Chapter 31

Gesture

Andy Lücking
Université de Paris, Goethe-Universität Frankfurt

The received view in (psycho)linguistics, dialogue theory and gesture studies is that co-verbal gestures, i.e. hand and arm movement, are part of the utterance and contribute to its content (Kendon 1980; McNeill 1992). The relationships between gesture and speech obey regularities that need to be defined in terms of not just the relative timing of gesture to speech, but also the linguistic form of that speech: for instance, prosody and syntactic constituency and headedness (Loehr 2007; Ebert et al. 2011; Alahverdzhieva et al. 2017). Consequently, speech-gesture integration is captured in grammar by means of a gesture-grammar interface. This chapter provides basic snapshots from gesture research, reviews constraints on speech-gesture integration and summarises their implementations into HPSG frameworks. Pointers to future developments conclude the exposition. Since there are already a couple of overviews on gesture such as Özyürek (2012), Wagner et al. (2014) and Abner et al. (2015), this chapter aims at distinguishing itself by providing a guided tour of research that focuses on using (mostly) standard methods for semantic composition in constraint-based grammars like HPSG to model gesture meanings.

1 Why gestures?

People talk with their whole body. A verbal utterance is couched in an intonation pattern that, via prosody, articulation speed or stress, function as *paralinguistic* signals (e.g. Birdwhistell 1970). The temporal dimension of paralinguistics

Andy Lücking. 2020. Gesture. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 1011–1059. Berlin: Language Science Press. DOI:??

gives rise to *chronemic* codes (Poyatos 1975; Bruneau 1980). *Facial expressions* are commonly used to signal emotional states (Ekman & Friesen 1978), even without speech (Argyle 1975), and are correlated to different illocutions of the speech acts performed by a speaker (Domaneschi et al. 2017). Interlocutors use *gaze* as a mechanism to achieve joint attention (Argyle & Cook 1976) or provide social signals (Kendon 1967). Distance and relative direction of speakers and addressees are organised according to culture-specific radii into social spaces (*proxemics*, Hall 1968). Within the inner radius of private space, tactile codes of *tacesics* (Kauffman 1971) are at work. Since the verbal and nonverbal communication means of face to face interaction may occur simultaneously, *synchrony* (i.e. the mutual overlap or relative timing of verbal vs. non-verbal communicative actions) is a feature of the multimodal utterance itself; it contributes, for instance, to identifying the word(s) that are affiliated to a gesture (Wiltshire 2007). A special chronemic case is signalling at the right moment – or, for that matter, missing the right moment (an aspect of communication dubbed *kairemics* by Lücking & Pfeiffer 2012: 600). Besides the manifold areas of language use, the conventionalised, symbolic nature of language secures language's primacy in communication, however (de Ruiter 2004). For thorough introductions into semiotics and multimodal communication see Nöth (1990), Posner et al. (1997–2004) or Müller et al. (2013–2014).

The most conspicuous non-verbal communication means of everyday interaction are hand and arm movements, known as *gestures* (in a more narrow sense which is also pursued from here on). In seminal works, McNeill (1985; 1992) and Kendon (1980; 2004) argue that co-verbal gestures, i.e. hand and arm movements, can be likened to words in the sense that they are part of a speaker's utterance and contribute to discourse. Accordingly, integrated speech-gesture production models have been devised (Kita & Özyürek 2003; de Ruiter 2000; Krauss et al. 2000) that treat utterance production as a multimodal process (see Section 4.4 for a brief discussion). Given gestures' imagistic and often spontaneous character, it is appealing to think of them as "postcards from the mind" (de Ruiter 2007: 21). Clearly, given this entrenchment in speaking, the fact that one can communicate meaning with non-verbal signals has repercussions to areas hitherto taken to be purely linguistic (in the sense of being related to the verbal domain). This section highlights some phenomena particularly important for grammar, including, for instance, *mixed syntax* (Slama-Cazacu 1976), or *pro-speech gesture*:

(1) He is a bit [*circular movement of index finger in front of temple*].

In (1), a gesture replaces a position that is usually filled by a syntactic constituent. The gesture is emblematically related to the property of *being mad* so that the mixed utterance from (1) is equivalent to the proposition that the referent of *he* is a bit mad.

Figure 1: *Die Skulptur die hat 'n* [*BETONsockel*] ('The sculpture has a [CONCRETE base]') [V5, 0:39]

The gesture shown in Figure 1 depicts the shape of a concrete base, which the speaker introduces into discourse as an attribute of a sculpture:[1]

(2) Die Skulptur die hat 'n [BETONsockel].
The sculpture it has a [CONCRETE base].
'The sculpture has a concrete base.'

The following representational conventions obtain: square brackets roughly indicate the portion of speech which overlaps temporally with the gesture (or more precisely, with the gesture stroke; see Figure 5 below) and upper case is used to mark main stress or accent. So both timing and intonation give clues that the gesture is related to the noun *Betonsockel* 'concrete base'. From the gesture, but not from speech, we get that the concrete base of the sculpture has the shape of a flat cylinder – thus, the gesture acts as a nominal modifier. There is a further complication, however: the gesture is incomplete with regard to its interpretation – it just depicts about half of cylinder. Thus, gesture interpretation may involve processes known from gestalt theory (see Lücking 2016 on a *good continuation* constraint relevant to (2)/Figure 1).

The speaker of the datum in Figure 2 uses just a demonstrative adverb in order to describe the shape of a building he is talking about:

[1]The examples in Figures 1, 2, 3, 4, 9 and 11 are drawn from the (German) *Speech and Gesture Alignment* corpus (SaGA, Lücking et al. 2010) and are quoted according to the number of the dialogue they appear in and their starting time in the respective video file (e.g. "V9, 5:16" means that the datum can be found in the video file of dialogue V9 at minute 5:16). Examples/Figures 4 and 9 have been produced especially for this volume; all others have also been used in Lücking (2013) and/or Lücking (2016).

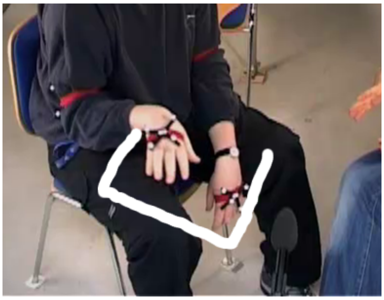


Figure 2: *Dann ist das Haus halt SO* [] (‘The house is like THIS []’) [V11, 2:32]

(3) Dann ist das Haus halt SO [].
Then is the house just like THIS [].
‘The house is like this [].’

The demonstrative shifts the addressee’s attention to the gesture, which accomplishes the full shape description, namely a cornered U-shape. In contrast to the example in Figure 1, the utterance associated with Figure 2 is not even interpretable without the gesture.

A lack of interpretability is shared by exophorically used demonstratives, which are *incomplete* without a demonstration act like a pointing gesture (Kaplan 1989: 490). For instance, Claudius would experience difficulties in understanding how serious Polonius is about his (Polonius’) conjecture about the reason of Hamlet’s (alleged) madness, if Polonius had not produced pointing gestures (Shakespeare, *Hamlet, Prince of Denmark* Act II, Scene 2; the third occurrence of *this* is anaphoric and refers back to Polonius’ conjecture):

(4) POLONIUS (*points to his head and shoulders*): Take this from this if this be otherwise.

In order for Claudius to interpret Polonius’ multimodal utterance properly, he has to associate correctly the two pointing gestures with the first two occurrences of *this* (cf. the problems discussed by Kupffer 2014). Polonius facilitates such an interpretation by means of a temporal coupling of pointing gestures and their associated demonstratives – a relationship that is called *affiliation*. The role of synchrony in multimodal utterances is further illustrated by the following example, (5), and Figure 3 (taken from Lücking 2013: 189):

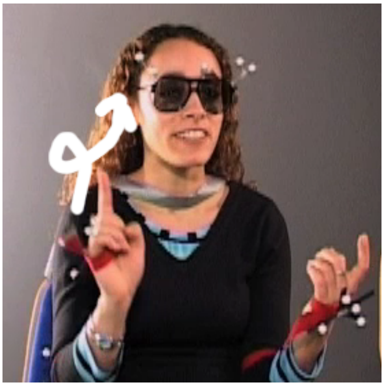


Figure 3: *Ich g[laube das sollen TREP]pen sein* (‘I think those should be STAIRcases’) [V10, 3:19]

(5) Ich g[laube das sollen TREP]pen sein.
I think those should STAIRcases be
‘I think those should be staircases.’

The first syllable of the German noun *Treppen* (*staircases*) carries main stress, indicated by capitalization. The square brackets indicate the temporal overlap between speech and gesture stroke, which is shown in Figure 3. The gesture attributes a property to the noun it attaches to: from the multimodal utterance, the observer retrieves the information that the speaker talks about spiral staircases. This interpretation assumes that the common noun is the affiliate of the gesture. Obviously, mere temporal synchrony is too weak to be an indicator of affiliation. In fact, there are speech-gesture affiliations without temporal overlap between gesture and verbal affiliate at all (e.g. Lücking et al. 2004). Therefore, temporal overlap or vicinity is just one indicator of affiliation. A second one is intonation: a gesture is usually related to a stressed element in speech (McClave 1994; Nobe 2000; Loehr 2004; 2007). As a result, multimodal communication gives rise to a complex “peak pattern” (Tuite 1993; Loehr 2004; Jannedy & Mendoza-Denton 2005).

The interpretation of a gesture changes with different affiliations. Suppose the gesture from Figure 3 is produced in company to stressed *glaube* (*think*) instead of *staircases*:

(6) Ich G[LAUbe das sollen Trep]pen sein.
I THINK those should staircases be
‘I think those should be staircases.’

Now the spiral movement is interpreted as a metaphorical depiction of a psychological process. Thus, the interpretation of a gesture depends on the integration point (affiliation), which in turn is marked by temporal vicinity, prosody and syntactic constituency of the candidate affiliate (Alahverdzhieva et al. 2017).

The crucial observations in any case are that gestures contribute to propositional content and take part in pragmatic processes. Interestingly, gestures share the latter aspect with laughter, which also has propositional content (Ginzburg et al. 2015), for instance, when referring to real world events. Thus, a multimodal utterance may express a richer content than speech alone, as in (5), or a content equivalent to speech, as in (6); it can even express less than speech or contradict speech:[2]

> The nonverbal act can repeat, augment, illustrate, accent, or contradict the words; it can anticipate, coincide with, substitute for or follow the verbal behavior; and it can be unrelated to the verbal behavior. (Ekman & Friesen 1969: 53)

Contradictions or speech-gesture mismatches can occur when saying "right" but pointing left (as can be observed in everyday life but also been found in SaGA, e.g. in dialogue V24, at 4:50). A more complex case is given in (7) and Figure 4, where the speaker talks about a "rectangular arch" (which is of course a *contradictio in adiecto* in itself), but produces a roundish movement with the extended index finger of her right hand (the object she talks about is an archway). Note that the gesture just overlaps with "rectangular": its temporal extension in (7) is again indicated by means of square brackets within the original German utterances. The main stress is on the first syllable of the adjective and the noun receives secondary stress. The dots ("..") mark a short pause, so the gesture starts before "rechteckiger".

(7)	so'n	so'ne Art	[.. RECHTecki]ger	BOgen
	such an	such kind of	.. RECTangular	ARrch
	'kind of rectangular arch'			

An obvious interpretation of this mismatch is that "rectangular" is a slip of the tongue; interestingly, we found no "slip of the hand" in our data so far (which may be a hint to a possibly imagistic origin of gestures, as assumed in some production models; cf. Section 4.4).

[2] In case of contradiction or speech-gesture mismatch, the resulting multimodal utterance is perceived as ill-formed and induces N400 effects (Wu & Coulson 2005; Kelly et al. 2004).

Figure 4: "so'n so'ne Art [.. RECHTecki]ger BOgen" (*kind of RECTangular ARch*) [V4, 1:47].

Moving from sentence to dialogue, *interactive gestures* are bound up with turn management, among other things (Bavelas et al. 1992; 1995). For instance, pointing gestures can be used to indicate the next speaker (Rieser & Poesio 2009). Interestingly, speaker-indicating pointings are typically not produced with an outstretched index finger, but with an open hand (an example is given in Figure 14 in Section 3.6). Thus, irrespective of the question whether grammar is inherently multimodal, dialogue theory has to deal at least with certain non-verbal interaction means in any case (see also Lücking, Ginzburg & Cooper (2020), Chapter 29 of this volume).

While there is ample evidence that at least some gestures contribute to the content of the utterance they co-occur with, does this also mean that they are part of the content *intended to be communicated*? A prominent counter-example is gesturing on the telephone (see Bavelas et al. 2008 for an overview of a number of respective studies). Since such gestures are not observable for the addressee, they cannot reasonably be taken to be a constituent of the content intended for communication. Rather, "telephone gestures" seem to be speaker-oriented, presumably facilitating word retrieval. The fact that it is difficult to suppress gesturing even in absence of an addressee speaks in favour of a multimodal nature if not of language, then at least of speaking and surely interacting. Furthermore, the lion's share of everyday gestures seems to consist of rather sloppy movements that do not contribute to the content of the utterance in any interesting sense, though they might signal other information like speaker states. In this sense they are contingent, as opposed to being an obligatory semantic component (Lücking 2013). Gestures (or some other demonstration act) can become obligatory when they are produced within the scope of a demonstrative expression (recall (3)/Figure 2). A concurrent use with demonstratives is also one of the hallmarks collected by Cooperrider (2017) in order to distinguish *foreground* from *background* gestures (the other hallmarks are absence of speech, co-organization with speaker gaze and speaker effort). This distinction reflects two traditions within gesture studies: according to one tradition most prominently bound up with the work of McNeill (1992), gesture is a *byproduct* of speaking and therefore opens a "window into the speaker's mind". The other tradition, represented early on by Goodwin (2003) and Clark (1996), conceives gestures as a *product* of speaking, that is, as interaction means designed with a communicative intention. Since a gesture cannot be both a byproduct and a product at the same time, as noted by Cooperrider (2017), a bifurcation that is rooted in the cause and the production process of the gesture has to be acknowledged (e.g. gesturing on the phone is only puzzling from the product view, but not from the byproduct one). We will encounter this distinction again when briefly reviewing speech-gesture production models in Section 4.4. Gestures of both species are covered in the following.

