the left describes the façade of the Berlin State Library: She uses the noun phrase *sone gelb-goldenen Tafeln* 'such yellow golden tiles' accompanied by a gesture modeling a rectangular shape. On the verbal level, the adjective *gelb-golden* expands the nuclear noun, modifying it at the same time by reducing its

extension to tiles with a specific characteristic of color. On the gestural level, the rectangular shape performed by the hands of the speaker fulfills an analogous function of modifying the nuclear noun. The resulting intersection of both extensions is a set of tiles with a specific characteristic of color (yellow golden) and a specific characteristic of shape (rectangular). This division of labor is a very frequent pattern in multimodal noun phrases: due to the particular medial capacity of both modes, speakers tend to use gestures for referring to aspects of shape, whereas the use of verbal adjectives provides information with respect to color (Fricke 2008, 2012). The noun phrase of this example shows a temporal overlap between the verbal adjective and the modifying co-speech gesture.

The crucial question is whether or not co-speech gestures are capable of instantiating independent syntactic constituents detached from the nuclear noun in a noun phrase. Co-speech gestures can only be expected to adopt an attributive function within verbal noun phrases on the syntactic level if this requirement is met. Research into this question so far did not get beyond the assumption that co-speech gestures can fill syntactic gaps in linear verbal constituent structures. Considering temporal overlaps as given in this example, the following alternative explanation with respect to the relation between the rectangular gesture and the nuclear noun *tiles* could be offered: The rectangular shape metonymically stands for the respective concept TILE which is associated with the word form *tiles* (e.g., Lakoff and Johnson 1980; Mittelberg 2006; Mittelberg and Waugh 2009). This explanation would also be in line with the assumption of a so-called "lexical affiliate" according to Schegloff (1984).

It is worth observing at this point that in colloquial German the article *son* 'such a' provides a syntactic integration of modifying gestures within verbal noun phrases as required above. According to Hole and Klumpp (2000), the qualitative deictic *son* is a fully grammaticalized article inflecting for case, gender and number, which is simultaneously used for definite type reference and indefinite token reference. They give convincing evidence that in German the article *son* is not just an optional contraction of *so* 'such' and *ein* 'a'. They emphasize that "*son* does not just narrow down the meaning of the indefinite article, it introduces a whole new dimension, namely, that of a necessary two-dimensional reference classification" (Hole and Klumpp 2000: 240). Exactly these two dimensions of reference apply also to multimodal noun phrases with *son* as article. In the following examples the speaker informs the addressee about the shape of the table he wants to buy within the next few days. The underlying pattern is "I want to by such a [quality] table". In contrast to noun phrases with definite articles the speaker refers in this case to an indefinite token of a definite type (Hole and Klumpp 2000: 234). With respect to example (10) this means that the speaker wants to buy a specific kind of table (definite type) that only looks like the table he is pointing at (indefinite token).

(8) *Ich will sonen [runden] Tisch kaufen.* (verbal description)
'I want to buy such a [circular] table.'

(9) *Ich will sonen Tisch kaufen.* (+ iconic gesture of a circle)
'I want to buy such a table.' (+ iconic gesture of a circle)

(10) *Ich will sonen Tisch kaufen.* (+ pointing to a circular table)
'I want to buy such a table.' (+ pointing to a circular table)

As we have seen, the German article *son* shows a multimodal integratability of a very high degree. As a fully grammaticalized article it is governed by the nuclear noun (first step), as a qualitative deictic *son* cataphorically requires the description of a quality, which can be instantiated either verbally or gesturally (second step). Both steps will be explicated and complemented by a third step in the following section with respect to Seiler's continuum of determination within complex noun phrases.

