Role of Gesture in Language Processing: Toward a unified account for production and comprehension

Asli Özyürek

The Oxford Handbook of Psycholinguistics (2 ed.)
Edited by Shirley-Ann Rueschemeyer and M. Gareth Gaskell

Print Publication Date: Aug 2018 Subject: Psychology, Cognitive Psychology
Online Publication Date: Sep 2018 DOI: 10.1093/oxfordhb/9780198786825.013.25

Abstract and Keywords

Use of language in face-to-face context is multimodal. Production and perception of speech take place in the context of visual articulators such as lips, face, or hand gestures which convey relevant information to what is expressed in speech at different levels of language. While lips convey information at the phonological level, gestures contribute to semantic, pragmatic, and syntactic information, as well as to discourse cohesion. This chapter overviews recent findings showing that speech and gesture (e.g. a drinking gesture as someone says, “Would you like a drink?”) interact during production and comprehension of language at the behavioral, cognitive, and neural levels. Implications of these findings for current psycholinguistic theories and how they can be expanded to consider the multimodal context of language processing are discussed.

Keywords: co-speech gesture, pointing, iconic gesture, imagery, integration, action, simulation, motor, communicative intent

25.1 Introduction

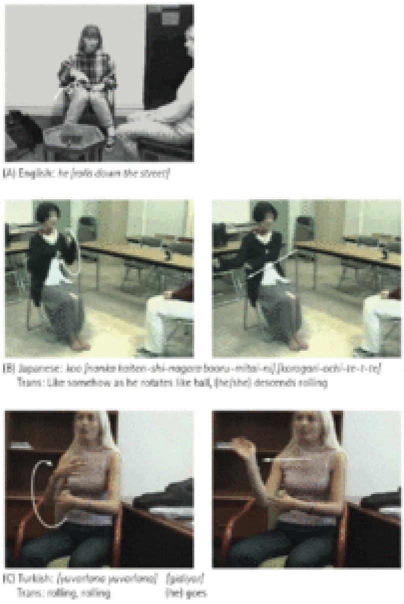
USE of language in the face-to-face context involves production and perception of speech using many visual articulators, such as the lips, face, or hand gestures. These visual articulators convey relevant information to what is expressed in speech and at different levels. For example, while lips convey information at the phonological level, hand gestures (and face and head movements to some extent) contribute to semantic, pragmatic, and even syntactic information (e.g., Bavelas et al., 2000; Floyd, 2016; Kendon, 2004; Kita et al.,2007; Krahmer & Swerts, 2007; McNeill, 1992). Gestures can have different forms and functions in communication such as to pick out (e.g., points) or depict concrete or absent referents, action, and motion (e.g., iconic gestures), highlight meaning in the speech channel (e.g., beats), or to coordinate communicative interactions during dialogue such as signaling turn-taking, agreements, and so on (e.g., interactive, pragmatic gestures). Although gestures reveal the information in a different representational format than speech due to radical differences in the visual and auditory modalities, the two are systematically related to each other and convey the speaker's meaning together as a "composite signal" (Clark, 1996; Enfield, 2009). This chapter focuses on the role of hand gestures, and more specifically of iconic gestures, and to some extent of pointing gestures, in the processing of language during production and comprehension based on spontaneous and elicited productions as well as experimental and neural data. The accumulated findings in this domain show that gestures interact with language, both during production and comprehension, and that any model of language processing that tries to account for contextual uses of language needs to consider the role of gesture. Similarities in interactions between speech and gesture during both production and comprehension should be considered in extending the psycholinguistic models of language.

(p. 593) 25.2 The role of gesture in language production

In considering the role of gesture in language production, it is essential to consider the semantic and temporal relatedness between speech and gesture during spontaneous productions. First, in most cases there is semantic overlap between the representation in gesture and the meaning expressed in concurrent speech. However, gesture usually also encodes additional information that is not expressed in speech due to the affordances of the modality. Consider the example of someone giving directions. He might say in his speech "you walk across the street" accompanied by an iconic hand gesture consisting of the hand moving from left to right while the fingers wiggle repetitively. In this example, a single gesture exhibits simultaneously the manner, the change of location, and the direction of the movement to the right. Speech expresses the manner and the path (walk and across) of the movement, but not the change in direction (left to right). Thus, there is informational overlap between speech and gesture, but also additional/non-overlapping information in the gesture (Holler & Beattie, 2003; Kita & Özyürek, 2003; Özyürek et al., 2005). Secondly, there is systematic temporal relationship between speech and gesture. A gesture phrase has three phases: preparation, stroke (semantically the most meaningful part of the gesture), and retraction or hold. All three phases together constitute a gesture phrase. McNeill (1992) has also shown that in 90% of speech-gesture pairs, the stroke coincides with the relevant speech segment, which might be a single lexical item or a phrase. For example, the stroke phase of the gesture in the aforementioned example is very likely to occur during the phrase “walk across” or “walk across the street” (see examples of spontaneous gestures in Fig. 25.1 from different languages).