2 Kinds of gestures

Pointing at an object seems to be a different kind of gesture than mimicking drinking by moving a bent hand (i.e. virtually holding something) towards the mouth while slightly rotating the back of hand upwards. And both seem to be different from actions like scratching or nose-picking. On such grounds, gestures are usually assigned to one or more classes of a taxonomy of gesture classes. Gestures that fulfil a physiological need (such as scratching, nose-picking, foot-shaking or pen-fiddling) have been called *adaptors* (Ekman & Friesen 1969) and are not dealt with further here (but see Żywiczyński et al. 2017 for evidence that adaptors may be associated with turn transition points in dialogue). Gestures that have an intrinsic relation to speech and what is communicated have been called *regulators* and *illustrators* (Ekman & Friesen 1969) and cover a variety of gesture classes. These gesture classes are characterized by the function performed by a gesture and the meaning relation the gesture bears to its content. A classic taxonomy consists of the following inventory (McNeill 1992):

- iconic (or representational) gestures. Spontaneous hand and arm movements that are commonly said to be based on some kind of resemblance relation.[3] Iconic gestures employ a mode of representation such as *draw-*

[3]But see footnote 8 in Section 3.5 for pointers to critical discussions of resemblance as a sign-bearing relation.

ing, *modelling*, *shaping* or *placing* (Streeck 2008; Müller 1998).

- deictic gestures (pointing). Typically hand and arm movements that perform a demonstration act. In which way pointing is standardly accomplished is subject to culture-specific conventions (Wilkins 2003). In principle, any extended body part, artefact or locomotor momentum will serve the demonstrative purpose. Accordingly, there are deictic systems that involve lip-pointing (Enfield 2001) and nose-pointing (Cooperrider & Núñez 2012). Furthermore, under certain circumstances, pointing with the eyes (gaze-pointing) is also possible (Hadjikhani et al. 2008). Note further that the various deictic means can be interrelated. For instance, manual pointing can be differentiated by cues of head and gaze (Butterworth & Itakura 2000). Furthermore, pointing with the hand can be accomplished by various hand shapes: Kendon & Versante (2003) distinguish *index finger pointing*, (with a *palm down* and a *palm vertical* variant) *thumb pointing*, and *open hand pointing* (again with various palm orientations). Kendon & Versante (2003: 109) claim that "the form of pointing adopted provides information about how the speaker wishes the object being indicated to be regarded". For instance, pointing with the thumb is usually used when the precise location of the intended referent is not important (Kendon & Versante 2003: 121–125), while the typical use of index finger palm down pointing is to single out an object (Kendon & Versante 2003: 115). Open hand pointing has a built-in metonymic function since the object pointed at is introduced as an example for issues related to the current discourse topic (what in semantic parlance can be conceived as the *question under discussion*; see, e.g. Ginzburg 2012). For instance, with 'open hand palm vertical', one indicates the *type* of the object pointed at instead of the object itself (Kendon & Versante 2003: 126).

- beats (rhythmic gestures, baton). Hand and arm movements that are coupled to the intonational or rhythmic contour of the accompanying speech. Beats lack representational content but are usually used for an emphasising effect. "The typical beat is a simple flick of the hand or fingers up and down, or back and forth" (McNeill 1992: 15). Hence, a beat is a gestural means to accomplish what is usually expressed by vocal stress, rhythm or speed in speech.

- emblem (lexicalized gestures). In contrast to the other classes, emblems are special in that they follow a fully conventionalized form-meaning relation. A common example in Western countries is the thumbs-up gesture, signalling "approval or encouragement" (Merriam Webster online dictionary[4]). Emblems may also be more local and collected within a dictionary like the dictionary of everyday gestures in Bulgaria (Kolarova 2011).

Reconsidering gestures that have been classified as beats, among other gestures, Bavelas et al. (1992) observed that many of the stroke movements accomplish functions beyond rhythmic structuring or emphasis. Rather, they appear to contribute to dialogue management and have been called *interactive gestures*. Therefore, these gestures should be added to the taxonomy:

- interactive gestures. Hand and arm movements that accomplish the function "of helping the interlocutors coordinate their dialogue" (Bavelas et al. 1995: 394). Interactive gestures include pointing gestures that serve turn allocation ("go ahead, it's your turn") and gestures that are bound up with speaker attitudes or the relationship between speaker and addressee. Examples can be found in 'open palm/palm upwards' gestures used to indicate the information status of a proposition ("as you know") or the mimicking of quotation marks in order to signal a report of direct speech (although this also has a clear iconic aspect).

The gesture classes should not be considered as mutually exclusive categories, but rather as dimensions according to which gestures can be defined, allowing for multi-dimensional cross-classifications (McNeill 2005; Gerwing & Bavelas 2004). For instance, it is possible to superimpose pointing gestures with iconic traits. This has been found in the study on pointing gestures described in Kranstedt et al. (2006a), where two participants at a time were involved in an identification game: one participant pointed at one of several parts of a toy airplane scattered over a table, the other participant had to identify the pointed object. When pointing at a disk (a wheel of the toy airplane), some participants used index palm down pointing, but additionally turned around their index finger in a circle – that is, the pointing gesture not only locates the disk (deictic dimension) but also depicted its shape (iconic dimension). See Özyürek (2012) for an overview of various gesture classification schemes.

In addition to classifying gestures according to the above-given functional groups, a further distinction is usually made with regard to the ontological place of their referent: representational and deictic gestures can relate to concrete or to

[4] https://www.merriam-webster.com/dictionary/thumbs-up, accessed 20th August 2018. The fact that emblems can be lexicalized in dictionaries emphasizes their special, conventional status among gestures.

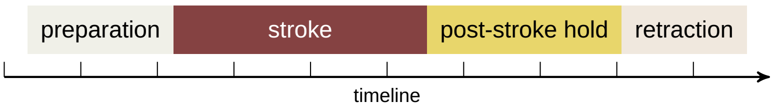


Figure 5: Gesture phases

abstract objects or scenes. For instance, an iconic drawing gesture can metaphorically display the notion “genre” via a conduit metaphor (McNeill 1992: 14):

(8) It [was a Sylves]ter and Tweety cartoon.
both hands rise up with open palm handshape, palms facing; brackets indicate segments concurrent with the gesture stroke (see Figure 5).

The gesture in (8) virtually holds an object, thus depicting the abstract concept of the genre of being a Sylvester and Tweety cartoon as a bounded container. Accordingly, gestures can be cross-classified into *concrete* and *abstract* or *metaphorical* ones (see the volume of Cienki & Müller 2008 on gesture and metaphor).

On the most basic, kinematic level, the movement of a prototypical gesture follows an “anatomic triple”: gestures have to be partitioned into at least a preparation, a stroke, and a retraction phase (Kendon 1972). The gesture phases are shown in the diagram in Figure 5. The stroke is the movement part that carries the gesture’s meaning. It can be “frozen”, leading to a post-stroke hold. If a stroke has to wait for its affiliated expression(s), a pre-stroke hold can also arise (Kita et al. 1998). The preparation and retraction phases bring hand and arms into and out of the stroke, respectively. Unless stated otherwise, when talking about gestures in what follows (and in hindsight concerning the examples given in Section 1), the stroke phase, which is the “gesture proper” or the “semantically interpretable” phase, is referred to.

Perhaps it should be noted that the spontaneous, usually co-verbal hand and arm movements considered in this chapter are different from the signed signs of sign languages (see Steinbach & Holler 2020, Chapter 30 of this volume) and pantomime (neither spontaneous nor co-verbal).[5]

3 Gestures in HPSG

Integrating a gesture’s contribution into speech was initiated in computer science (Bolt 1980). Coincidentally, these early works used typed feature structure

[5] In languages like German, the difference between free gesticulation and sign language signs is also reflected terminologically: the former are called *Gesten*, the latter *Gebärden*.

descriptions akin to the descriptive format used in HPSG grammars. Though linguistically limited, the crucial invention has been a *multimodal chart parser*, that is, an extension of chart parsing that allows the processing of input in two modalities (namely speech and gesture). Such approaches are reviewed in Section 3.2. Afterwards, a more elaborate gesture representation format is introduced that makes it possible to encode the observable form of a gesture in terms of kinematically derived attribute-value structures (Section 3.3). Following the basic semiotic distinction between deictic (or indicating or pointing) gestures and iconic (or representational or imagistic) gestures, the analysis of each class of gestures is exemplified in Sections 3.4 and 3.5, respectively. To begin with, however, some basic phenomena that should be covered by a multimodal grammar are briefly summarized in Section 3.1.

3.1 Basic empirical phenomena of grammatical gesture integration

With regard to grammar-gesture integration, three main phenomena have to be dealt with:

- What is the meaning of a gesture? On which grounds should semantic representations or truth conditions be assigned to hand and arm movements?
- What is the affiliate of a gesture, that is, its verbal attachment site?
- What is the result of multimodal integration, that is, the outcome of composing verbal and non-verbal meanings?

Given the linguistic significance of gestures as sketched in the preceding sections, formal grammar- and semantic-oriented accounts of speech-gesture integration have recently been developed that try to deal with (at least one of) the three basic phenomena, though with different priorities, including Alahverdzhieva (2013), Alahverdzhieva & Lascarides (2010), Ebert (2014), Giorgolo (2010), Giorgolo & Asudeh (2011), Lücking (2013; 2016), Rieser (2008; 2011; 2015), Rieser & Poesio (2009) and Schlenker (2018). It should be noted that the first basic question does not have to be considered a question for grammar, but can be delegated to a foundational theory of gesture meaning. Here gestures turn out to be like words again, where "semantic theory" can refer to explaining meaning (foundational) or specifying meaning (descriptive) (Lewis 1970: 19). In any case, the HPSG-related approaches are briefly reviewed below.

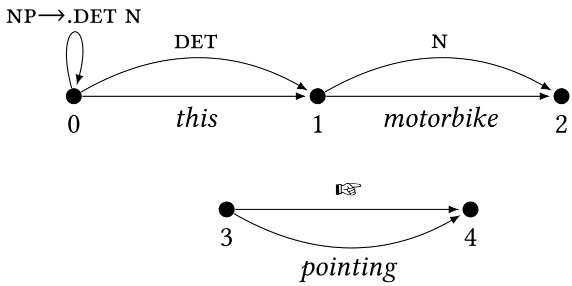
3.2 Precursors

Using typed feature structure descriptions to represent the form and meaning of gestures goes back to computer science approaches to human-computer interaction. For instance, the *QuickSet* system (Cohen et al. 1997) allows users to operate on a map and move objects or lay out barbed wires (the project was funded by a grant from the US army) by giving verbal commands and manually indicating coordinates. The system processes voice and pen (gesture) input by assigning signals from both media representations in the form of attribute-value matrices (AVMs) (Johnston 1998; Johnston et al. 1997). For instance, *QuickSet* will move a vehicle to a certain location on the map when asked to *Move this*[☞] *motorbike to here*[☞], where '☞' represents an occurrence of touch gesture (i.e. pen input).

Since a conventional constrained-based grammar for speech-only input rests on a "unimodal" parser, Johnston (1998) and Johnston et al. (1997) developed a *multimodal chart parser*, which is still a topic of computational linguistics (Alahverdzhieva et al. 2012) (see also Bender & Emerson 2020, Chapter 28 of this volume). A multimodal chart parser consists of two or more layers and allows for layer-crossing charts. The multimodal NP *this*[☞] *motorbike*, for instance, is processed in terms of a multimodal chart parser covering a speech (s) and a gesture (g) layer:

(9) s:

g:



A multimodal chart or *multichart* is defined in terms of sets of identifiers from both layers. Possible multicharts from (9) include the following ones:

(10) multichart 1: {[s,0,1], [g,3,4]}
multichart 2: {[s,1,2], [g,3,4]}
…

The basic rule for integrating spatial gestures with speech commands is the *basic integration scheme* (Johnston 1998; Johnston et al. 1997), reproduced in (11):

(11)

$$
\left[\begin{array}{ll}
\text{LHS}: & \left[\begin{array}{ll} \text{CAT}: & \textit{command}\\ \text{MODALITY}: & \boxed{2}\\ \text{CONTENT}: & \boxed{1}\\ \text{TIME}: & \boxed{3}\end{array}\right]\\
\text{RHS}: & \left[\begin{array}{ll}
\text{DTR1}: & \left[\begin{array}{ll} \text{CAT}: & \textit{located\_command}\\ \text{MODALITY}: & \boxed{6}\\ \text{CONTENT}: & \boxed{1}[\textit{location}\ \boxed{5}]\\ \text{TIME}: & \boxed{7}\end{array}\right]\\
\text{DTR2}: & \left[\begin{array}{ll} \text{CAT}: & \textit{spatial\_gesture}\\ \text{CONTENT}: & \boxed{5}\\ \text{MODALITY}: & \boxed{9}\\ \text{TIME}: & \boxed{10}\end{array}\right]
\end{array}\right]\\
\text{CONSTRAINTS}: & \left\{\begin{array}{l} \text{OVERLAP}(\boxed{7},\boxed{10}) \vee \text{FOLLOW}(\boxed{7},\boxed{10},4\text{S})\\ \text{TOTAL-TIME}(\boxed{7},\boxed{10},\boxed{3})\\ \text{ASSIGN-MODALITY}(\boxed{6},\boxed{9},\boxed{2})\end{array}\right\}
\end{array}\right]
$$

The AVM in (11) implements a mother-daughter structure along the lines of a context-free grammar rule, where a left-hand side (LHS) expands to a right-hand side (RHS). The right-hand side consists of two constituents (daughters DTR1 and DTR2), a verbal expression (*located_command*) and a gesture. The semantic integration between both modalities is achieved in terms of structure sharing, see tag $\boxed{5}$: the spatial gesture provides the location coordinate for the verbal command.