6.4. Seiler's continuum of determination: The German article *son* as the turning point between specification and characterization

Son as an article with two dimensions of reference (indefinite token of a definite type) instantiates exactly the turning point in Seiler's continuum of determination from "determination of reference to determination of concept" (Seiler 1978: 310) by epitomizing both principles. Seiler assumes the following two main rules for the serialization of determiners in the broad sense (including adjectives, numerals, quantifiers, etc.) within noun phrases:

(i) Specification (determination of reference): "The range of head nouns for which a determiner D is potentially applicable increases with the potential distance of that determiner from the head noun N". (Seiler 1978: 308)
(ii) Characterization (determination of concept): "Determiners indicate properties implied in the concept represented by the head noun. The degree of naturalness of such an implication of Dn_i vs. Dn_j decreases proportionally to the distance of Dn_i vs. Dn_j with regard to the head noun". (Seiler 1978: 310)

In the following illustration (Fig. 46.7), Seiler's continuum with both its domains is highlighted by two grey-shaded rectangles with a bilaterally oriented arrow between them. Moving from the left side towards the right, the determination of reference declines and the determination of concept increases, while moving from right to left determination of concept declines and determination of reference increases. On the gestural level, the deictic gesture is attached to the domain of specification and the iconic gesture to the domain of characterization.

Because *son* as article is (Fig. 46.7), according to Eisenberg, governed by the nuclear noun with respect to its gender, and because in specific contexts the existence of a gesture, either deictic or iconic, is a precondition for the possibility of using *son*, *son* within a noun phrase instantiates an additional turning point, namely, between linguistic monomodality and linguistic multimodality. *Son* is the syntactic integration point on the level of the linguistic system for gestures accompanying speech in noun phrases. Gestures structurally integrated to such an extent can also be integrated functionally as attributes in verbal noun phrases. Thus, because *son* in the noun phrase requires a qualitative description, which can be gesturally instantiated as well, it is shown that iconic gestures in noun phrases constitute autonomous syntactic units detached from the nuclear noun. Furthermore, they can establish syntactic relations with the nuclear noun.

If the gestural qualitative determination takes place through an iconic gesture then there follows a categorial selection of the gestural modes of representation (Müller 1998) by the article *son* (Fricke 2012): in noun phrases with the article *son*, iconic

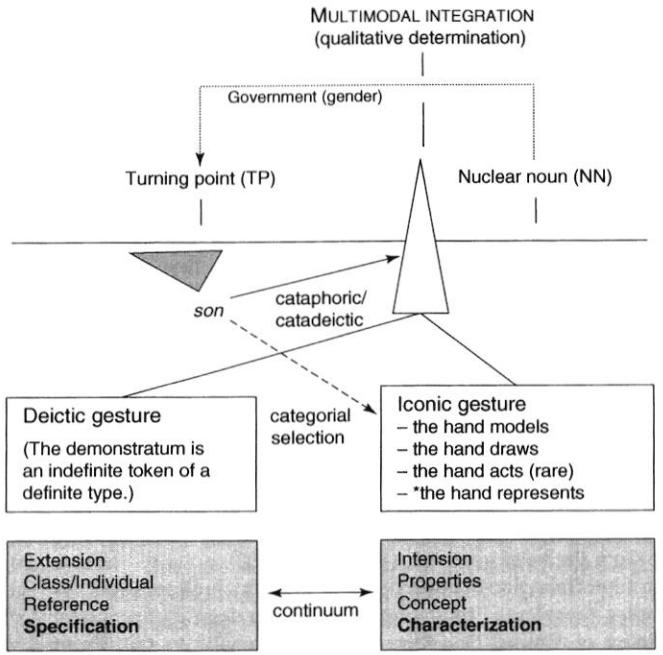

Fig. 46.7: The article *son* as turning point and syntactic integration point for co-speech gestures in noun phrases (Fricke 2012: 228)

gestures primarily occur in the modes of representation "the hand models" and "the hand draws", with rare occurrences in the mode "the hand acts". The mode "the hand represents", by contrast, does not appear once in a corpus of instructions for reaching a destination around Potsdam Square. The nuanced manual depiction of particular characteristics of objects might be hindered because the whole hand represents an object in this mode.