25.2.1 Different models of speech and gesture production

Even though speech and gesture seem to be tightly coordinated to achieve semantic and temporal congruity for communicative effectiveness, there is controversy in the literature regarding gesture’s underlying origin of representational format, especially of iconic gestures and their interactions with speech during the language production process (see de Ruiter, 2007; Wagner et al., 2014, for a review of different models). Iconic gestures depicting visual, imagistic aspects of actions and referents convey perceptual, motoric, and analogic mappings between gestures and the conceptual content they evoke. While some view and explain the production of such gestures as being generated and executed independent of the spoken linguistic utterances they accompany, others see their processing as intrinsically interwoven with production of spoken language. Another controversy exists regarding the communicative nature of gesture production; that is, whether gestures are produced for the speakers themselves or are designed for the informational needs of the addressees (i.e., with communicative intent or not). These two issues have been crucial for designing different models of gesture production and its relation to processing of spoken language production. These models will be briefly reviewed next. (p. 594)



Click to view larger

Fig. 25.1 Differences between English, Japanese, and Turkish speakers' gestures depicting simultaneous manner and path of a motion event paralleling differences in syntactic encoding (brackets indicate where the stroke of the gestures overlap with the speech segment).

According to some views (Krauss et al., 1995; Wesp et al., 2001) speech and iconic gestures originate and are processed independently and are executed in a parallel fashion (i.e., to explain their overt coordination at the behavior level). According to these views iconic gestures are generated and processed directly and solely from the spatial, motoric action representations, whereas speech is generated from abstract propositional representations. This model also assumes that gestures are not communicatively intended, and thus not a necessary part of the speaker's intended message expressed in the spoken utterance. Gestures are (p. 595) generated from spatial representations, "prelinguistically," and independent from how certain information is linguistically formulated. The function of gestures is to keep memories of such representations active and/or facilitate lexical retrieval through cross-modal priming (i.e., from gesture to speech).

Also according to another framework, Gesture as Simulated Action (GSA) (Beilock & Goldin-Meadow, 2010; Cook & Tannenhous, 2009; Hostetter & Alibali, 2008), and a recent Action Generation Hypothesis (Chu & Kita, 2016) gestures *can* arise directly out of simulations of actions (action representations) without requiring explicit interactions between speech and gesture[1]. As such, gestures can be seen as a direct window into "simulated" cognition of speakers. For example, in Cook and Tannenhous (2009), participants were asked to solve a tower of Hanoi problem either by moving real objects with their hands or by moving objects on a computer screen with a mouse. They then described their solutions to a listener who would be solving the same problems later. Participants who solved the problem with real objects produced more gestures with grasping hand shapes and more gestures with higher and more curved trajectories than those who solved the computerized version of the problem. In a recent study, Chu and Kita (2016) have asked participants to imagine mentally rotating "smooth" faced or "spiky" mugs, and to think aloud as they did so in a non-communicative setting. They have found participants to gesture less in the "spiky" mug condition than in the "smooth" mug condition. According to authors these findings reflect speakers' action representations about how they would likely manually interact with the mugs. Finally, a recent study shows gestures being sensitive to “affordances” of objects mentioned in speech (Masson-Carro et al., 2016). Note that in these models no explicit interactions between speech and gesture are *necessary*, but possible (see footnote 1). Finally, the GSA model gestures—even though not inherently communicatively intended—can be suppressed if context does not require it. However, communicative intent does not shape the form or the choice of gesturing or not.

Other models (e.g., Interface Hypothesis, Kita & Özyürek, 2003) on the other hand propose more close interactions between imagistic/action representations that give rise to gestural representations and linguistic conceptualization during the generation and execution of coordinated speech and gesture units (e.g., clause). In this framework, these interactions are almost inevitable due to the notion that gestures function as a communicative device as does language, and that there is close semantic and temporal coordination between the two. As such, gestures are generated from the same communicative intention used during conceptualization of speech production (de Ruiter, 2007; Melinger & Levelt, 2004; Peeters et al., 2015, shown for pointing gesture) and go hand-in-hand taking addressee’s knowledge state into account (i.e., common ground such as shared knowledge between interlocutors, visibility of the gestures, or the shared space among the interlocutors) (Alibali et al., 2001; Campisi & Özyürek, 2013; Schubotz et al., 2015; Özyürek, 2002). A recent study manipulating production of pointing gestures accompanying demonstrative speech (e.g., *this, that*) as a function of addressee’s knowledge state has also identified possible neural correlates of (p. 596) communicative intent in the brain during the planning of pointing gestures (Peeters et al., 2015). Finally, when one considers the larger discourse context, it is also shown that gesture production is influenced by the accessibility of the referent (i.e., old vs. new) in discourse (e.g., Debreslioska et al., 2013; Perniss & Özyürek, 2015; So et al., 2009), a finding that generalizes across different types of languages. Speakers are more likely to gesture with new or pragmatically marked referents in discourse context paralleling such discourse markers in speech, within and across languages (Azar et al., 2017).