The bimodal integration is constrained by a set of restrictions, mainly regulating the temporal relationship between speech and gesture (see tags $\boxed{7}$ and $\boxed{10}$ in the CONSTRAINTS set): the gesture may overlap with its affiliated word in time, or follow it in at most four seconds (see the 4S under CONSTRAINTS). An integration scheme highly akin to that displayed in (11) also underlies current grammar-oriented approaches to deictic and iconic gestures (see Sections 3.4 and 3.5 below).

3.3 Representing gestures with AVMs

Representing the formal features of gestures in terms of attribute-value matrices has been initiated in robotics (Kopp et al. 2004). A representation format that captures the “phonological”, physical-kinematic properties of a gesture is designed according to the moveable junctions of arms and hands. For instance, the representation of the gesture in Figure 3 according to the format used in Lücking et al. (2010) is given in (12):

(12)
$$\begin{bmatrix} \textit{right hand} & \\ \textsc{handshape} & \begin{bmatrix} \textsc{shape} & \textit{G} \\ \textsc{path} & \textit{0} \\ \textsc{dir} & \textit{0} \end{bmatrix} \\ \textsc{palm} & \begin{bmatrix} \textsc{orient} & \textit{PAB>PAB/PUP>PAB} \\ \textsc{path} & \textit{0} \\ \textsc{dir} & \textit{0} \end{bmatrix} \\ \textsc{boh} & \begin{bmatrix} \textsc{orient} & \textit{BUP>BTB/BUP>BUP} \\ \textsc{path} & \textit{arc>arc>arc} \\ \textsc{dir} & \textit{MR>MF>ML} \end{bmatrix} \\ \textsc{wrist} & \begin{bmatrix} \textsc{position} & \textit{P-R} \\ \textsc{path} & \textit{line} \\ \textsc{dir} & \textit{MU} \\ \textsc{dist} & \textit{D-EK} \\ \textsc{extent} & \textit{small} \end{bmatrix} \\ \textsc{sync} & \begin{bmatrix} \textsc{config} & \textit{BHA} \\ \textsc{rel.mov} & \textit{LHH} \end{bmatrix} \end{bmatrix}$$

The formal description of a gestural movement is given in terms of the handshape, the orientations of the palm and the back of the hand (BOH), the movement trajectory (if any) of the wrist and the relation between both hands (synchronicity, SYNC). The handshape is drawn from the fingerspelling alphabet of American Sign Language, as illustrated in Figure 6. The orientations of palm and back of hand are specified with reference to the speaker's body (e.g. *PAB* encodes "palm away from body" and *BUP* encodes "back of hand upwards"). Movement features for the whole hand are specified with respect to the wrist: the starting position is given and the performed trajectory is encoded in terms of the described path and the direction and extent of the movement. Position and extent are given with reference to the *gesture space*, that is, the structured area within the speaker's immediate reach (McNeill 1992: 86–89) – see the left-hand side of Figure 7. Originally, McNeill considered the gesture space as "a shallow disk in front of the speaker, the bottom half flattened when the speaker is seated" (McNeill 1992: 86). However, also acknowledging the distance of the hand from the speaker's body (feature DIST) turns the shallow disk into a three-dimensional space, giving rise to the three-dimensional model displayed on the right-hand side of Figure 7. The gesture space regions known as *center-center*, *center* and *periphery*, possibly changed by location modifiers (*upper right*, *right*, *lower right*, *upper left*, *left*, *lower left*), are now modelled as nested cuboids. Thus, gesture space is structured according to all three body axes: the sagittal, the longitudinal and the transverse axes. Annotations straightforwardly transfer to the three-dimensional gesture

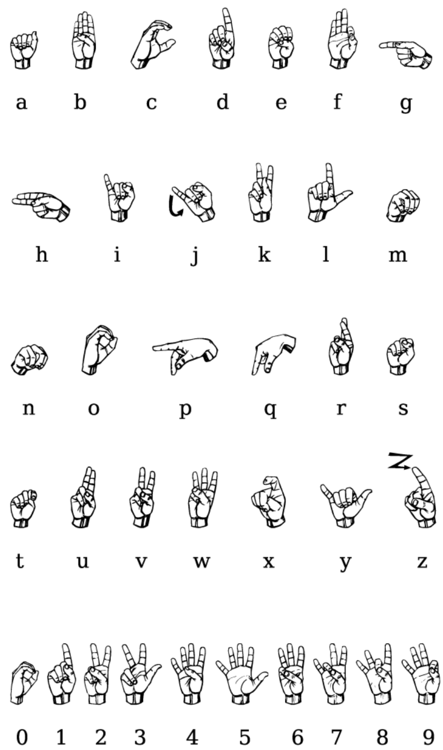


Figure 6: American Sign Language fingerspelling alphabet (image released to Public Domain by user Ds13 in the English Wikipedia at 18th December 2004, https://commons.wikimedia.org/wiki/File:Asl_alphabet_gallaudet.png)

space model. Such a three-dimensional gesture space model is assumed throughout this chapter. Complex movement trajectories through the vector space can describe a rectangular or a roundish path (or mixtures of both). Both kinds of movements are distinguished in terms of *line* or *arc* values of feature PATH. An example illustrating the difference is given in Figure 8. A brief review of gesture annotation can be found in Section 4.1.

3.4 Pointing Gestures

Pointing gestures are *the* prototypical referring device: they probably pave a way to reference in both evolutionary and language acquisition perspectives (Bruner

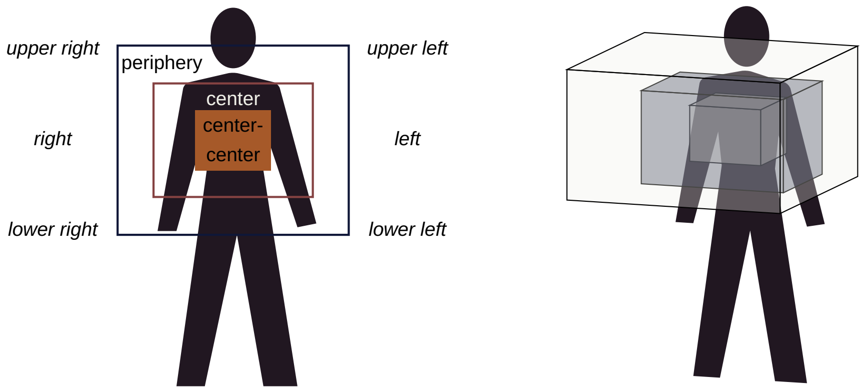


Figure 7: Gesture Space (left hand side is simplified from McNeill 1992: 89). Although originally conceived as a structured "shallow disk" McNeill (1992: 86), adding distance information gives rise to a three-dimensional gesture space model as illustrated on the right-hand side.

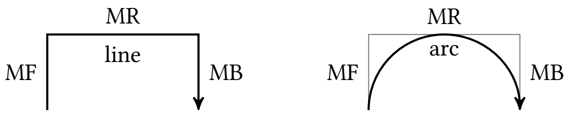


Figure 8: The same sequence of direction labels can give rise to an open rectangle or a semicircle, depending on the type of concatenation (Lücking 2016: 385).

1998; Masataka 2003; Matthews et al. 2012); they are predominant inhabitants of the "deictic level" of language, interleaving the symbolic (and the iconic) levels (Levinson 2008, see also Bühler 1934); they underlie reference in *Naming Games* in computer simulation approaches (Steels 1995) (for a semantic assessment of naming and categorisation games, see Lücking & Mehler 2012).

With regard to deictic gestures, Fricke (2012: Sec. 5.4) argues that deictic words within noun phrases – her prime example is German *so* 'like this' – provide a *structural*, that is, *language-systematic* integration point between the vocal plane of conventionalized words and the non-vocal plane of body movement. Therefore, with this conception, not only utterance production but *grammar* is inherently multimodal.

The referential import of the pointing gesture has been studied experimentally in some detail (Bangerter & Oppenheimer 2006; Kranstedt et al. 2006b,a; van der Sluis & Krahmer 2007). As a result, it turns out that pointings do not rely on a direct "laser" or "beam" mechanism (McGinn 1981). Rather, they serve a (more or less rough) locating function (Clark 1996) that can be modelled in terms of a *pointing cone* (Kranstedt et al. 2006b; Lücking et al. 2015). These works provide an answer to the first basic question (cf. Section 3.1): pointing gestures have a “spatial meaning” which focuses or highlights a region in relation to the direction of the pointing device. Such a spatial semantic model has been introduced in Rieser (2004) under the name of *region pointing*, where the gesture adds a locational constraint to the restrictor of a noun phrase. In a related way, two different functions of a pointing gesture have been distinguished by Kühnlein et al. (2002), namely singling out an object (13a) or making an object salient (13b).

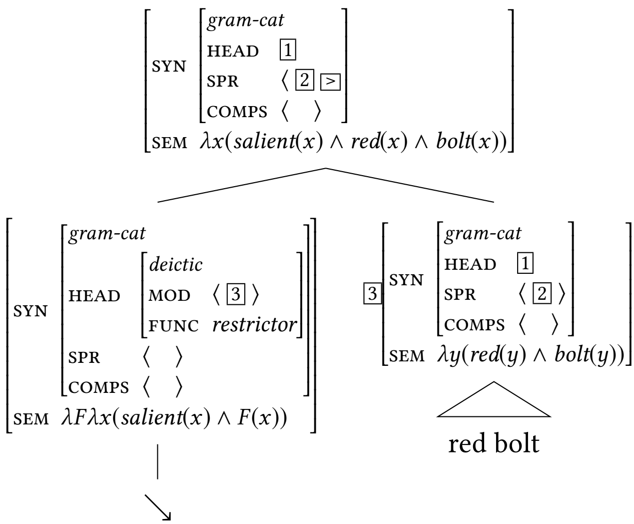
(13) a. $\lambda F \lambda x(x = c \wedge F(x))$

b. $\lambda F \lambda x(\mathit{salient}(x) \wedge F(x))$

The approach is expressed in lambda calculus and couched in an HPSG framework. The derivation of the instruction *Take the red bolt* plus a pointing gesture is exemplified in (14). A pointing gesture is represented by means of “$\searrow$” and takes a syntactic position within the linearized inputs according to the start of the stroke phase. For instance, the pointing gesture in (14a) occurred after *the* has been articulated but before *red* is finished. The derivation of the multimodal N′ constituent is shown in (14b).

(14) a. Take [the $\searrow$[N′ [N′ red bolt]]]

b.



The spatial model is also adopted in Lascarides & Stone (2009), where the region denoted by pointing is represented by a vector $\vec{p}$. This region is an argument

Figure 9: *Und man*[*chmal ist da auch ein EISverkäufer*] ('and some[times there's an ICE cream guy]'), [V5, 7:20]

to function v, however, which maps the projected cone region to $v(\vec{p})$, the space-time talked about, which may be different from the gesture space (many more puzzles of local deixis are collected by Klein 1978 and Fricke 2007).

Let us illustrate some aspects of pointing gesture integration by means of the real world example in (15) and Figure 9, taken from dialogue V5 of the SaGA corpus.

(15) Und man[chmal ist da auch ein EISverkäufer].
And some[times] is there also an ICE cream guy.
'And sometimes there's an ice cream guy'

The context in which the gesture appears is the following: the speaker describes a route which goes around a pond. He models the pond with his left hand, a post-stroke hold (cf. Figure 5) held over several turns. After having drawn the route around the pond with his right hand, the pointing gesture in Figure 9 is produced. The pointing indicates the location of an ice cream vendor in relation to the pond modelled in gesture space. Such instances of indirect or proxy pointing have been interpreted as *dual points* by Goodwin (2003); in standard semantics they are analysed in terms of *deferred reference*, where one thing is indicated but another but related thing is referred to (Quine 1950; Nunberg 1993). The "duality" or "deference" involved in the datum consists of a mapping from the location indicated in gesture space onto a spatial area of the described real world situation. Such mappings are accounted for by the function v that shifts the pointing cone area from gesture space $\vec{p}$ to some other space $v(\vec{p})$ (Lascarides & Stone 2009). So, the deictic gesture locates the ice cream vendor. Since it is held during nearly the whole utterance, its affiliate expression "Eisverkäufer" (*ice cream guy*) is picked out due to carrying primary accent (indicated by capitalization).[6] Within HPSG, such constraints can be formulated within an interface to metrical trees from the phonological model of Klein (2000) or phonetic information packing from Engdahl & Vallduví (1996) – see also De Kuthy (2020), Chapter 24 of this volume. The well-developed basic integration scheme of Alahverdzhieva et al. (2017: 445) rests on a strict speech and gesture overlap and is called the *Situated Prosodic Word Constraint*, which allows the combination of a speech daughter (S-DTR) and a gesture daughter (G-DTR) – see Figure 10. The Situated Prosodic Word Constraint applies to both deictic and iconic gestures. Under certain conditions, including when a deictic gesture is direct (i.e. $\vec{p} = v(\vec{p})$), however, the temporal and prosodic constraints can be relaxed for pointings.

$$
\begin{bmatrix}
\textit{word} & \\
\text{OVERLAP} & \langle @7, @8 \rangle \\
\text{TIME} & @7 \cup @8 \\
\text{PHON} & @3 \\
\text{SYNSEM} & \begin{bmatrix} \text{CAT} & @5 \\ \text{CONT} & \begin{bmatrix} \text{RELS} & \boxed{C_{rel}} \oplus \boxed{S_{rel}} \oplus \boxed{G_{rel}} \\ \text{HCONS} & \boxed{S_{hc}} \oplus \boxed{G_{hc}} \end{bmatrix} \end{bmatrix} \\
\text{S-DTR} & \begin{bmatrix} \textit{word} & \\ \text{TIME} & @7 \\ \text{PHON} & @3\textit{nuclear\_or\_pre-nuclear} \\ \text{SYNSEM} & \begin{bmatrix} \text{CAT} & @5 \\ \text{CONT} & \begin{bmatrix} \text{RELS} & \boxed{S_{rel}} \\ \text{HCONS} & \boxed{S_{hc}}\textit{hcons} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\text{G-DTR} & \begin{bmatrix} \textit{depicting\_or\_deictic} & \\ \text{TIME} & @8 \\ \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{G-FEATURE} & \textit{value} \\ \ldots & \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{RELS} & \boxed{G_{rel}} \\ \text{HCONS} & \boxed{G_{hc}} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\text{C-CONT.RELS} & \boxed{C_{rel}}
\end{bmatrix}
$$

Figure 10: *Situated Prosodic Word Constraint* (Alahverdzhieva et al. 2017: 445)

[6] Semantically, other integration points are possible, too, most notably with "da" (*there*). However, the intonation-based integration point patterns well with observations of the affiliation behaviour of iconic gestures, as indicated with respect to examples (5) and (6) in Section 1. Concerning deictic gestures, a constraint that favours affiliation to deictic words over affiliation to stressed words (if they differ at all) seems conceivable nonetheless.