The fact that there seems to occur a categorial selection of iconic gestures through *son* basically distinguishes a qualitative determination through iconic gestures generated during speaking from a qualitative determination through extralinguistic objects as demonstrata of deictic gestures. These objects are given in a concrete situation and are interpreted by the speaker and the addressee according to a specific quality by means of deictic guidance of attention. Taken together, we therefore deal with a syntactical integration of gestures into noun phrases emerging in three consecutive steps (Fricke 2012: 230): The first step is constituted by the government of *son* by the nuclear noun of the noun phrase (dotted arrow), the second step consists of the cataphoric integration of a gestural qualitative determination required by *son* (solid arrow). In the case of an iconic gesture providing the qualitative determination, the third step accomplishes a categorial selection with respect to the four gestural modes of representation "the hand models", "the hand draws", "the hand acts", and "the hand represents" (dashed arrow).

7. Conclusion: Why we need a multimodal approach to grammar

"Verbal and nonverbal activity is a unified whole, and theory and methodology should be organized or created to treat it as such" (Pike 1967: 26). Therefore, we need a multimodal approach to grammar contributing to a description of language in all its structural, functional as well as medial and cognitive particularities. As we have seen, this kind of approach is not only basically feasible but also necessary to do justice to language in general and to its function as a medium of individual communication. This argument has been substantiated by the analysis of typification and semantization of gestures as potential syntactic constituents, by giving the rules of a generative context-free phrase structure grammar of co-speech gestures which displays recursion and self-embedding, and by the grammatical analysis of multimodal attribution in German noun phrases. The notion of multimodality so far has been constrained to specific types of utterances such as gesture-speech relations or language-image relations. If we conceive of multimodality as a global dimension of linguistic and semiotic analysis which is generally applicable to language and other systems of signs then we have to broaden our perspective by also including grammars of single languages and the human faculty of language. With respect to linguistics, this extension of perspective on multimodality reveals two basic principles: multimodal code manifestation of the language faculty and multimodal processes of code integration within grammars of single languages on the level of the language system. With regard to the objectives of a multimodal approach to grammar, we have distinguished between linguistic multimodality in the narrow and the broad sense. Multimodality in the narrow sense occurs when the media involved in an expression belong to different sense modalities and are structurally or functionally integrated in the same code or, alternatively, manifest the same code, e.g., "gesture-speech ensembles" (Kendon 2004). In the broad sense of multimediality, the media involved belong to the same sense modality, e.g., "language-image ensembles". It is worth pointing out that both kinds of multimodal ensembles differ with respect to their specific potential for establishing and instantiating grammatical structures and functions. According to Goodman's *Languages of Art* (1976), non-linear images are essentially syntactically and semantically dense whereas we must give due recognition to the fact that linear co-speech gestures are not. These findings offer new perspectives for comparative studies on multimodality and grammaticalization combining both areas of research.

8. References

Bolinger, Dwight L. 1975. *Aspects of Language.* New York: Harcourt Brace Jovanovich. First published [1968].

Bressem, Jana 2012. *Repetitions in gesture: Structures, functions, and cognitive aspects.* Ph.D. dissertation, European University Viadrina, Frankfurt (Oder).

Bressem, Jana and Silva H. Ladewig 2011. Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica* 184: 53–91.

Bühler, Karl 2011. *Theory of Language. The Representational Function of Language.* Amsterdam: Benjamins. First published [1934].

Cienki, Alan 2008. Why study metaphor and gesture? In: Alan Cienki and Cornelia Müller (eds.), *Metaphor and Gesture*, 5–24. Amsterdam: Benjamins.

Cienki, Alan and Cornelia Müller (eds.) (2008). *Metaphor and Gesture.* Amsterdam: Benjamins.