In addition to postulating that gestures are as communicatively intended as speech, the Interface Hypothesis also proposes that speech and gesture processing interact during production. The evidence for this comes from studies showing that iconic gestures of the same event (i.e., similar imagery) differ according the language-specific semantic and grammatical encoding of spatial information in different languages. The independence models mentioned here would predict that the way certain elements of an event are encoded linguistically will *not* change the form of gestures, since gestures are generated from and are shaped *solely* by spatial, imagistic, motoric, action representations (i.e., which would be similar across speakers of different languages with different encoding possibilities). However according to interaction models (i.e., specifically the Interface Model; see Kita & Özyürek, 2003), the linguistic encoding of the event would change the shape of gestures, due to an interaction between linguistically formulating the message (i.e., specific semantic, linguistic, and discourse for requirements of each language) and the spatiomotoric imagery that underlies formation of gesture during the conceptualization phase of the online language production. That is, the spatiomotoric imagery would be influenced by the linguistic conceptualization that is specific for each language, giving rise to differences in gestures (mostly for iconic gestures) for the same event.

One domain where there are particular differences between spoken languages and their corresponding gestures is in the realm of expressions of motion events. Talmy (1985) has proposed a typology in the expression of motion event across the world's languages, based on how path of motion is expressed syntactically: in *satellite-framed languages* (S-languages, English, German, and so on), manner of motion is typically expressed in the verb, while path of motion appears in a particle outside the verb (e.g., "The boy ran down the stairs"). Whereas in *verb-framed languages* (V-languages; Turkish, Spanish, and so on), the main verb usually encodes the path of motion, while manner information is encoded with gerunds (e.g., Spanish), adverbs, or subordinate clauses (e.g., Turkish) outside the verb. In the next example, "*in*" (descend) is the main verb encoding path but manner, "*koş" (*run) is expressed in the subordinate clause. Because of these differences, speakers of V-framed languages express mostly path of motion but omit the manner in speech.

[1] Turkish:

çocuk koş-arak merdiven-den in-di

child run-CONN stairs-ABL descend-PAST

"The boy descended the stairs while running"

Researchers examining speakers' gesture production across a variety of languages show that the content and type of the iconic gestures covary with the aforementioned preferences made in different languages. For example in V-framed languages, adult and child speakers prefer to express only the path of motion both in their speech and gesture; for example, French (Gullberg et al., 2008), and Turkish (Özyürek et al., 2008, 2014).The congruency between (p. 597) speech and gesture patterns is also found in another study where Turkish and English speakers were asked to talk about 10 different motion events that involved different types of manner (jump, roll, spin, rotate) and path (descend, ascend, go around). In cases where only manner or only path was expressed in an utterance in either language, speakers of both languages were more likely to express congruent information in gesture to what is expressed with speech (e.g., he went down the slope: Gesture: index finger moving down expressing just the path information).

Also in line with the view that what can *not* be habitually expressed in speech is also omitted in gesture comes from a study (Kita & Özyürek, 2003) that compares how Japanese, Turkish, and English speakers speak and gesticulate about an event, and where languages differ in the lexical items available to encode a certain part of the event. In this case, speakers of all three languages were shown a Sylvester and Tweety cartoon. In one scene, Sylvester grabs a rope and tries to swing from one building to another to catch

Tweety. It was found that English speakers all used the verb *swing across*, and encoded the arc shape of Sylvester's trajectory. On the other hand, Japanese and Turkish speakers used verbs such as *go across*, which does not encode the arc trajectory. In their conceptual planning phase of the utterance describing this event, Japanese and Turkish speakers presumably got feedback from speech formulation processes and created a mental representation of the event that does not include the trajectory shape. If gestures reflect this planning process, the gestural contents should differ cross-linguistically in a way analogous to the difference in speech. It was indeed found that Japanese and Turkish speakers were more likely to produce a straight gesture, which does not encode the trajectory shape, and most English speakers produced just gestures with an arc trajectory.

More evidence demonstrating that what is represented in iconic gestures seems to vary according to verb semantics of the specific language comes from a study comparing French and Dutch speakers' speech and gesture patterns. Placement events are encoded using the simple verb *mettre* "put" in French. In contrast, speakers of Dutch encode these events by using positional verbs such as *leggen* "lay" and *zetten "*set/stand," depending on the shape of the object that is placed. Paralleling these distinctions, adult French speakers have been found to use iconic gestures that encode only the path or direction of movement in their placement descriptions, whereas Dutch speakers' gestures represent the shape of the moved object (i.e., via the hand shape as if holding the object), as well as the direction of movement (Gullberg, 2011; Gullberg & Narasimhan, 2010). Note that these results speak against the idea that action representation system *alone* cannot be the origin of iconic gestures, as assumed by the GSA model of gesture production. Otherwise French speakers would also be expected to gesture, representing the shape of the objects as per Dutch speakers.