In order to deal with gestures that are affiliated with expressions that are larger than single words, Alahverdzhieva et al. (2017) also develop a phrase or sentence level integration scheme, where the stressed element has to be a semantic head (in the study of Mehler & Lücking 2012, 18.8% of the gestures had a phrasal affiliate). In this account, the affiliation problem (the second desideratum identified in Section 3.1) has a well-motivated solution on both the word and the phrasal levels, at least for temporally overlapping speech-gesture occurrences (modulo the conditioned relaxations for pointings). Semantic integration of gesture location and verbal meaning (the third basic question from Section 3.1) is brought about using the underspecification mechanism of *Robust Minimal Recursion Semantics* (RMRS), a refinement of *Minimal Recursion Semantics* (MRS) (Copestake et al. 2005), where basically scope as well as arity of elementary expressions is underspecified (Copestake 2007) – see the RELS and HCONS features in Figure 10. For some background on (R)MRS see the above given references, or see Koenig & Richter (2020), Chapter 23 of this volume.

A dialogue-oriented focus on pointing is taken in Lücking (2018): here, pointing gestures play a role in formulating processing instructions that guide the addressee in where to look for the referent of demonstrative noun phrases.

3.5 Iconic Gestures

There is nearly no semantic work on the grounds according to which the meanings assigned to iconic gestures should be assigned to them in the first place (this is the first basic question from Section 3.1). Semantic modelling usually focuses on the interplay of (in this sense presumed) gesture content with speech content, that is, on the third of the basic questions from Section 3.1. Schlenker (2018: 296) is explicit in this respect: "It should be emphasized that we will not seek to explain how a gesture […] comes to have the content that it does, but just ask how this content interacts with the logical structure of a sentence".[7] Two exceptions, however, can be found in the approaches of Rieser (2010) and Lücking (2013; 2016). Rieser (2010) tries to extract a "depiction typology" out of a speech-and-gesture corpus where formal gesture features are correlated with topological clusters consisting of geometrical constructs. Thus, he tries to address the first basic question from Section 3.1 in terms of an empirically extracted gesture typology. These geometrical objects are used in order to provide a possibly underspecified semantic representation for iconic gestures, which is then integrated into word meaning via lambda calculus (Hahn & Rieser 2010; Rieser 2011). The work of Lücking (2013;

[7] The omission indicated by "[…]" just contains a reference to an example in the quoted paper.

2016) is inspired by Goodman's notion of *exemplification* (Goodman 1976), that is, iconic gestures are connected to semantic predicates in terms of a reversed denotation relation: the meaning of an iconic gesture is given in terms of the set of predicates which have the gesture event within their denotation. In order to make this approach work, common perceptual features for predicates are extracted from their denotation and represented as part of a lexical extension of their lexemes, serving as an interface between hand and arm movements and word meanings. This conception in turn is motivated by psychophysic theories of the perception of biological events (Johansson 1973), draws on philosophical similarity conceptions beyond isomorphic mappings (Peacocke 1987),[8] and, using a somewhat related approach, has been proven to work in robotics by means of imagistic description trees (Sowa 2006). These perceptual features serve as the integration locus for iconic gestures, using standard unification techniques. The integration scheme for achieving this is the following one (Lücking 2013: 249) (omitting the time constraint used in the basic integration scheme in 11):

(16)
$$
\begin{bmatrix}
\textit{sg-ensemble} & \\
\textsc{phon} & \boxed{12} \\
\textsc{cat} & \boxed{2} \\
\textsc{cont} & \boxed{3}\begin{bmatrix}\textsc{restr} & \left\langle \ldots, \boxed{5}\begin{bmatrix}\textit{pred} & \\ \textsc{cvm} & \boxed{1}\end{bmatrix}, \ldots \right\rangle\end{bmatrix} \\
\textsc{s-dtr} & \begin{bmatrix}\textit{verbal-sign} & \\ \textsc{phon} & \boxed{12}\begin{bmatrix}\textsc{accent} & \boxed{6}\end{bmatrix} \\ \textsc{cat} & \boxed{2} \\ \textsc{cont} & \boxed{3}\end{bmatrix} \\
\textsc{g-dtr} & \begin{bmatrix}\textit{gesture-vec} & \\ \textsc{aff} & \left\langle \begin{bmatrix}\textsc{phon} & \begin{bmatrix}\textsc{accent} & \boxed{6}\,\textit{marked}\end{bmatrix}\end{bmatrix}\right\rangle \\ \textsc{traj} & \boxed{1} \\ \textsc{cont} & \begin{bmatrix}\textsc{mode} & \textit{exemplification} \\ \textsc{ex-pred} & \boxed{5}\end{bmatrix}\end{bmatrix}
\end{bmatrix}
$$

Comparable to a modifier, a gesture attaches to an affiliate via feature AFF, which in turn is required to carry intonational accent, expressed in terms of information packaging developed by Engdahl & Vallduví (1996) (cf. De Kuthy 2020, Chapter 24 of this volume). The semantic contribution of a gesture is contributed via the new semantic mode *exemplification*, that is, a gesture displays a predication from the RESTR list of its affiliate. The exemplification interface is established

[8] That mere resemblance, usually associated with iconic signs, is too empty a notion to provide the basis for a signifying relation has been emphasised on various occasions (Burks 1949; Bierman 1962; Eco 1976; Goodman 1976; Sonesson 1998).

using the format of vector semantics developed by Zwarts & Winter (2000) and Zwarts (2003) in order to capture the semantic contribution of locative prepositions, motion verbs and shape adjectives, among other things. This involves two steps: on the one hand, the representation of a gesture (cf. Section 3.3) is mapped onto a vectorial representation; on the other hand, the content of place and form predicates is enriched by abstract psychophysic information in the sense of Johansson (1973) (see above), also spelled out in terms of vector representations. Both steps are illustrated by means of the simple example shown in Figure 11, where the speaker produces a semicircle in both speech and gesture.

Figure 11: “und [oben haben die so’n HALBkreis]” (*and on the top they have such a SEMIcircle*), [V20, 6:36].

The kinematic gesture representation of the movement carried out (CARRIER) by the wrist – *move up*, *move left*, *move down*, which are concatenated (“$\oplus$”) by movement steps in a bent (“$\oplus_{\smile}$”, as opposed to rectangular “$\oplus_{\llcorner}$”) way (cf. also Figure 8) – is translated via vectorising function **V** into a vector trajectory (TRAJ(ECTORY)) from the three-dimensional vector space, cf. Figure 7:[9]

(17)
$$\begin{bmatrix} \textit{gesture-vec} & \\ \text{TRAJ} & \left[\mathbf{V}\boxed{1}\mathbf{=UP} \oplus_{\smile} \mathbf{-RT} \oplus_{\smile} \mathbf{-UP}\right] \\ \text{CARRIER} & \begin{bmatrix} \textit{gesture} & \\ \text{MORPH} & \left[\text{WRIST.MOV}\ \boxed{1}\textit{mu} \oplus_{\smile} \textit{ml} \oplus_{\smile} \textit{md}\right] \end{bmatrix} \end{bmatrix}$$

The lexical entry for *semicircle* is endowed with a *conceptual vector meaning* attribute CVM. Within CVM it is specified (or underspecified) what kind of vector (VEC) is at stake (axis vector, shape vector, place vector), and how it looks, that

[9] Vectors within gesture space can be conceived of as equivalence classes over concrete movement annotation predicates.

is, which PATH it describes. A semicircle can be defined as an axis vector whose path is a 180° trajectory. Accordingly, 180° is the root of a type hierarchy which hosts all vector sequences within gesture space that describe a half circle. This information is added in terms of a form predicate to the restriction list of *semicircle*, as shown in the speech daughter's (S-DTR) content (CONT) value in (18). Licensed by the speech-gesture integration scheme in (16), the half-circular gesture trajectory from (17) and its affiliate expression *semicircle* can enter into an ensemble construction, as shown in (18):

(18)
$$
\left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{sg-ensemble}} \\
\text{PHON} & \left\langle \boxed{1}\left[\begin{array}{ll}\text{SEG} & \textit{semicircle}\end{array}\right]\right\rangle \\
\text{CAT} & \boxed{2}\left[\begin{array}{ll}\text{HEAD} & \textit{noun}\end{array}\right] \\
\text{CONT} & \boxed{3}\left[\begin{array}{ll}
\text{INDEX} & i \\
\text{RESTR} & \left\langle \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{geom-obj-pred}} \\ \text{RELN} & \textit{semicircle} \\ \text{INST} & i\end{array}\right], \boxed{5}\left[\begin{array}{ll}\multicolumn{2}{l}{\textit{form-pred}} \\ \text{RELN} & \textit{round2} \\ \text{ARG} & i \\ \text{CVM} & \left[\begin{array}{ll}\text{VEC} & \textit{axis-path}(i,\boldsymbol{v}) \\ \text{PATH} & \boxed{7}\end{array}\right]\end{array}\right]\right\rangle
\end{array}\right] \\
\text{S-DTR} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{verbal-sign}} \\
\text{PHON} & \boxed{1}\left[\begin{array}{ll}\text{ACCENT} & \boxed{6}\end{array}\right] \\
\text{CAT} & \boxed{2} \\
\text{CONT} & \boxed{3}
\end{array}\right] \\
\text{G-DTR} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{gesture-vec}} \\
\text{AFF} & \left\langle\left[\begin{array}{ll}\text{PHON} & \left[\begin{array}{ll}\text{ACCENT} & \boxed{6}\textit{marked}\end{array}\right]\end{array}\right]\right\rangle \\
\text{TRAJ} & \boxed{7}\boldsymbol{UP} \oplus_{\smile} -\boldsymbol{RT} \oplus_{\smile} -\boldsymbol{UP} \\
\text{CONT} & \left[\begin{array}{ll}\text{MODE} & \textit{exemplification} \\ \text{EX-PRED} & \boxed{5}\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

By extending lexical entries with frame information from frame semantics (Fillmore 1982), also the exemplification of non-overtly-expressed predicates becomes feasible (Lücking 2013: Sec. 9.2.1); a datum showing this case has already been given with the *spiral staircases* in (5)/Figure 3. A highly improved version of the "vectorisation" of gestures with a translation protocol has been spelled out in Lücking (2016), but within the semantic framework of a *Type Theory with Records* (Cooper 2019; Cooper & Ginzburg 2015; cf. also Lücking, Ginzburg & Cooper 2020, Chapter 29 of this volume).

The richer formal, functional and representational features of iconic gestures as compared to deictic gestures (cf. Section 3.4) is accounted for in Alahverdzhieva et al. (2017) by assigning a formal predicate to each "phonological" feature of a gesture representation (cf. Section 3.3). These formal gesture predicates are highly underspecified, using *Robust Minimal Recursion Semantics* (RMRS) (Copestake 2007). That is, they can be assigned various predications (which are assumed to be constrained by iconicity with differing arity in the gesture resolution process).

Let us illustrate this by means of Example 1 from Alahverdzhieva et al. (2017), which is due to Loehr (2004) and re-given in (19), adapted to the representational conventions followed in this chapter.

(19) [So he mixes MUD]
The speaker performs a circular movement with the right hand over the upwards, open palm of the left hand

Using a variant of a kinematic representation format for gestures (cf. Section 3.3), the right hand from example 19 is notated as follows (Alahverdzhieva et al. 2017: 440):

(20)
$$\begin{bmatrix} \textit{depict-literal} & \\ \textsc{hand-shape} & \textit{bent} \\ \textsc{palm-orient} & \textit{towards-down} \\ \textsc{finger-orient} & \textit{towards-down} \\ \textsc{hand-location} & \textit{lower-periphery} \\ \textsc{hand-movement} & \textit{circular} \end{bmatrix}$$

Each feature value pair from the gesture's representation in (20) is mapped onto an RMRS-based underspecified representation (Alahverdzhieva et al. 2017: 442):

(21)
$$\begin{array}{l} l_0 : a_0 : [\mathcal{G}](h) \\ l_1 : a_1 : \textit{hand\_shape\_bent}(i_1) \\ l_2 : a_2 : \textit{palm\_orient\_towards\_down}(i_2) \\ l_3 : a_3 : \textit{finger\_orient\_towards\_down}(i_3) \\ l_4 : a_4 : \textit{hand\_location\_lower\_periphery}(i_4) \\ l_5 : a_5 : \textit{hand\_movement\_circular}(i_5) \\ h =_q l_n \textit{ where } 1 \leq n \leq 5 \end{array}$$

Note that all predicates mapped from the gesture in (21) fall within the scope of the scopal operator $[\mathcal{G}]$; this prevents an individual introduced by a depicting gesture from being an antecedent of a pronoun in speech.