Corballis, Michael C. 2007. The uniqueness of human recursive thinking. *American Scientist* 95: 240–248.
Efron, David 1972. *Gesture, Race and Culture*. The Hague: Mouton. First published [1941].
Ehlich, Konrad 1987. *so* – Überlegungen zum Verhältnis sprachlicher Formen und sprachlichen Handelns, allgemein und an einem widerspenstigen Beispiel. In: Inger Rosengren (ed.), *Sprache und Pragmatik. Lunder Symposium 1986*, 279–313. Stockholm: Almqvist and Wiksell.
Eisenberg, Peter 1999. *Grundriß der deutschen Grammatik. Volume 2: Der Satz*. Stuttgart: Metzler.
Ekman, Paul and Wallace V. Friesen 1969. The repertoire of nonverbal behavior. Categories, origins, usage, and coding. *Semiotica* 1: 49–98.
Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã. Another look at the design features of human language. *Current Anthropology* 46: 621–646.
Firth, John Rupert 1957. The use and distribution of certain English sounds. In: John Rupert Firth, *Papers in Linguistics 1934–1951*, 34–46. London: Oxford University Press. First published [1935].
Fitch, W. Tecumseh 2010. Three meanings of "recursion": Key distinctions for biolinguistics. In: Richard K. Larson, Viviane Déprez and Hiroko Yamakido (eds.), *The Evolution of Human Language*, 73–90. Cambridge: Cambridge University Press.
Fricke, Ellen 2007. *Origo, Geste und Raum: Lokaldeixis im Deutschen*. Berlin: De Gruyter.
Fricke, Ellen 2008. Grundlagen einer multimodalen Grammatik: syntaktische Strukturen und Funktionen. Habilitation thesis, European University Viadrina, Frankfurt (Oder).
Fricke, Ellen 2010. Phonaestheme, Kinaestheme und multimodale Grammatik. *Sprache und Literatur* 41: 69–88.
Fricke, Ellen 2012. *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin: De Gruyter.
Fricke, Ellen, Jana Bressem and Cornelia Müller in preparation. Gesture families and gestural fields. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeill and Sedinha Teßendorf (eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction* (HSK 38.2). Berlin and Boston: De Gruyter Mouton.
Goodman, Nelson 1976. *Languages of Art. An Approach to a Theory of Symbols*. London: Oxford University Press. First published [1968].
Harrison, Simon 2008. The expression of negation through grammar and gesture. In: Jordan Zlatev, Mats Andrén, Marlene Johansson Falck and Carita Lundmark (eds.), *Studies in Language and Cognition*, 405–409. Newcastle upon Tyne: Cambridge Scholars Publishing.
Harrison, Simon 2009. Grammar, gesture, and cognition: The case of negation in English. Ph.D. dissertation, Université Bordeaux 3.
Hauser, Marc D., Noam Chomsky and W. Tecumseh Fitch 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(4): 1569–1579.
Hjelmslev, Louis 1969. *Prolegomena to a Theory of Language*. Madison: University of Wisconsin Press. First published [1943].
Hole, Daniel and Gerson Klumpp 2000. Definite type and indefinite token: The article *son* in colloquial German. *Linguistische Berichte* 182: 231–244.
Karlsson, Fred 2010. Syntactic recursion and iteration. In: Harry van der Hulst (ed.), *Recursion and Human Language*, 43–67. Berlin: De Gruyter Mouton.
Kendon, Adam 1972. Some relationships between body motion and speech. An analysis of an example. In: Aron W. Siegman and Benjamin Pope (eds.), *Studies in Dyadic Communication*, 177–210. New York: Pergamon Press.
Kendon, Adam 1980. Gesticulation and speech: Two aspects of the process of utterance. In: Mary R. Key (ed.), *The Relationship of Verbal and Nonverbal Communication*, 207–227. The Hague: Mouton.
Kendon, Adam 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
Kendon, Adam and Laura Versante 2003. Pointing by hand in "Neapolitan". In: Sotaro Kita (ed.), *Pointing: Where Language, Culture, and Cognition Meet*, 109–137. Mahwah, NJ: Erlbaum.