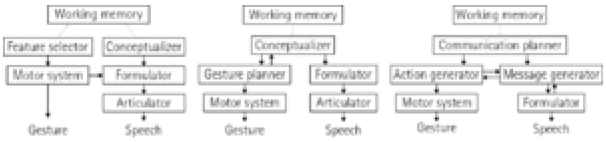
Finally, another way linguistic encoding can shape gestural representation has been found in expressions of events that include both manner and path. Here the influence is found not at the lexical level but more at the level of how information is *syntactically* packaged. As mentioned here, verbal descriptions differ cross-linguistically in terms of how manner and path information is lexicalized. English speakers used a manner verb and a path particle or preposition to express the two pieces information within one clause (e.g., he *rolled down* the hill). In contrast, Japanese and Turkish speakers separate manner and path expressions over two clauses; path as in the main clause and manner as in the subordinated clause (e.g., he descended as he rolled). Given the assumption that a clause approximates a unit of processing in speech production (Levelt, 1989) presumably English speakers were (p. 598) likely to process both manner and path within a single processing unit, whereas Japanese and Turkish speakers were likely to need two processing units. Consequently, Japanese and Turkish speakers should be more likely to separate the imagistic representations of manner and path in preparation for speaking so that two pieces of information could be dealt with in turn, unlike as in English speakers. The gesture data confirmed this prediction. In depicting how an animated figure rolled down a hill having swallowed a bowling ball in the cartoon, Japanese and Turkish speakers were more likely to use separate gestures, one for manner and one for path, and

English speakers were more likely to use just one gesture to express both manner and path (see Fig. 25.1). Note that as reported in Kita and Özyürek (2003), these patterns are tendencies but not absolute parallels between gesture and speech and might be modulated with what is salient and prominent in the real event or discourse context.

Based on these findings one could argue, however, that gestural variation across speakers of different languages is not due to online interaction between linguistic and imagistic thinking, but rather either due to cultural patterns of gestures learned from others independent of linguist encoding or deep effects of language, which then determine gesture production directly. Two recent studies rule out these possibilities. First of all, Özçalışkan et al. (2016a) have replicated findings from Kita and Özyürek (2003) with blind speakers of English and Turkish –showing that differences in gesture patterns are not learned by seeing others but are influenced by the specific language used. Secondly in another study Özçalışkan et al. (2016b) asked speakers of Turkish and English first to talk about events containing simultaneous manner and path (e.g., an animated figure hopping into a house) and later to depict them with gesture, only without speaking. While in the speech condition gestures have differed as predicted by Kita and Özyürek, in the silent condition, both groups were similar and used conflated gestures. Different patterning of gestures in silent conditions than in accompanying speech context has also been found at the level of ordering of semantic elements such as agent, patient, and action within English (Goldin-Meadow, McNeill, & Singleton, 1996) and also across speakers of different languages that use different word orders (Goldin-Meadow, So, Özyürek, & Mylander, 2008). These experimental findings then argue against the claims that differences in co-speech gestures across different languages arise simply from culturally or prelinguistically shaped conceptualization of events, but rather they point to online influence of linguistic conceptualization on gesture processing.

25.3 Summary: Role of gesture in language production

Gestures that speakers use during multimodal utterances serve multiple functions (i.e., cognitive, communicative) and are shaped by multiple representations (imagery, action simulation, abstract propositional) during production. While much research has emphasized how action, motoric, and spatial representations shape gesture form and content directly and independent of the linguistic processing, there is considerable evidence showing that gesture is shaped also by the speaker's language system and by the communicative needs of the addressee (e.g., to emphasize new and pragmatically marked information, knowledge state, visibility, location of the addressee) and the discourse context. After all, gestures are communicative acts, produced with communicative functions (as language) and are produced to fit semantically and temporally to verbal utterances. (p. 599)



Click to view larger

Fig. 25.2 Adapted schematic overview of different models in relation to speech and gesture production (Morrel-Samuels & Krauss, 1992; de Ruiter, 2000; Kita & Özyürek, 2003; from left to right).

Reprinted from *Speech Communication*, 57, Petra Wagner, Zofia Malisz, and Stefan Kopp, Gesture and speech in interaction: An overview, pp. 209–32, doi.org/10.1016/j.specom.2013.09.008, Copyright © 2013 Elsevier B.V., with permission from Elsevier.

Next are some sketches of different models proposed for speech and gesture production, as taken from Wagner, Malisz, and Kopp (2014)) schematizing their link to the language system and to the communicative intent in message formulation (Fig. 25.2).