Regimented by the *Situated Prosodic Word Constraint* from Figure 10, the underspecified semantic description of the gesture in (21) and its affiliated noun *mud* can enter into the multimodal construction given in Figure 12 (where the gesture features are partly omitted for the sake of brevity).

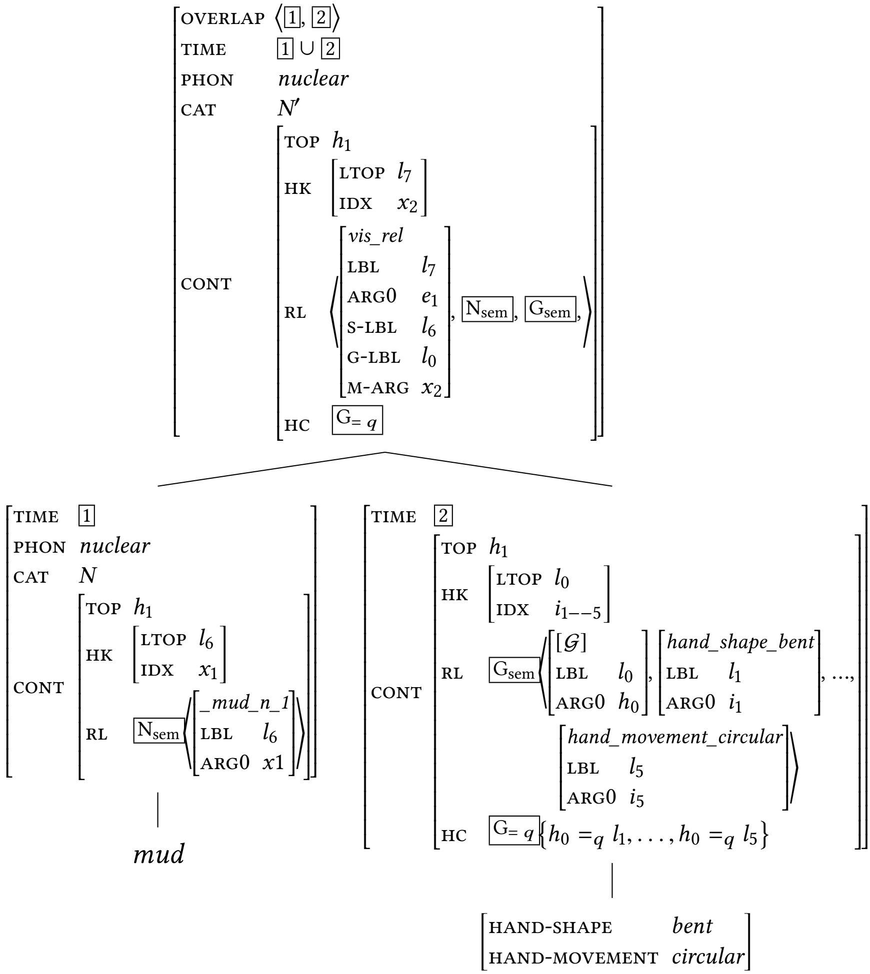


Figure 12: Derivation tree for depicting gesture and its affiliate noun *mud* (Alahverdzhieva et al. 2017: 447)

The underspecified RMRS predicates derived from gesture annotations are interpreted according to a type hierarchy rooted in those underspecified RMRS predicates. For example, the circular hand movement of the "mud gesture" can give rise to two slightly different interpretations: on the one hand, the circular hand movement can depict – in the context of the example – that mud is being mixed from an observer viewpoint (McNeill 1992). This reading is achieved by following the left branch of (22), where the gesture contributes a conjunction of predications that express that a substance rotates. When integrated with speech, the substance resolves to the mud and the rotating event to the mixing. On the other hand, the gesture can depict seen from the character viewpoint (McNeill 1992), which corresponds to the predication from the right branch of (22). Here the rotating event is brought about by agent j' which is required to be coreferential with *he*, the subject of the utterance.

(22)

$$\mathit{hand\_movement\_circular}(\mathrm{i})$$

$$\mathit{substance}(x') \wedge \mathit{rotate}(e', x') \qquad \mathrm{rotate}(e', j', x')$$

In addition to addressing (solving) the three basic questions identified in Section 3.1 – roughly, foundation of gesture meaning, regimenting affiliation, and characterisation of semantic integration – another issue has received attention recently, namely the projection behaviour of gestures when interacting with logical operators (Ebert 2014; Schlenker 2018). For instance, the unembedded gesture in (23) triggers the inference that the event being described actually happened in the manner in which it was gesticulated (Schlenker 2018: 303):

(23) John [*slapping gesture*] punished his son.
⇒ John punished his son by slapping him.

That is, (23) more or less corresponds to what semantic speech-gesture integration approaches, as briefly reviewed above, would derive as the content of the multimodal utterance.

Embedding the slapping gesture under the *none*-quantifier triggers, according to Schlenker (2018: 303), the following inference:

(24) None of these 10 guys [*slapping gesture*] punished his son.
⇒ for each of these 10 guys, if he had punished his son, this would have involved some slapping.

The universal inference patterns with presupposition. Unlike presupposition, however, Schlenker (2018: 303) claims that the inference is conditionalized on the at-issue contribution of (24), expressed by the *if*-clause. He then develops a notion of "cosupposition", which rests on an expression's local context that entails the content of its affiliated gesture. So far, there is no connection from such projections to HPSG, however.

Beyond being involved in pragmatic processes like inferring, gestures also take part in "micro-evolutionary" developments. Iconic gestures in particular are involved in a short-term dynamic phenomenon: on repeated co-occurrence, iconic gestures and affiliated speech can fuse into a *multimodal ensemble* (Kendon 2004; Lücking et al. 2008; Mehler & Lücking 2012). The characteristic feature of such an ensemble is that their gestural part, their verbal part, or even both parts can be simplified without changing the meaning of the ensemble. Ensembles, thus, are the result of a process of sign formation as studied, for instance, in experimental semiotics (Galantucci & Garrod 2011). Such grammaticalisation processes eventually might lead to conventional signs. However, most conventional, emblematic everyday gestures seem to be the result of circumventing a taboo: something you should not name is gesticulated (Posner 2002).

3.6 Other gestures

As noted in the taxonomy reviewed in Section 2, there are gestures that, unlike the deictic and iconic ones discussed in the previous sections, do not contribute to propositional content, but serve functions bound up with dialogue management. Such gestures have been called *interactive gestures* (Bavelas et al. 1992). Two examples are given in Figures 13 and 14, which have been discussed by Bavelas et al. (1995).

The "delivery gesture" in Figure 13 is used to underline an argument, or to refer to the fact that the current issue is known to the interlocutors. In the latter function, the gesture is also termed *shared information gesture*.

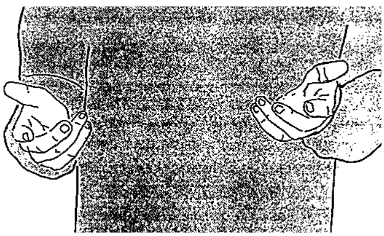


Figure 13: "Here's my point."

The 'open hand' pointing gesture in Figure 14 acts as a turn-taking device: it can function as a turn-assigning gesture (underlined by the caption of Figure 14), or, when used to point at the current speaker, it can also indicate that the gesturer wants to take the turn and address the current turn holder.

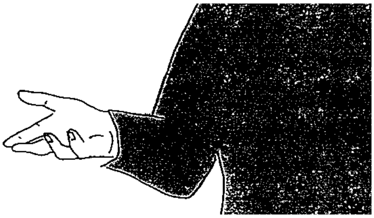


Figure 14: "You go ahead."

So far there is no account of interactive gestures in HPSG. Given their entrenchment in dialogue processes, their natural home seems to be in a dialogue theory, anyway (see Lücking, Ginzburg & Cooper 2020, Chapter 29 of this volume). Accordingly, what is presumably the only formal approach to some of these gestures has been spelled out within the dialogical framework PTT in Rieser & Poesio (2009).

4 Gesture and …

Besides being of a genuine linguistic, theoretical interest, gesture studies are a common topic in various areas of investigation, some of which are briefly pointed at below.

4.1 … tools, annotation, corpora

Since gestures are signs in the visual modality, they have to be videotaped. Gesture annotation is carried out on the recorded video films. The main tools that allow for video annotation are, in alphabetical order, Anvil[10], ELAN[11] and EXMARaLDA[12].

Annotation should follow an annotation standard which is specified in an annotation scheme. Various annotation schemes for gestures and speech-gesture integration have been proposed, partly differing in annotation foci, including the

[10] https://www.anvil-software.org/, Kipp 2014.

[11] https://tla.mpi.nl/tools/tla-tools/elan/, Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands, Sloetjes & Wittenburg 2008.

[12] https://exmaralda.org/, Schmidt 2012.

following ones: annotation schemes that focus on form description and gestures classification in terms of a taxonomy like the one introduced in Section 2 have been developed by R. Breckenridge Church, published in the appendix of McNeill (1992); CoGEST (Gibbon et al. 2003); FORM (Martell et al. 2002) and the SaGA annotation (Lücking et al. 2013). The form of gestures and their timing with speech is the object of the coding scheme of Kipp et al. (2007). An interaction-oriented scheme has been proposed by Allwood et al. (2007), which is formulated on the level of turns and dialogue management. A detailed annotation scheme for the form and function of gestures has been developed in terms of “annotation decision trees” within the NEUROGES system (Lausberg & Sloetjes 2009).

Annotated videos of real life interactions give rise to multimodal corpora. Among those that include data on gestures are the following ones.

The multimodal SmartKom Corpus (Schiel et al. 2003), which grew out of the SmartKom project (Wahlster 2006), comprises recording sessions of various Wizard-of-Oz experiments (that is, human-computer interaction where the human participant is made to believe that the system she or he interacts with is autonomous while in fact it is, at least partly, operated by another human). Recordings are extended basically by a transliteration and labelling of natural speech, labelling of gestures and annotation of user states (in the corpus’ first release). The first public release, SKP 1.0, contains 90 recording sessions of 45 users. The multimodal SmartKom corpus as well as further SmartKom resources are hosted at the *Bavarian Archive for Speech Signals* (https://www.bas.uni-muenchen.de/Bas/).

The AMI Meeting Corpus (Carletta et al. 2006) consists of 100 hours of meeting recordings. The meetings were recorded in English but include mostly non-native speakers. The AMI Meeting Corpus provides orthographic transcriptions, but also has a couple of further annotations, including dialogue acts, named entities, head gesture, hand gesture, gaze direction, movement and emotional states.

The SaGA (“Speech and Gesture Alignment”) corpus consists of 24 German route direction dialogues obtained after a bus ride through a virtual town (Lücking et al. 2010). Audio and video data from the direction-giver were recorded. The SaGA corpus consists of 280 minutes of video material containing 4,961 iconic/deictic gestures, approximately 1,000 discourse gestures and 39,435 word tokens (Lücking et al. 2013). Gesture annotation has been carried out in great detail, following a kinematic, form-based approach (cf. the above-given remark on annotation schemes). Part of the SaGA corpus is available from the *Bavarian Archive for Speech Signals* (https://www.bas.uni-muenchen.de/Bas).

The DUEL (“Disfluency, exclamations and laughter in dialogue”) corpus (Hough et al. 2016) comprises 24 hours of natural, face-to-face dialogue in German, French and Mandarin Chinese. It includes audio, video and body tracking data and is transcribed and annotated for disfluency, laughter and exclamations.

The FIGURE (derived from "Frankfurt Image GestURE") corpus (Lücking et al. 2016) is built on recordings of 50 participants with various mother tongues (though mostly German) spontaneously producing gestures in response to five or six terms from a total of 27 stimulus terms, which have been compiled mainly from image schemata (Lakoff 1987). The gestures have been kinematically annotated by means of a variant of the SaGA annotation scheme. The FIGURE annotation is available from the Text Technology Lab Frankfurt (https://www.texttechnologylab.org/applications/corpora).

4.2 … robots and virtual agents

In the context of Human-Computer Interaction (HCI) or Human-Robot Interaction (HRI), gesture plays an important role (in fact, the formal modelling of deictic and iconic gestures has been initiated in these fields, cf. Section 3.2). One reason for this prominence of gesture in technical areas is that people who interact with a robot evaluate it more positively when the robot displays non-verbal behaviours such as hand and arm gestures along with speech (see e.g. Salem et al. 2012). Within HCI/HRI, two kinds of distinctions have to be made. The first is a distinction between "robot" in the sense of virtual avatars and "robot" in the (probably more common) sense of physical devices (only the latter will be henceforth called a "robot"). The second distinction discerns gesture generation from gesture recognition. Given this simple systematization, altogether four divisions of gesture and virtual avatars/robots arise (references are just exemplary and preferably from earlier HCI/HRI times): (i) gesture generation by robots (e.g. Le et al. 2011); (ii) gesture recognition by robots (e.g. Triesch & von der Malsburg 1998); (iii) gesture generation by virtual avatars (e.g. Cassell et al. 2000); and (iv) gesture recognition in VR/AR (e.g. Weissmann & Salomon 1999). For a more detailed overview see Lücking & Pfeiffer (2012). Enabling humans to act and interact in virtual rooms (e.g. Pfeiffer et al. 2018) can be seen as recent extension of gesture use in HCI/HRI.