Kinsella, Anna 2010. Was recursion the key step in the evolution of the human language faculty? In: Harry van der Hulst (ed.), *Recursion and Human Language*, 179–191. Berlin: De Gruyter Mouton.
Kita, Sotaro, Ingeborg van Gijn and Harry van der Hulst 1998. Movement phases in signs and co-speech gestures, and their transcription by human encoders. In: Ipke Wachsmuth and Martin Fröhlich (eds.), *Gesture and Sign Language in Human-Computer Interaction*, 23–35. Berlin: Springer.
Krämer, Sybille 1998. Das Medium als Spur und Apparat. In: Sybille Krämer (ed.), *Medien, Computer, Realität: Wirklichkeitsvorstellungen und Neue Medien*, 73–94. Frankfurt am Main: Suhrkamp.
Kress, Gunter 2011. What is mode? In: Carey Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis*, 54–67. London: Routledge. First published [2009].
Kress, Gunter and Theo van Leeuven 2006. *Reading Images. The Grammar of Visual Design*. London: Routledge. First published [1996].
Ladewig, Silva H. 2010. Beschreiben, suchen und auffordern. Varianten einer rekurrenten Geste. *Sprache und Literatur* 41(1): 89–111.
Ladewig, Silva H. 2011. Putting the cyclic gesture on a cognitive basis. *CogniTextes* 6. http://cognitextes.revues.org/406.
Ladewig, Silva H. 2012. *Syntactic and semantic integration of gestures into speech: structural, cognitive, and conceptual aspects*. Ph.D. dissertation, European University Viadrina, Frankfurt (Oder).
Lakoff, George and Mark Johnson 1980. *Metaphors We Live By*. Chicago: Chicago University Press.
Lobina, David J. 2011 "A running back" and forth: A review of recursion and human language. *Biolinguistics* 5(1–2): 151–169. http://www.biolinguistics.eu/index.php/biolinguistics/article/view/198.
Lobina, David J. and José Eugenio García-Albea 2009. Recursion and cognitive science: Data structures and mechanisms. In: Niels Taatgen and Hedderik van Rijn (eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 1347–1352. Austin, Texas: Cognitive Science Society
Loehr, Daniel P. 2004. Gesture and intonation. Ph.D. thesis, Georgetown University, Washington, DC.
Lyons, John 1977. *Semantics*. Volume 2. Cambridge: Cambridge University Press.
McClave, Evelyn Z. 1991. Intonation and gestures. Ph.D. thesis, Georgetown University, Washington, DC.
McNeill, David 1985. So you think gestures are nonverbal? *Psychological Review* 92: 350–371.
McNeill, David 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: Chicago University Press.
McNeill, David 2005. *Gesture and Thought*. Chicago: University of Chicago Press.
Merten, Klaus 1999. *Einführung in die Kommunikationswissenschaft*. Volume 1/1: Grundlagen der Kommunikationswissenschaft. Münster, Germany: LIT.
Mittelberg, Irene 2006. *Metaphor and metonymy in language and gesture: Discourse evidence for multimodal models of grammar*. Ph.D. dissertation, Cornell University, Ithaca, NY.
Mittelberg, Irene 2008. Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In: Alan Cienki and Cornelia Müller (eds.), *Metaphor and Gesture*, 115–154. Amsterdam: Benjamins.
Mittelberg, Irene and Linda R. Waugh 2009. Metonymy first, metaphor second: A cognitive-semiotic approach to multimodal figures of speech in co-speech gesture. In: Charles Forceville and Eduardo Urios-Aparisi (eds.), *Multimodal Metaphor*, 329–356. Berlin: De Gruyter Mouton.
Müller, Cornelia 1998. *Redebegleitende Gesten: Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Arno Spitz.
Müller, Cornelia 2004. Forms and uses of the palm up open hand: A case of a gesture family? In: Cornelia Müller and Roland Posner (eds.), *The Semantics and Pragmatics of Everyday Gestures. Proceedings of the Berlin Conference April 1998*, 233–256. Berlin: Weidler.