25.4 The role of gesture in language comprehension

Even though the abovementioned production models of gesture have been split with regard to consideration of gestures as part of speakers' communicative intent and whether they are linked to the language production system, when it comes to comprehension a more unified view emerges. Growing research shows that gestures are interpreted as communicative. As part of a speaker's message, speech and gesture processing mutually influence each other, recruiting similar semantic processing areas in the brain. Furthermore, the perceived communicative intent of the gesture also influences its integration with speech, also at the neural level.

It has been a longstanding finding that addressees pick up semantic information from gestures that accompany speech. That is, gestures are not perceived by comprehenders simply as handwaving or as attracting attention to what is conveyed in speech. For example, Kelly et al. (1999) showed participants video stimuli where gestures conveyed additional information to that conveyed in speech (gesture pantomiming drinking while speech is "I stayed up all night") and asked them to write what they heard. In addition to the speech they heard, participants' written text contained information that was conveyed only in gesture but not in speech (i.e., "I stayed up drinking all night"). In another study, Beattie and Shovelton (1999) showed that listeners answer questions about the size and relative position of objects in a speaker's message more accurately when gestures were part of the description and conveyed additional information than speech.

Furthermore, findings show that gesture is not semantically perceived as an independent system of representation, but it also influences speech comprehension. In a priming study by Kelly et al. (2010) participants were presented with action primes (e.g., someone chopping vegetables) followed by bimodal speech and gesture targets. They were asked to press a button if what they heard in speech or gesture depicted the action prime. Participants related primes to targets more quickly and accurately when they contained congruent (p. 600) information (speech: “chop”; gesture: chop) than when they contained incongruent information (speech: “chop”; gesture: twist). Moreover, the strength of the incongruence between overlapping speech and gesture affected processing, with fewer errors for weak incongruities (speech: “chop”; gesture: cut) than for strong incongruities (speech: “chop”: gesture “open”). This indicates that in comprehension, the relative semantic relations between the two channels are considered, providing evidence against independent processing of the two channels. Furthermore and crucially, this effect was bidirectional and was found to be similar when either speech or gesture targets matched or mismatched the action primes. That is, gesture influenced processing of speech and speech influenced processing of gesture. Further research has shown that gestures also show semantic priming effects. Yap et al. (2011) have shown that iconic gestures—shown without speech—(highly conventionalized ones such as flapping both hands on the side meaning bird) prime sequentially presented words.

The evidence for semantic integration between representational gestures and speech has also been corroborated in many neurocognitive studies. They have shown that comprehension of iconic gestures involves brain activations known to be involved in semantic processing of speech (i.e., modulation of the electrophysiological recording component, N400, which is sensitive to the ease of semantic integration of a word to previous context). For example, Wu and Coulson (2007) found that semantically incongruous gestures (shown without speech), when presented after cartoon images, elicited a negative-going event-related potential (ERP) effect around 450 ms, in comparison to gestures that were congruent with the cartoon image. Furthermore, unrelated words followed by gestures (shown without their accompanying speech) also elicited a more negative N400 than related words.

Holle and Gunter (2007) extended the use of the ERP paradigm to investigate the semantic processing of gestures in a speech context. They asked whether manual gestures presented earlier in the sentence could disambiguate the meaning of an otherwise ambiguous word presented later in the sentence and investigated the brain’s neural responses to this disambiguation. An electroencephalograph (EEG) was recorded as participants watched videos of a person gesturing and speaking simultaneously. The experimental sentences contained an unbalanced homonym in the initial part of the sentence (e.g., *She controlled the ball* . . .) and were disambiguated at a target word in the subsequent clause (*which during the game* . . . versus *which during the dance* . . .). Coincident with the homonym, the speaker produced an iconic gesture that supported either the dominant or the subordinate meaning. ERPs were time-locked to the onset of the target word. The N400 to target words was found to be smaller after a congruent gesture and larger after an incongruent gesture, suggesting that listeners can use the semantic information from gesture to disambiguate upcoming speech.

In another ERP study, Özyürek et al. (2007) examined directly whether ERPs measured as a response to semantic processing evoked by iconic gestures are comparable to those evoked by words. This ERP study investigated the integration of co-speech gestures and spoken words to a previous sentence context. Participants heard sentences in which a critical word was accompanied by a gesture. Either the word or the gesture was semantically anomalous with respect to the previous sentence context. Both the semantically anomalous gestures and anomalous words to previous sentence context elicited identical N400 effects, in terms of the latency and the amplitude.