In order to plan and design the speech/gesture output of a virtual avatar or a robot, a multimodal representation format is required. To this end, the *Multimodal Utterances Representation Markup Language* for conversational robots (MURML) has been developed (Kranstedt et al. 2002). A similar purpose is served by the *Extensible MultiModal Annotation* (EMMA; Johnston 2009).

4.3 … learning

Following a "gesture as a window to the mind" view, gestures must be a prime object of educational theory and practice, and they are indeed, as demonstrated by research of Cook & Goldin-Meadow (2006) and colleagues. Effectiveness of gestures has been studied in math lessons (Goldin-Meadow et al. 2001), in the acquisition of counting competence (Alibali & DiRusso 1999) and in bilingual education (Breckinridge Church et al. 2004), among other areas. The fairly unanimous result is that gestures can indeed reflect students' conceptualisations and provide insights into cognitive processes involved in learning. Therefore, they can be used as a teaching device as well as an indicator of learning progress and understanding.

4.4 … aphasia

Current models of utterance production are speech-gesture production models, assuming a (more or less) integrated generation of multimodal utterances. Based on such models, one expects an effect on gesture performance when speech production is impaired, as is the case with aphasic speakers. Aphasia is an acquired speech disorder, which can be caused by a stroke, ischaemia, haemorrhage, craniocerebral trauma and further brain-damaging diseases. Different speech-gesture production models make slightly different predictions for speakers suffering from aphasia and can be evaluated accordingly (de Ruiter & de Beer 2013). Indeed, observing the gesture behaviour of aphasic speakers is one aspect of gesture and aphasia (Jakob et al. 2011; Kong et al. 2017; Sekine & Rose 2013). With the exception of the growth point theory, speech-gesture production models are based on Levelt's (1989) model.

The *Growth Point model* (McNeill & Duncan 2000) assumes that the "seed" of an utterance is an inherently multimodal idea unit that comprises imagistic as well as symbolic proto-representations which unfold into gesture and speech respectively in the process of articulation (see also Röpke 2011 on the growth point's entrenchment in contexts and frames).

The *Sketch model* (de Ruiter 2000) reflects explicitly different kinds of gestures (see Section 2). Its name is due to the sketch component, an abstract spatio-temporal representation alongside Levelt's preverbal message. Independently from each other, the sketch is sent to a gesture planner, while the preverbal message is processed by the formulator.

According to the *Lexical Access model* of Krauss et al. (2000), iconic gestures are related to words and are used in order to facilitate speaker-internal word retrieval rather than communicating pictorial information.

The *Interface model* (Kita & Özyürek 2003) assumes that the processes for speech and gesture generation negotiate with each other and therefore can influence each other during the production phase.

Other aspects include the use of gesture in speech therapy. Very much in line with the lexical access model, gestures have been used in order to facilitate word retrieval in what can be called *multimodal therapy* (Rose 2006). Following a different strategy, gestures are also used in order to enhance the communicative range of patients: they learn to employ gestures instead of words in order to communicate at least some of their needs and thoughts more fluently (Cubelli et al. 1991; Caute et al. 2013).

However, just counting on gestures in therapy does not automatically lead to success (Auer & Bauer 2011). The type and severity of aphasia, the individual traits of the aphasic speaker and the kinds of gestures impaired or still at disposal, among other factors, seem to constitute a complex network for which currently no generally applicable clinical pathway can be given.

5 Outlook

What are (still) challenging issues with respect to grammar-gesture integration, in particular from a semantic point of view? Candidates include:

- gestalt phenomena: the trajectories described by a gesture are often incomplete and have to be completed by drawing on gestalt principles or everyday knowledge (Lücking 2016).
- negligible features: not all formal features of a gesture are meaning-carrying features in the context of utterance. For instance, in a dynamic gesture the handshape often (though not always) does not provide any semantic information (cf. also examples (17) and (21)/(22)). How can we distinguish between significant and negligible gesture features?
- "semantic endurance": due to holds, gestures can show their meaning contributions for some period of time and keeps available for semantic attachment. This may call for a more sophisticated algebraic treatment of speech-gesture integration than offered by typed feature structures (Rieser 2015).

Finally, the empirical domain of "gesture" has to be extended to other non-verbal signals, in particular propositional ones such as laughter (Ginzburg et al.

2015), facial expressions or gaze (see Section 1 for a brief list of non-verbal signals), in isolation as well as in mutual combination. Thus, there is still some way to go in order to achieve a fuller understanding of natural language interaction and thereby natural languages.

Acknowledgements

This work is partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the program "Investissements d'Avenir" (reference: ANR-10-LABX-0083). It contributes to the IdEx Université de Paris – ANR-18-IDEX-0001. For insightful comments on earlier drafts I want to thank Anne Abeillé, Jonathan Ginzburg, Alex Lascarides, Stefan Müller, Hannes Rieser, and Markus Steinbach. They helped to improve the chapter a lot. Needless to say that all remaining oddities or shortcomings are my own. Furthermore, I am grateful to Elizabeth Pankratz for attentive remarks and for proofreading.

References

Abner, Natasha, Kensy Cooperrider & Susan Goldin-Meadow. 2015. Gesture for linguists: A handy primer. *Language and Linguistics Compass* 9(11). 437–451. DOI:10.1111/lnc3.12168

Alahverdzhieva, Katya. 2013. *Alignment of speech and co-speech gesture in a constraint-based grammar*. School of Informatics, University of Edinburgh dissertation.

Alahverdzhieva, Katya, Dan Flickinger & Alex Lascarides. 2012. Multimodal grammar implementation. In Eric Fosler-Lussier, Ellen Riloff & Srinivas Bangalore (eds.), *Proceedings of the 2012 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies* (NAACL-HLT 2012), 582–586. Montreal, Canada.

Alahverdzhieva, Katya & Alex Lascarides. 2010. Analysing language and co-verbal gesture in constraint-based grammars. In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot*, 5–25. Stanford, CA: CSLI Publications. http://csli-publications.stanford.edu/HPSG/2010/, accessed 2018-2-25.

Alahverdzhieva, Katya, Alex Lascarides & Dan Flickinger. 2017. Aligning speech and co-speech gesture in a constraint-based grammar. *Journal of Language Modelling* 5(3). 421–464.

Alibali, Martha W. & Alyssa A. DiRusso. 1999. The function of gesture in learning to count: More than keeping track. *Cognitive Development* 14(1). 37–56. DOI:10.1016/S0885-2014(99)80017-3

Allwood, Jens, Loredana Cerrato, Kristiina Jokinen, Costanza Navarretta & Patrizia Paggio. 2007. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation* 41. 273–287. DOI:10.1007/s10579-007-9061-5

Argyle, Michael. 1975. *Bodily communication*. New York, NY: Methuen & Co.

Argyle, Michael & Mark Cook. 1976. *Gaze and mutual gaze*. Cambridge, UK: Cambridge University Press.

Auer, Peter & Angelika Bauer. 2011. Multimodality in aphasic conversation: Why gestures sometimes do not help. *Journal of Interactional Research in Communication Disorders* 2(2). 215–243. DOI:10.1558/jircd.v2i2.215

Bangerter, Adrian & Daniel M. Oppenheimer. 2006. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture* 6(1). 85–102.

Bavelas, Janet B., Nicole Chovil, Linda Coates & Lori Roe. 1995. Gestures specialized for dialogue. *Personality and Social Psychology Bulletin* 21(4). 394–405.

Bavelas, Janet B., Nicole Chovil, Douglas A. Lawrie & Allan Wade. 1992. Interactive gestures. *Discourse Processes* 15(4). 469–489. DOI:10.1080/01638539209544823

Bavelas, Janet B., Jennifer Gerwing, Chantelle Sutton & Danielle Prevost. 2008. Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language* 58(2). 495–520. DOI:10.1016/j.jml.2007.02.004

Bender, Emily M. & Guy Emerson. 2020. Computational linguistics and grammar engineering. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 917–961. Berlin: Language Science Press. DOI:??

Bierman, Arthur K. 1962. That there are no iconic signs. *Philosophy and Phenomenological Research* 23(2). 243–249.

Birdwhistell, Ray L. 1970. *Kinesics and context: Essays on body motion communication*. Philadelphia: University of Pennsylvania Press.

Bolt, Richard A. 1980. "put-that-there": voice and gesture at the graphics interface. *SIGGRAPH Comput. Graph.* 14. 262–270. DOI:http://doi.acm.org/10.1145/965105.807503

Breckinridge Church, Ruth, Saba Ayman-Nolley & Shahrzad Mahootian. 2004. The role of gesture in bilingual education: Does gesture enhance learning?

International Journal of Bilingual Education and Bilingualism 7(4). 303–319. DOI:10.1080/13670050408667815

Bruneau, Thomas J. 1980. Chronemics and the verbal-nonverbal interface. In Mary Ritchie Key (ed.), *The relationship of verbal and nonverbal communication*, 101–118. The Hague, The Netherlands: Mouton.

Bruner, Jerome S. 1998. Routes to reference. *Pragmatics & Cognition* 6(1/2). 209–227. Special Issue: The Concept of Reference in the Cognitive Sciences.

Bühler, Karl. 1934. *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Stuttgart: Gustav Fischer Verlag. Re-edition Stuttgart: UTB, Lucius & Lucius, 1999.

Burks, Arthur W. 1949. Icon, index, and symbol. *Philosophy and Phenomenological Research* 9(4). 673–689.

Butterworth, George & Shoji Itakura. 2000. How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology* 18(1). 25–50.

Carletta, Jean, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma & Pierre Wellner. 2006. The AMI Meeting Corpus: a pre-announcement. In Steve Renals & Samy Bengio (eds.), *Machine learning for multimodal interaction: second international workshop, mlmi 2005, edinburgh, uk, july 11-13, 2005, revised selected papers* (Lecture Notes in Computer Science 3869), 28–39. Berlin & Heidelberg: Springer. DOI:10.1007/11677482_3

Cassell, Justin, Matthew Stone & Hao Yan. 2000. Coordination and context-dependence in the generation of embodied conversation. In *Proceedings of the first international conference on natural language generation*, 171–178.

Caute, Anna, Tim Pring, Naomi Cocks, Madeline Cruice, Wendy Best & Jane Marshall. 2013. Enhancing communication through gesture and naming therapy. *Journal of Speech, Language, and Hearing Research* 56(1). 337–351. DOI:10.1044/1092-4388(2012/11-0232)

Cienki, Alan J. & Cornelia Müller (eds.). 2008. *Metaphor and gesture* (Gesture studies 3). Amsterdam: John Benjamins.

Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.

Cohen, Philip R., Michael Johnston, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen & Josh Clow. 1997. QuickSet: Multimodal interaction for distributed applications. In *Proceedings of the fifth ACM international conference on multimedia* (MULTIMEDIA '97), 31–40. Seattle, Washington, USA. DOI:10.1145/266180.266328

Cook, Susan Wagner & Susan Goldin-Meadow. 2006. The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development* 7(2). 211–232. DOI:10.1207/s15327647jcd0702_4

Cooper, Robin. 2019. *From perception to linguistic communication: Using a theory of types with records (TTR) to model linguistic action and content.* https://github.com/robincooper/ttl. MS Gothenburg University.

Cooper, Robin & Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. In Shalom Lappin & Chris Fox (eds.), *The handbook of contemporary semantic theory*, 2nd edn., chap. 12, 375–407. Oxford, UK: Wiley-Blackwell.

Cooperrider, Kensy. 2017. Foreground gesture, background gesture. *Gesture* 16(2). 176–202. DOI:10.1075/gest.16.2.02coo

Cooperrider, Kensy & Rafael Núñez. 2012. Nose-pointing: Notes on a facial gesture of Papua New Guinea. *Gesture* 12(2). 103–129. DOI:doi:10.1075/gest.12.2.01coo

Copestake, Ann. 2007. Semantic composition with (Robust) Minimal Recursion Semantics. In *Proceedings of the workshop on deep linguistic processing* (DeepLP'07), 73–80. Prague, Czech Republic.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An introduction. *Research on Language and Computation* 3(4). 281–332.

Cubelli, Roberto, Piera Trentini & Carmelo G. Montagna. 1991. Re-education of gestural communication in a case of chronic global aphasia and limb apraxia. *Cognitive Neuropsychology* 8(5). 369–380. DOI:10.1080/02643299108253378

De Kuthy, Kordula. 2020. Information structure. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 825–859. Berlin: Language Science Press. DOI:??

Domaneschi, Filippo, Marcello Passarelli & Carlo Chiorri. 2017. Facial expressions and speech acts: Experimental evidences on the role of the upper face as an illocutionary force indicating device in language comprehension. *Cognitive Processing* 18(3). 285–306. DOI:10.1007/s10339-017-0809-6

Ebert, Cornelia. 2014. *The non-at-issue contributions of gestures.* Workshop on Demonstration and Demonstratives, April 11–12 2014, Stuttgart.

Ebert, Cornelia, Stefan Evert & Katharina Wilmes. 2011. Focus marking via gestures. In Ingo Reich, Eva Horch & Dennis Pauly (eds.), *Proceedings of sinn & bedeutung 15*, 193–208. Saarbrücken, Germany: Saarland University Press.

Eco, Umberto. 1976. *A theory of semiotics.* Bloomington: Indiana University Press.

Ekman, Paul & Wallace V. Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1(1). 49–98.

Ekman, Paul & Wallace V. Friesen. 1978. *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Enfield, Nick J. 2001. Lip-pointing: A discussion of form and function with reference to data from Laos. *Gesture* 1(2). 185–212.

Engdahl, Elisabet & Enric Vallduví. 1996. Information packaging in HPSG. In Elisabet Engdahl & Enric Vallduví (eds.), *Edinburgh working papers in cognitive science* (Studies in HPSG 12), 1–31. Edinburgh: University of Edinburgh.