Müller, Cornelia 2008. *Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View.* Chicago: University of Chicago Press.
Müller, Cornelia 2010. Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. *Sprache und Literatur* 41/1: 37–68.
Müller, Cornelia submitted. How gestures mean. The construal of meaning in gestures with speech.
Peirce, Charles Sanders 1931–58. *Collected Papers.* Charles Hawthorne and Paul Weiss (eds.) Volumes 1–6; Arthur W. Burks (ed.) Volumes 7–8. Cambridge, MA: Harvard University Press.
Pike, Kenneth L. 1967. *Language in Relation to a Unified Theory of the Structure of Human Behavior.* The Hague: Mouton.
Pinker, Steven and Ray Jackendoff 2005. The faculty of language: What's special about it? *Cognition* 95(2): 201–236.
Posner, Roland 1986. Zur Systematik der Beschreibung verbaler und non-verbaler Kommunikation. In: Hans-Georg Bosshardt (ed.), *Perspektiven auf Sprache: Interdisziplinäre Beiträge zum Gedenken an Hans Hörmann*, 267–313. Berlin: De Gruyter.
Posner, Roland 2004. Basic tasks of cultural semiotics. In: Gloria Withalm and Josef Wallmannsberger (eds.), *Signs of Power – Power of Signs. Essays in Honor of Jeff Bernard*, 56–89. Vienna: INST.
Ruesch, Jurgen and Weldon Kees 1972. *Nonverbal Communication: Notes on the Visual Perception of Human Relations.* Berkeley: University of California. First published [1956].
Sauerland, Uli and Andreas Trotzke 2011. Biolinguistic perspectives on recursion: Introduction to the special issue. *Biolinguistics* 5(1–2): 1–9. http://www.biolinguistics.eu/index.php/biolinguistics/article/view/201/210.
Schegloff, Emanuel A. 1984. On some gestures' relation to talk. In: J. Maxwell Atkinson and John Heritage (eds.), *Structures of Social Action: Studies in Conversational Analysis*, 266–296. Cambridge: Cambridge Universtity Press.
Scherer, Klaus R. and Harald G. Walbott 1984. *Nonverbale Kommunikation. Forschungsberichte zum Interaktionsverhalten.* Weinheim: Belz.
Seiler, Hansjakob 1978. Determination: A functional dimension for interlanguage comparison. In: Hansjakob Seiler (ed.), *Language Universals*, 301–328. Tübingen: Narr.
Stetter, Christian 2005. *System und Performanz. Symboltheoretische Grundlagen von Medientheorie und Sprachwissenschaft.* Weilerswist: Velbrück Wissenschaft.
Stöckl, Hartmut 2004. In between modes: Language and image in printed media. In: Eija Ventola, Cassily Charles and Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 9–30. Amsterdam: Benjamins.
Streeck, Jürgen 2002. Grammars, words, and embodied meanings: On the uses and evolution of *so* and *like. Journal of Communication* 52(3): 581–596.
Streeck, Jürgen 2009. *Gesturecraft. The Manu-facture of Meaning.* Amsterdam: Benjamins.
Stukenbrock, Anja 2010. Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks *so. InLiSt – Interaction and Linguistic Structures* 47. http:// www.inlist.uni-bayreuth.de/issues/47/InLiSt47.pdf.
Watzlawick, Paul, Janet Beavin Bavelas and Don D. Jackson 1967. *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes.* New York: Norton.
Wundt, Wilhelm 1904. *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte.* Volume 1: *Die Sprache.* Leipzig: Engelmann. First published [1900].
Wundt, Wilhelm 1973. *The Language of Gestures.* The Hague: Mouton. First published [1900].
Zelinsky-Wibbelt, Cornelia 1983. *Die semantische Belastung von submorphematischen Einheiten im Englischen: Eine empirisch-strukturelle Untersuchung.* Frankfurt am Main: Peter Lang.

Ellen Fricke, Freiburg i. Br. and Chemnitz (Germany)