fMRI studies also show that perceiving gestures in a speech context involves the recruitment of the left-lateralized frontal-posterior temporal network (left inferior frontal gyrus (p. 601) (IFG), medial temporal gyrus (MTG), and superior temporal gyrus/sulcus (STG/S)). These brain areas are known to be sensitive to semantic processing linguistic information (see Özyürek, 2014, for a broader overview). Using a functional magnetic resonance imaging (fMRI) method, Straube et al. (2012) isolated the brain's activation in response to iconic gestures to see whether it overlaps with areas involved in processing verbal semantics. fMRI measures brain activity by detecting associated changes in blood flow (i.e., blood-oxygen-level-dependent (BOLD) response), relying on the fact that blood flow and neural activation are coupled. In this study, they compared the brain's activation triggered by meaningful spoken sentences, with sentences from an unknown language, and they also compared activation for co-speech gestures presented without their accompanying speech, and meaningless gestures also without speech. Meaningful iconic gestures activated the left IFG, bilateral parietal cortex, and bilateral temporal areas. The overlap of activations for meaningful speech and meaningful gestures occurred in the left IFG and bilateral MTG. These findings are consistent with another study by Xu et al. (2009) showing that left IFG and posterior MTG are involved in the comprehension of communicative gestures (i.e., pantomimes such as opening a jar without speech) as well as speech glosses of the same gestures (i.e., open jar) presented separately.

Further fMRI studies have attempted to locate the brain areas involved in *integrating* information from speech and gesture. Perceiving iconic gestures mismatching or complementing information or sensitivity to bimodal matching information comparing to speech or gesture alone recruits left IFG, bilateral posterior superior temporal sulcus (STSp), and middle temporal gyrus (MTGp). Interestingly, these are the areas that are also involved when increased semantic processing is required during speech comprehension (see Dick et al., 2012). Dick et al. (2012) for example found left IFG to be sensitive to meaning modulation by iconic gestures; that is, more activation in this area for complementary (speech: "I worked all night"; gesture: type) than redundant gestures accompanying speech (speech: "I typed all night"; gesture: type). Complementary gestures add information and require more semantic processing than redundant gestures. Finally, Skipper et al. (2009, 2015) found that when hand movements (iconic gestures) were related to the accompanying speech, left IFG (pars triangularis and pars opercularis) exhibited a weaker influence on other motor- and language-relevant cortical areas compared with when the hand movements were meaningless (i.e., grooming gestures or "self-adaptors") or when there were no accompanying hand movements. In a recent paper Skipper (2014) has also proposed a model (i.e., NOLB model) according to which gestures can be seen as a predictive context (through their activations of semantic and motor cortices) for speech comprehension, especially for auditory cortex.

If listeners/perceivers are integrating gestures into the speech context, then the next question is how robust is this integration process? Can it be modulated, is it obligatory/automatic, or is this a system unique to gesture? Recent research suggests that this integration can be modulated by several factors. First, the interactions between the two modalities seem to be sensitive to the temporal synchrony of the two channels as well as to the perceived communicative intent of the speakers, and thus seem to be flexible rather than obligatory depending on the communicative context. After all, spontaneous speech is not always accompanied by gestures; gestures might sometimes be asynchronous with the relevant speech segment (Chui, 2005), and the frequency or the informativeness of the representations in gestures can vary depending on the communicative nature of the situation (i.e., whether there is shared common ground between the listener and the addressee or not, and so on).

(p. 602) Habets et al. (2011) investigated the degree of synchrony in speech and gesture onsets that is optimal for semantic integration of the concurrent gesture and speech. Videos of a person gesturing were combined with speech segments that were either semantically congruent or incongruent with the gesture. The onset of the gesture strokes (i.e., the meaningful part of the gesture, but not the preparation) and speech were presented with three different degrees of synchrony: a stimulus onset asynchrony (SOA) 0 condition (the gesture stroke onset and the speech onset were simultaneous) and two delayed SOAs, where speech was delayed by 160 ms (partial overlap with speech) or 360 ms (speech onset presented after gesture stroke was executed; no overlap between the two) in relation to the gesture stroke onset. ERPs time-locked to the speech onset showed a significant difference between semantically congruent versus incongruent gesture—speech combinations for the N400 component with SOAs of 0 and 160 ms, respectively, but not for the 360 ms SOA. Therefore, the closer speech and gesture are temporally to each other (or at least when some temporal overlap is possible), the more likely they are to be integrated with each other

Not only the synchrony, but also the perceived communicative intent of the speakers seems to modulate the speech–gesture integration or the semantic processing of gestures. ERP studies by Kelly et al. (2007) have demonstrated that our brain integrates speech and gesture less strongly when the two modalities are perceived as not intentionally coupled (i.e., gesture and speech being produced by two different persons) than when they are perceived as being produced by the same person. In this study, adults watched short videos of gesture and speech that conveyed semantically congruous and incongruous information. In half of the videos, participants were told that the two modalities were intentionally coupled (i.e., produced by the same communicator), and in the other half, they were told that the two modalities were not intentionally coupled (i.e., produced by different communicators). When participants knew that the same communicator produced the speech and gesture, there was a larger bilateral frontal and central N400 effect to words that were semantically incongruous versus congruous with gesture. However, when participants knew that different communicators produced the speech and gesture—that is, when gesture and speech were not intentionally meant to go together—the N400 effect was present only in right-hemisphere frontal regions. The results demonstrate that pragmatic knowledge about the intentional relationship between gesture and speech modulates neural processes during the integration of the two modalities.