Fillmore, Charles. 1982. Frame Semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin Publishing Co.

Fricke, Ellen. 2007. *Origo, Geste und Raum* (Linguistik – Impulse & Tendenzen 24). Berlin: De Gruyter.

Fricke, Ellen. 2012. *Grammatik multimodal. Wie Wörter und Gesten zusammenwirken* (Linguistik – Impulse und Tendenzen 40). Berlin: De Gruyter.

Galantucci, Bruno & Simon Garrod. 2011. Experimental semiotics: A review. *Frontiers in Human Neuroscience* 5(11). DOI:10.3389/fnhum.2011.00011

Gerwing, Jennifer & Janet B. Bavelas. 2004. Linguistic influences on gesture's form. *Gesture* 4(2). 157–195.

Gibbon, Dafydd, Ulrike Gut, Benjamin Hell, Karin Looks, Alexandra Thies & Thorsten Trippel. 2003. A computational model of arm gestures in conversation. In ISCA Archive (ed.), *Proceedings of the 8th European conference on speech communication and technology* (EUROSPEECH 2003), 813–816. Geneva, Switzerland. http://www.isca-speech.org/archive/eurospeech_2003, accessed 2018-9-10.

Ginzburg, Jonathan. 2012. *The interactive stance: Meaning for conversation*. Oxford, UK: Oxford University Press.

Ginzburg, Jonathan, Ellen Breitholtz, Robin Cooper, Julian Hough & Ye Tian. 2015. Understanding laughter. In Thomas Brochhagen, Floris Roelofsen & Nadine Theiler (eds.), *Proceedings of the 20th Amsterdam Colloquium*, 137–146. Amsterdam, Netherlands. http://semanticsarchive.net/Archive/mVkOTk2N/AC2015-proceedings.pdf, accessed 2018-9-10.

Giorgolo, Gianluca. 2010. A formal semantics for iconic spatial gestures. In Maria Aloni, Harald Bastiaanse, Tikitu de Jager & Katrin Schulz (eds.), *Logic, language and meaning* (Lecture Notes in Computer Science 6042), 305–314. Berlin: Springer.

Giorgolo, Gianluca & Ash Asudeh. 2011. Multimodal communication in LFG: gestures and the correspondence architecture. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG 2011 conference*, 257–277. Stanford, CA: CSLI Publications. http://csli-publications.stanford.edu/LFG/16/, accessed 2018-9-30.

Goldin-Meadow, Susan, Howard Nusbaum, Spencer D. Kelly & Susan Wagner. 2001. Explaining math: Gesturing lightens the load. *Psychological Science* 12(6). 516–522.

Goodman, Nelson. 1976. *Languages of art. An approach to a theory of symbols.* 2nd edn. Idianapolis: Hackett Publishing Company, Inc.

Goodwin, Charles. 2003. Pointing as situated practice. In Sotaro Kita (ed.), *Pointing: Where language, culture, and cognition meet*, chap. 2, 217–241. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Hadjikhani, Nouchine, Rick Hoge, Josh Snyder & Beatrice de Gelder. 2008. Pointing with the eyes: The role of gaze in communicating danger. *Brain and Cognition* 68(1). 1–8. DOI:http://dx.doi.org/10.1016/j.bandc.2008.01.008

Hahn, Florian & Hannes Rieser. 2010. Explaining speech gesture alignment in MM dialogue using gesture typology. In Paweł Łupowski & Matthew Purver (eds.), *Proceedings of the 14th workshop on the semantics and pragmatics of dialogue: Aspects of semantics and pragmatics of dialogue* (SemDial 2010), 99–111. Poznan, Poland.

Hall, Edward T. 1968. Proxemics. *Current Anthropology* 9(2-3). 83–108.

Hough, Julian, Ye Tian, Laura de Ruiter, Simon Betz, Spyros Kousidis, David Schlangen & Jonathan Ginzburg. 2016. DUEL: a multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter. In *Proceedings of the 10th international conference on language resources and evaluation* (LREC 2016), 1784–1788. Portorož, Slovenia.

Jakob, Hanna, Daniela Bartmann, Georg Goldenberg, Wolfram Ziegler & Katharina Hogrefe. 2011. Zusammenhang von Spontansprachproduktion und Gesten bei Patienten mit Aphasie. *Aphasie und verwandte Gebiete* 30(3). 20–38.

Jannedy, Stefanie & Norma Mendoza-Denton. 2005. Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure (ISIS)* (3). 199–244. http://opus.kobv.de/ubp/volltexte/2006/877/, accessed 2018-9-30. URN: urn:nbn:de:kobv:517-opus-8774.

Johansson, Gunnar. 1973. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* 14(2). 201–211. DOI:10.3758/BF03212378

Johnston, Michael. 1998. Unification-based multimodal parsing. In *Proceedings of the 36th annual meeting of the Association for Computational Linguistics and 17th international conference on computational linguistics – Volume 1* (ACL '98), 624–630. Montreal, Quebec, Canada: Association for Computational Linguistics. DOI:10.3115/980845.980949

Johnston, Michael. 2009. Building multimodal applications with EMMA. In *Proceedings of the 2009 international conference on multimodal interfaces* (ICMI-MLMI '09), 47–54. Cambridge, Massachusetts, USA. DOI:10.1145/1647314.1647325

Johnston, Michael, Philip R. Cohen, David McGee, Sharon L. Oviatt, James A. Pittman & Ira Smith. 1997. Unification-based multimodal integration. In *Proceedings of the 35th annual meeting of the Association for Computational Linguistics and eighth conference of the european chapter of the association for computational linguistics*, 281–288. Madrid, Spain: Association for Computational Linguistics. DOI:10.3115/976909.979653

Kaplan, David. 1989. Demonstratives. In Joseph Almog, John Perry & Howard Wettstein (eds.). In collab. with Ingrid Deiwiks & Edward N. Zalta, *Themes from Kaplan*, 481–563. New York & Oxford: Oxford University Press.

Kauffman, Lynn E. 1971. Tacesics, the study of touch: A model for proxemic analysis. *Semiotica* 4(2). 149–161. DOI:10.1515/semi.1971.4.2.149

Kelly, Spencer D., Corinne Kravitz & Michael Hopkins. 2004. Neural correlates of bimodal speech and gesture comprehension. *Brain and Language* 89(1). 253–260. DOI:10.1016/S0093-934X(03)00335-3

Kendon, Adam. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26(1). 22–63. DOI:10.1016/0001-6918(67)90005-4

Kendon, Adam. 1972. Some relationships between body motion and speech. An analysis of an example. In Aron Wolfe Siegman & Benjamin Pope (eds.), *Studies in dyadic communication*, chap. 9, 177–210. Elmsford, NY: Pergamon Press.

Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key (ed.), *The relationship of verbal and nonverbal communication* (Contributions to the Sociology of Language 25), 207–227. The Hague: Mouton.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge, MA: Cambridge University Press.

Kendon, Adam & Laura Versante. 2003. Pointing by hand in "neapolitan". In Sotaro Kita (ed.), *Pointing: where language, culture, and cognition meet*, chap. 6, 109–137. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Kipp, Michael. 2014. ANVIL: A universal video research tool. In Jacques Durand, Ulrike Gut & Gjert Kristofferson (eds.), *Handbook of corpus phonology*, chap. 21, 420–436. Oxford, UK: Oxford University Press.

Kipp, Michael, Michael Neff & Irene Albrecht. 2007. An annotation scheme for conversational gestures: how to economically capture timing and form. *Journal on Language Resources and Evaluation - Special Issue on Multimodal Corpora* 41(3-4). 325–339.

Kita, Sotaro & Aslı Özyürek. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48(1). 16–32. DOI:10.1016/S0749-596X(02)00505-3

Kita, Sotaro, Ingeborg van Gijn & Harry van der Hulst. 1998. Movement phases in signs and co-speech gestures, and their transcription by human coders. In Ipke Wachsmuth & Martin Fröhlich (eds.), *Gesture and sign language in human-computer interaction*, 23–35. Berlin & Heidelberg: Springer.

Klein, Ewan. 2000. Prosodic constituency in HPSG. In Ronnie Cann, Claire Grover & Philip Miller (eds.), *Grammatical interfaces in HPSG* (Studies in Constraint-Based Lexicalism), chap. 10, 169–200. Stanford, CA: CSLI Publications.

Klein, Wolfgang. 1978. Wo ist hier? Präliminarien zu einer Untersuchung der lokalen Deixis. *Linguistische Berichte* 58. 18–40.

Koenig, Jean-Pierre & Frank Richter. 2020. Semantics. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 785–823. Berlin: Language Science Press. DOI:??

Kolarova, Zornitza. 2011. *Lexikon der bulgarischen Alltagsgesten.* Technische Universität Berlin dissertation.

Kong, Anthony Pak-Hin, Sam-Po Law & Gigi Wan-Chi Chak. 2017. A comparison of coverbal gesture use in oral discourse among speakers with fluent and non-fluent aphasia. *Journal of Speech, Language, and Hearing Research* 60(7). 2031–2046. DOI:10.1044/2017_JSLHR-L-16-0093

Kopp, Stefan, Paul Tepper & Justine Cassell. 2004. Towards integrated microplanning of language and iconic gesture for multimodal output. In *Proceedings of the 6th international conference on multimodal interfaces* (ICMI ’04), 97–104. State College, PA, USA: ACM. DOI:10.1145/1027933.1027952

Kranstedt, Alfred, Stefan Kopp & Ipke Wachsmuth. 2002. MURML: a multimodal utterance representation markup language for conversational agents. In *Pro-*

ceedings of the aamas02 workshop on Embodied Conversational Agents – let's specify and evaluate them. Bologna, Italy.

Kranstedt, Alfred, Andy Lücking, Thies Pfeiffer, Hannes Rieser & Ipke Wachsmuth. 2006a. Deictic object reference in task-oriented dialogue. In Gert Rickheit & Ipke Wachsmuth (eds.), *Situated communication*, 155–207. Berlin: Mouton de Gruyter.

Kranstedt, Alfred, Andy Lücking, Thies Pfeiffer, Hannes Rieser & Ipke Wachsmuth. 2006b. Deixis: How to determine demonstrated objects using a pointing cone. In Sylvie Gibet, Nicolas Courty & Jean-Francois Kamp (eds.), *Gesture in human-computer interaction and simulation: 6th International Gesture Workshop, gw 2005, Berder Island, France, May 18-20, 2005, revised selected papers* (Lecture Notes in Computer Science 3881), 300–311. Berlin: Springer.

Krauss, Robert M., Yihsiu Chen & Rebecca F. Gottesmann. 2000. Lexical gestures and lexical access: A process model. In David McNeill (ed.), *Language and gesture*, chap. 13, 261–283. Cambridge, UK: Cambridge University Press.

Kühnlein, Peter, Manja Nimke & Jens Stegmann. 2002. *Towards an HPSG-based formalism for the integration of speech and co-verbal pointing*. Talk presented at The first congress of the ISGS, *Gesture – The Living Medium*, University of Texas at Austin.

Kupffer, Manfred. 2014. Does context change? In Daniel Gutzmann, C'ecile Meier & Jan Köpping (eds.), *Approaches to meaning. Composition, value, and interpretation* (Current Research in the Semantics/Pragmatics Interface 1), 25–44. Leiden, NL: Brill.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press.

Lascarides, Alex & Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics* 26(4). 393–449. DOI:10.1093/jos/ffp004

Lausberg, Hedda & Han Sloetjes. 2009. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods* 41(3). 841–849.

Le, Quoc Anh, Souheil Hanoune & Catherine Pelachaud. 2011. Design and implementation of an expressive gesture model for a humanoid robot. In *Proceedings of the 11th international conference on humanoid robots* (IEEE-RAS 2011), 134–140. DOI:10.1109/Humanoids.2011.6100857

Levelt, Willem J. M. 1989. *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Levinson, Stephen C. 2008. Deixis. In *The handbook of pragmatics*, chap. 5, 97–121. Blackwell.

Lewis, David. 1970. General semantics. *Synthese*. Semantics of Natural Language II 22(1/2). 18–67.

Loehr, Daniel. 2004. *Gesture and intonation*. Washington, D.C.: Georgetown University dissertation.

Loehr, Daniel. 2007. Apects of rhythm in gesture in speech. *Gesture* 7(2). 179–214.

Lücking, Andy. 2013. *Ikonische Gesten. Grundzüge einer linguistischen Theorie*. Berlin: De Gruyter. Zugl. Diss. Univ. Bielefeld (2011).

Lücking, Andy. 2016. Modeling co-verbal gesture perception in type theory with records. In Maria Ganzha, Leszek Maciaszek & Marcin Paprzycki (eds.), *Proceedings of the 2016 federated conference on computer science and information systems* (Annals of Computer Science and Information Systems 8), 383–392. IEEE. DOI:10.15439/2016F83

Lücking, Andy. 2018. Witness-loaded and witness-free demonstratives. In Marco Coniglio, Andrew Murphy, Eva Schlachter & Tonjes Veenstra (eds.), *Atypical demonstratives. Syntax, semantics and pragmatics* (Linguistische Arbeiten 568). Berlin: De Gruyter.

Lücking, Andy, Kirsten Bergman, Florian Hahn, Stefan Kopp & Hannes Rieser. 2013. Data-based analysis of speech and gesture: the Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications. *Journal on Multimodal User Interfaces* 7(1-2). 5–18. DOI:10.1007/s12193-012-0106-8

Lücking, Andy, Kirsten Bergmann, Florian Hahn, Stefan Kopp & Hannes Rieser. 2010. The Bielefeld speech and gesture alignment corpus (SaGA). In *Multimodal corpora: Advances in capturing, coding and analyzing multimodality* (LREC 2010), 92–98. Malta. DOI:10.13140/2.1.4216.1922

Lücking, Andy, Jonathan Ginzburg & Robin Cooper. 2020. Grammar in dialogue. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 963–1007. Berlin: Language Science Press. DOI:??