Finally, Holler et al. (2014) have investigated how listeners/viewers comprehend speech–gesture pairs in a simulated triadic communication setting where the speakers' eye gaze is directed at them versus to another addressee (i.e., away from them). Participants were scanned (fMRI) while taking part in triadic communication involving two recipients and a speaker. The speaker uttered sentences that were accompanied by complementary iconic gestures (speech: "she cleaned the house"; gesture: mopping) or with speech only. Crucially, the speaker alternated her gaze direction toward or away from the participant in the experiment, thus rendering him/her in two recipient roles: addressed (direct gaze) versus unaddressed (averted gaze) recipient. "Speech and gesture" utterances, but not "speech only" utterances, produced more activity in the right MTG, one of the brain areas found consistently involved in speech–gesture integration, when participants were addressed than when not addressed. Thus, when the eye gaze of the speaker is averted away from the listener/viewer, indexing decrease in the perception of communicative intent, integration of the (p. 603) two channels and/or semantic processing gesture might be reduced (also see Holler et al., 2014, for similar effects shown by behavioral measures).

Finally, one study has investigated to what extent perception of information from gesture is special by comparing integration of gesture to that of manipulable actions (Kelly et al., 2015). This study shows that listeners/viewers are less likely to integrate overlapping action (e.g., somebody actually drinking from a glass) information to a speech context than a gesture (e.g., someone performing a drink gesture). This suggests that the communicative nature of gesture might be triggering more integration with speech than non-communicative actions, corroborating findings from the aforementioned studies.

25.5 Summary: Role of gesture in language comprehension

When it comes to processing of gestures in speech context, there is a robust involvement of semantic processing, similar to that involved in processing spoken language and recruiting similar brain areas. Studies show further this is not an independent system, but gestures are processed in relation to the speech context they occur in and they in turn influence speech comprehension. The Habets et al. (2011) study also shows that the temporal overlap between speech and gesture is crucial for their integration, pointing to the role of not only gesture in speech processing but also of speech in gesture processing.

Thus, one can suggest that there is an incremental meaning interpretation occurring between the two channels. This is in line with the Integrated Systems hypothesis by Kelly et al. (2010) according to which speech and gesture, mutually constrain each other's meaning interpretation in an online manner. Skipper (2014) has also proposed gesture as a predictive context for speech comprehension consistent with this view.

25.6 General conclusions: Toward a unified account of the role of gesture

Both the results of the production and the comprehension studies reported here suggest that information from speech and gesture, is processed in an interactive way during production *and* comprehension; recruiting similar semantic processing and neural correlates in the brain, rather than being processed in a distinct, modular, or modality-specific fashion. Even though many current models view gesture production as mere action simulations and arising originally independent of the language system, postulating gesture processing as linked to language processing seems more plausible when we consider converging evidence from both production and comprehension. Furthermore, this approach is more in line with the genuinely communicative nature of gesture use; that it is sensitive to context, discourse, and the listener's knowledge status. This is not to say that gestures are completely independent of action processing, but any account of gesture processing should consider the (p. 604) interactions between speech and gesture, and their communicative nature both during production and comprehension to give a unified account of their role in language processing. This will offer unique insights into understanding language processing in context in general.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*, 169–88.

Azar, Z., Backus, A., & Özyürek, A. (2017). Highly proficient bilinguals maintain language-specific pragmatic constraints on pronouns: Evidence from speech and gesture. In: Gunzelmann, G., Howes, A., Tenbrink, T., & Davelaar, E. (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (CogSci 2017) (pp. 81–6). Cognitive Science Society, Austin, TX.

Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning. An integrated message model of language use in face-to-face dialogue. *Journal of Language and Social Psychology*, *19*, 163–94.

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, *123*, 1–30.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, *21*, 1605–10.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*, 14–27.

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *The Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 257–70.

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, *37*, 871–87.

Clark, H. (1996). *Using Language*. Cambridge University Press, Cambridge.

Cook, S. W., & Tannenhous, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, *113*, 98–104.

Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. M. (2013). Gestural viewpoint signals referent accessibility. *Discourse Processes*, *50*(7), 431–56.

Dick, A. S., Goldin-Meadow, S., Solodkin, A., & Small, S. L. (2012). Gesture in the developing brain. *Developmental Science*, *15*, 165–80.

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

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge University Press, Cambridge, UK.

Floyd, S. (2016). Modally hybrid grammar? Celestial pointing for time-of-day reference in Nheengatú. *Language*, *92*(1), 31–64.

Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review*, *103*, 34–55.

Goldin-Meadow, S., So, C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163–8.

(p. 605) Gullberg, M. (2011). Language-specific encoding of placement events in gestures. In: Bohnemeyer, J., & Pederson, E. (Eds.), *Event Representation in Language and Cognition* (pp. 166–88). Cambridge University Press, Cambridge, MA.

Gullberg, M., Hickmann, M., & Hendriks, H. (2008). Learning to talk and gesture about motion in French. *First Language*, *28*(2), 2000–236.

Gullberg, M., & Narasimhan, B. (2010). What gestures reveal about how semantic distinctions develop in Dutch children's placement verbs. *Cognitive Linguistics*, *21*(2), 239–62.

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience*, *23*, 1845–54.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*, 1175–92.

Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S., & Özyürek, A. (2014). Eye'm talking to you: Speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Social, Cognitive and Affective Neuroscience*, *10*(2), 255–61.

Holler, J., & Beattie, G. (2003). How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process? *Semiotica*, *146*, 81–116.

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*, 692–7.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*, 495–514.

Kelly, S. D., Barr, D., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, *40*, 577–92.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*, 260–7.

Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain Language*, *101*, 222–33.

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture and action during language comprehension. *Psychonomic Bulletin & Review*, *22*, 517–23.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press, Cambridge, MA.

Kita, S., & Özyürek, A. (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.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, *22*(8), 1212–36.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*, 396–414.

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, *31*, 533–52.

Levelt, P. (1989). *Speaking*. MIT Press, Cambridge.

McNeill, D. (1992). *Hand and Mind*. University of Chicago Press, Chicago, IL.

(p. 606) Masson-Carro, I., Goudbeek, M. B., & Krahmer, E. J. (2016). Can you handle this?: The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, *31*(3), 430–40.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*(2), 119–41.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts the temporal asynchrony of hand gesture and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 615–23.

Özçalışkan, S., Lucero, C., & Goldin-Meadow, S. (2016a). Is seeing gesture necessary to gesture like a native speaker? *Psychological Science*, *27*(5), 737–47.

Özçalışkan, S., Lucero, C., & Goldin-Meadow, S. (2016b). Does language shape silent gesture? *Cognition*, *148*, 10–18.

Özyürek, A. (2002). Do speakers design their co-speech gestures for their addresees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*(4), 688–704.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *369*(1651), 20130296.

Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures? Insights from crosslinguistic variations and similarities. *Gesture*, *5*(1/2), 219–40.

Özyürek, A., Kita, S., Allen, S., Furman, R., Brown, A., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental Psychology*, *44*(4), 1040–54.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–16.

Peeters, D., Chu, M., Holler, J., Hagoort, P., & Özyürek, A. (2015). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *Journal of Cognitive Neuroscience*, *27*(12), 2352–68.

Perniss, P. M., & Özyürek, A. (2015). Visible cohesion: A comparison of reference tracking in sign, speech, and co-speech gesture. *Topics in Cognitive Science*, *7*(1), 36–60.

Schubotz, L., Holler, J., & Özyürek, A. (2015). Age-related differences in multi-modal audience design: Young, but not old speakers, adapt speech and gestures to their addressee's knowledge. In: Ferré, G., & Tutton, M. (Eds.), *Proceedings of the 4th GESPIN —Gesture & Speech in Interaction Conference*. Université of Nantes, Nantes.

Skipper, J. (2014). Echoes of the spoken past: How auditory cortex hears context during speech perception. *Philosophical Transactions of the Royal Society of London. B-Biological Science*s, *369*(1651), 20130297.

Skipper, J. (2015). The NOLB model: A model of the natural organization of language and the brain. In: Willems, R. (Ed.), *Cognitive Neuroscience of Natural Language Use* (pp. 101–34). Cambridge University Press, Cambridge.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Current Biology*, *19*, 661–7.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand in hand. *Cognitive Science*, *33*, 115–25.

(p. 607) Straube, B., Green, A., Weis, S., & Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: An fMRI study. *PLoS One*, *7*, e51207.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In: Shopen, T. (Ed.), *Language Typology and Semantic Description. Vol. III: Grammatical categories and the lexicon* (pp. 36–149). Cambridge University Press, Cambridge.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–32.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, *114*, 591–600.

Wu, Y., & Coulson, S. (2007). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*, *14*(1), 57–63.

Xu, J., Gannon, P., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(49), 20664–9.

Yap, D. F., So, W. C., Yap, M., & Tan, Y. Q. (2011). Iconic gestures prime words. *Cognitive Science*, *35*, 171–83.

Notes:

([1]) Note that in the GSA model, gesture's link to speech is seen best as gesture production, helping conceptualization for speech. It does not propose explicit bidirectional communication between speech and gesture unlike what is assumed in other models such as Interface Hypothesis (see next). The GSA model also accepts the possibility that linguistic planning involves simulations of perceptual events, which in turn can influence production of gestures (Hostetter & Alibali, 2008; p. 508).

Asli Özyürek

Asli Özyürek, Radboud University Nijmegen and Max Planck Institute for Psycholinguistics, the Netherlands