Lücking, Andy & Alexander Mehler. 2012. What's the scope of the Naming Game? Constraints on semantic categorization. In Thomas C. Scott-Phillips, Mónica Tamariz, Erica A. Cartmill & James R. Hurford (eds.), *Proceedings of the 9th international conference on the evolution of language* (Evolang IX), 196–203. Kyoto, Japan.

Lücking, Andy, Alexander Mehler & Peter Menke. 2008. Taking fingerprints of speech-and-gesture ensembles: Approaching empirical evidence of intrapersonal alignmnent in multimodal communication. In *Proceedings of the 12th workshop on the semantics and pragmatics of dialogue* (LonDial'08), 157–164. King's College London.

Lücking, Andy, Alexander Mehler, Désirée Walther, Marcel Mauri & Dennis Kurfürst. 2016. Finding recurrent features of image schema gestures: the FIGURE corpus. In *Proceedings of the 10th international conference on language resources and evaluation* (LREC 2016), 1426–1431. Portorož (Slovenia).

Lücking, Andy & Thies Pfeiffer. 2012. Framing multimodal technical communication. In Alexander Mehler & Laurent Romary (eds.). In collab. with Dafydd Gibbon, *Handbook of technical communication* (Handbooks of Applied Linguistics 8), chap. 18, 591–644. Berlin: De Gruyter Mouton.

Lücking, Andy, Thies Pfeiffer & Hannes Rieser. 2015. Pointing and reference reconsidered. *Journal of Pragmatics* 77. 56–79. DOI:10.1016/j.pragma.2014.12.013

Lücking, Andy, Hannes Rieser & Jens Stegmann. 2004. Statistical support for the study of structures in multi-modal dialogue: Inter-rater agreement and synchronization. In Jonathan Ginzburg & Enric Vallduví (eds.), *Proceedings of the eighth workshop on the semantics and pragmatics of dialogue* (Catalog '04), 56–63. Barcelona.

Martell, Craig, Chris Osborn, Jesse Friedman & Paul Howard. 2002. FORM: a kinematic annotation scheme and tool for gesture annotation. In *Proceedings of multimodal resources and multimodal systems evaluation*, 15–22. Las Palmas, Spain.

Masataka, Nobuo. 2003. From index-finger extension to index-finger pointing: Ontogenesis of pointing in preverbal infants. In Sotaro Kita (ed.), *Pointing: Where language, culture, and cognition meet*, chap. 4, 69–84. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Matthews, Danielle, Tanya Behne, Elena Lieven & Michael Tomasello. 2012. Origins of the human pointing gesture: A training study. *Developmental Science* 15(6). 817–829. DOI:10.1111/j.1467-7687.2012.01181.x

McClave, Evelyn. 1994. Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research* 23(1). 45–66.

McGinn, Colin. 1981. The mechanism of reference. *Synthese* 49(2). 157–186.

McNeill, David. 1985. So you think gestures are nonverbal? *Psychological Review* 92(3). 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.

McNeill, David & Susan D. Duncan. 2000. Growth points in thinking-for-speaking. In David McNeill (ed.), *Language and gesture*, chap. 7, 141–161. Cambridge, MA: Cambridge University Press.

Mehler, Alexander & Andy Lücking. 2012. Pathways of alignment between gesture and speech: Assessing information transmission in multimodal ensembles. In Gianluca Giorgolo & Katya Alahverdzhieva (eds.), *Proceedings of the international workshop on formal and computational approaches to multimodal communication under the auspices of ESSLLI 2012, Opole, Poland, 6–10 August.*

Müller, Cornelia. 1998. *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich* (Körper – Kultur – Kommunikation 1). Berlin: Berlin Verlag. Zugl. Diss. Frei Universität Berlin (1996).

Müller, Cornelia, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeill & Sedinha Tessendorf (eds.). 2013–2014. *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction.* 2 vols. (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1 & 2). Berlin: De Gruyter Mouton.

Nobe, Shuichi. 2000. Where do *most* spontaneous representational gestures actually occur with respect to speech? In David McNeill (ed.), *Language and gesture*, 186–198. Cambridge, MA: Cambridge University Press.

Nöth, Winfried. 1990. *Handbook of semiotics.* Bloomington, Indianapolis: Indiana University Press.

Nunberg, Geoffrey. 1993. Indexicality and deixis. *Linguistics and Philosophy* 16(1). 1–43. DOI:10.1007/BF00984721

Özyürek, Aslı. 2012. Gesture. In Roland Pfau, Markus Steinbach & Bencie Woll (eds.), *Sign language: An international handbook*, vol. 37 (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK)), chap. 27, 626–646. Berlin: De Gruyter Mouton.

Peacocke, Christopher. 1987. Depiction. *The Philosophical Review* 96(3). 383–410.

Pfeiffer, Thies, Carolin Hainke, Leonard Meyer, Maik Fruhner & Moritz Niebling. 2018. Proceedings der pre-conference-workshops der 16. e-learning fachtagung informatik co-located with 16th e-learning conference of the german computer society. In Daniel Schiffner (ed.) (DeLFI 2018 2250). Frankfurt, Germany: CEUR Workshop Proceedings.

Posner, Roland. 2002. Alltagsgesten als Ergebnis von Ritualisierung. In Jan C. Joerden (ed.), *Ritualisierte Tabuverletzung, Lachkultur und das Karnevaleske*, vol. 6 (Studien zur Ethik in Ostmitteleuropa), 395–421. Frankfurt am Main: Peter Lang.

Posner, Roland, Klaus Robering & Thomas A. Sebeok (eds.). 1997–2004. *Semiotik. Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur / Semiotics. A Handbook on the Sign-Theoretic Foundations of Nature and Culture.*

4 vols. (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 13/1–4). Berlin: De Gruyter.

Poyatos, Fernando. 1975. Cross-cultural study of paralinguistic "alternants" in face-to-face interaction. In Adam Kendon, Harris M. Richards & Mary Ritchie Key (eds.), *Organization of behavior in face-to-face interaction*, 285–314. The Hague: Mouton.

Quine, Willard Van Orman. 1950. Identity, ostension, and hypostasis. *The Journal of Philosophy* 47(22). 621–633.

Rieser, Hannes. 2004. Pointing in dialogue. In *Proceedings of the eighth workshop on the semantics and pragmatics of dialogue* (Catalog '04), 93–100. Barcelona.

Rieser, Hannes. 2008. Aligned iconic gesture in different strata of mm route-description. In *LonDial 2008: The 12th workshop on the semantics and pragmatics of dialogue* (SemDial), 167–174. King's College London.

Rieser, Hannes. 2010. On factoring out a gesture typology from the Bielefeld Speech-Gesture-Alignment Corpus. In Stefan Kopp & Ipke Wachsmuth (eds.), *Proceedings of GW 2009: Gesture in embodied communication and human-computer interaction*, 47–60. Berlin & Heidelberg: Springer.

Rieser, Hannes. 2011. How to disagree on a church window's shape using gesture. In Klaus Hölker & Carla Marello (eds.), *Dimensionen der Analyse von Texten und Diskursen. Festschrift für János Sándor Petöfi*, 231–246. Berlin: LIT Verlag.

Rieser, Hannes. 2015. When hands talk to mouth. Gesture and speech as autonomous communicating processes. In Christine Howes & Staffan Larsson (eds.), *Proceedings of the 19th workshop on the semantics and pragmatics of dialogue* (SemDial 2015: goDIAL), 122–130. Gothenburg, Sweden.

Rieser, Hannes & Massimo Poesio. 2009. Interactive gestures in dialogue: A PTT model. In *Proceedings of the 10th annual meeting of the special interest group in discourse and dialogue* (SIGDIAL 2009), 87–96. Queen Mary University of London.

Röpke, Insa. 2011. Watching the growth point grow. In *Proceedings of the second conference on gesture and speech in interaction* (GESPIN 2011). http://coral2.spectrum.uni-bielefeld.de/gespin2011/final/Roepke.pdf.

Rose, Miranda L. 2006. The utility of arm and hand gestures in the treatment of aphasia. *International Journal of Speech-Language Pathology* 8(2). 92–109. DOI:10.1080/14417040600657948

de Ruiter, Jan P. & Carola de Beer. 2013. A critical evaluation of models of gesture and speech production for understanding gesture in aphasia. *Aphasiology* 27(9). 1015–1030. DOI:10.1080/02687038.2013.797067 de Ruiter, Jan Peter. 2000. The production of gesture and speech. In David McNeill (ed.), *Language and gesture*, chap. 14, 284–311. Cambridge, UK: Cambridge University Press.

de Ruiter, Jan Peter. 2004. On the primacy of language in multimodal communication. In Jean-Claude Martin, Elisabeth Den, Peter Kühnlein, Lou Boves, Patrizia Paggio & Roberta Catizone (eds.), *Proceedings of the fourth international conference on language resources and evaluation and the workshop on multimodal corpora: Models of human behaviour for the specification and evaluation of multimodal input and output interfaces* (LREC 2004), 38–41. Lisbon, Portugal.

de Ruiter, Jan Peter. 2007. Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture* 7(1). 21–38.

Salem, Maha, Stefan Kopp, Ipke Wachsmuth, Katharina Rohlfing & Frank Joublin. 2012. Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics* 4(2). 201–217. DOI:10.1007/s12369-011-0124-9

Schiel, Florian, Silke Steininger & Uli Türk. 2003. *The SmartKom multimodal corpus at BAS*. Tech. rep. 34. München: Ludwig-Maximilians-Universität München.

Schlenker, Philippe. 2018. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41(3). 295–365. DOI:10.1007/s10988-017-9225-8

Schmidt, Thomas. 2012. EXMARaLDA and the FOLK tools. In *Proceedings of lrec*. ELRA. http://www.lrec-conf.org/proceedings/lrec2012/pdf/529_Paper.pdf.

Sekine, Kazuki & Miranda L. Rose. 2013. The relationship of aphasia type and gesture production in people with aphasia. *American Journal of Speech-Language Pathology* 22(4). 662–672. DOI:10.1044/1058-0360(2013/12-0030)

Slama-Cazacu, Tatiana. 1976. Nonverbal components in message sequence: “Mixed Syntax”. In William C. McCormick & Stephan A. Wurm (eds.), *Language and man. Anthropological issues* (World Anthropology), 217–227. The Hague & Paris: Mouton.

Sloetjes, Han & Peter Wittenburg. 2008. Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th international conference on language resources and evaluation* (LREC 2008).

van der Sluis, Ielka & Emiel Krahmer. 2007. Generating multimodal references. *Discourse Processes* 44(3). 145–174. Special Issue: Dialogue Modelling: Computational and Empirical Approaches. DOI:10.1080/01638530701600755

Sonesson, Göran. 1998. That there are many kinds of iconic signs. *Visio* 1(1). 33–54.

Sowa, Timo. 2006. *Understanding coverbal iconic gestures in shape descriptions*. Berlin: Akademische Verlagsgesellschaft. Zugl. Diss. Univ. Bielefeld.

Steels, Luc. 1995. A self-organizing spatial vocabulary. *Artificial Life* 2(3). 319–332.

Steinbach, Markus & Anke Holler. 2020. Sign languages. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 1009–1010. Berlin: Language Science Press. DOI:??

Streeck, Jürgen. 2008. Depicting by gesture. *Gesture* 8(3). 285–301. DOI:10.1075/gest.8.3.02str

Triesch, Jochen & Christoph von der Malsburg. 1998. Robotic gesture recognition. In Ipke Wachsmuth & Martin Fröhlich (eds.), *Gesture and sign language in human-computer interaction*, 233–244. Berlin & Heidelberg: Springer.

Tuite, Kevin. 1993. The production of gesture. *Semiotica* 93(1/2). 83–105.

Wagner, Petra, Zofia Malisz & Stefan Kopp. 2014. Gesture and speech in interaction: An overview. *Speech Communication* 57. 209–232. DOI:10.1016/j.specom.2013.09.008

Wahlster, Wolfgang (ed.). 2006. *SmartKom: foundations of multimodal dialogue systems*. Berlin & Heidelberg: Springer.

Weissmann, John & Ralf Salomon. 1999. Gesture recognition for virtual reality applications using data gloves and neural networks. In *Proceedings of the international joint conference on neural networks* (IJCNN'99. 3), 2043–2046. DOI:10.1109/IJCNN.1999.832699

Wilkins, David. 2003. Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In Sotaro Kita (ed.), *Pointing: Where language, culture, and cognition meet*, chap. 8, 171–215. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Wiltshire, Anne. 2007. Synchrony as the underlying structure of gesture: The relationship between speech sound and body movement at the micro level. In Robyn Loughnane, Cara Penry Williams & Jana Verhoeven (eds.), *In between wor(l)ds: Transformation and translation*, vol. 6 (Postgraduate Research Papers on Language and Literature), 235–253. Melbourne: School of Languages & Linguistics, The University of Melbourne.

Wu, Ying Choon & Seanna Coulson. 2005. Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 42(6). 654–667.

Zwarts, Joost. 2003. Vectors across spatial domains: from place to size, orientation, shape, and parts. In *Representing direction in language and space* (Explorations in Language and Space 1), chap. 3, 39–68. Oxford, NY: Oxford University Press.

Zwarts, Joost & Yoad Winter. 2000. Vector space semantics: a model-theoretic analysis of locative prepositions. *Journal of Logic, Language, and Information* 9(2). 169–211.

Żywiczyński, Przemysław, Sławomir Wacewicz & Sylwester Orzechowski. 2017. Adaptors and the turn-taking mechanism. The distribution of adaptors relative to turn borders in dyadic conversation. *Interaction Studies* 18(2). 276–298. DOI:10.1075/is.18.2.07zyw

Part V

The broader picture