Liesbet Quaeghebeur

University of Antwerp

The ‘All-at-Onceness’ of Embodied, Face-to-Face Interaction

Much has been written on the cognitive architecture behind verbal interaction, yet little is to be found about how language users successfully produce and understand multimodal expressions: i.e., face-to-face talk involving co-expressive and synchronized facial expressions, gesticulation, gaze, posture, and other bodily means of conveying information. I survey existing theories about this matter – mainly coming from gesture studies – and extend McNeill’s (1992, 2005) intrapersonal language-thought-hand nexus to the interpersonal realm: the speech/gesture unit does not transport thought from one mind into another; rather, embodied interactions constitute shared thoughts. Therefore – opposing the sender-receiver, coding-decoding picture – I argue for the ‘all-at-onceness’ of face-to-face interaction.

Keywords: multimodality, gesture, expression, cognition, face-to-face interaction

INTRODUCTION

The purpose of this paper is to discuss the cognitive architecture behind the production and understanding of multimodal language use. Traditionally defined, multimodal language use involves not only the verbal medium (i.e., speech), but also one or more other bodily means of conveying information: e.g., facial expression, intonation, gesticulation, posture, stance, and gaze. While there is an abundance of theories on how verbal (‘uni-modal’) expression and understanding works, I start from the finding that face-to-face interaction is, in fact, always multimodal, and ask the question how these different modes are cognitively related to each other.

The account I present defends the idea that expressive multimodality is characterized by an ‘all-at-onceness’ in the intrapersonal as well as interpersonal realm. Because thought coincides with expression, the multimodal display is produced and understood immediately and as a whole. This account stands in clear opposition to cognitive-processing accounts, which, on the one hand, postulate a distinction between mental content and its expression/reception; and, on the other, assume that the

Address for correspondence: Department of Philosophy, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium, liesbetq@gmail.com.

multimodal nature of the expression/reception is processed by integrating information from different cognitive channels. These presuppositions are related. That is why, in investigating how different expressive modes cognitively relate to each other, one must deal with more than just that question. Equally important is the issue of how multimodal expressions – irrespectively of how these are organized internally – relate to thought, or the speaker’s and listener’s mental content.

In an attempt to show these distinctions and nuances and their importance for cognitive theory clearly, I will proceed by sustaining the – in my view artificial – distinction between expressive subjectivity and expressive intersubjectivity: i.e., between “outward” expression and mutual “inward” understanding, as dealt with in the following two sections, both of which have a similar outline. Looking at multimodal language use from the speaker’s viewpoint, I begin my investigations by first going through the different expressive possibilities that bodies harbor. I move on to the actual cognitive question by reviewing existing theories on multimodal language production, originating in the field of gesture studies, that regard multimodal expression as the externalization of thought produced for the purpose of communication. Then – still looking from the speaker’s viewpoint – I briefly discuss the work of David McNeill who, under the heading of *co-expressive non-redundancy*, defends – in the intrapersonal realm – exactly the kind of ‘all-at-once’ theory I wish to extend to the interpersonal realm. Moving on to the intersubjective section – which looks at multimodal language use from the viewpoint of several conversational participants in interaction – I first take a look at some additional expressive modalities that the theoretical introduction of the social ‘other’ brings along. Surprisingly, there is little theory available in the literature regarding the cognitive architecture behind understanding multimodal language use. Using what is found there, I reconstruct two general theoretical possibilities, both of which assume that a (now shared) mental content is inferred from multimodal expression through cognitive processing. Then, elaborating on McNeill’s framework and incorporating insights from Merleau-Ponty’s phenomenology, I discuss what I call the ‘all-at-onceness’ of multimodal linguistic interaction, with conversational participants engaging in the joint construction of meaning: i.e., thinking and expressing themselves to each other – rendering the sender-receiver, cognitive-coding-decoding model unnecessary.

THE EMBODIMENT OF EXPRESSIVE SUBJECTIVITY

The multimodal expressive system

When it comes to expressing oneself to a social ‘other’, what natural means for doing so does one have at one’s disposal? Using words – i.e., the verbal modality – immediately springs to the fore: people consider it the most efficient channel for externalizing mental content and transferring information. So Loncke *et al.* (2006:16) write, “for most communicators, speech is so much in the foreground and so precise that the other modalities and channels appear redundant or even superfluous.” Subsumed under *non-verbal communication* or *body language*, these other modalities are popularly regarded as the body’s sneaky way of divulging one’s hidden true intentions: one really notices other people’s non-

verbal signs only when they do not fully accord with the picture of herself that the speaker is verbally trying to put forward. That said, non-verbal communication is omnipresent in ordinary face-to-face conversation; estimates are that an important part of rich, pragmatic meaning is conveyed and understood through non-verbal signs¹.

Put aside the verbal/non-verbal distinction for the moment to take an as-unbiased-as-possible descriptive view of the human expressive system. From head to toe, basically every body part that can be moved is used to convey meaning, in multimodal social interaction. The muscular organization of the human face allows for a variety of expressions that are unknown anywhere else in the animal kingdom (Cole 1998). Though they have a limited action radius, nevertheless the eyebrows, ears, and nose play a definite role in facial expression. Meanwhile, the eyes and eyelids are famous for their expressivity, which they achieve through the intricate manipulation of gaze direction, eye muscle tension, squint, and speed of blinking.

The mouth harbors three expressive modes. First, there is its outward appearance – opening, shape, and tension of the lips – through which the mouth takes part in facial expression. Closely related is the verbal function of the speech organs: i.e., the verbal mode. Verbal language is never produced in a neutral manner but necessarily expresses meaning through modulation of loudness, pitch, tempo, rhythm, melody, articulation, stress, resonance, and nasality.

It is time to leave the face and continue downwards. The body as a whole assumes a posture, indicative of a person's general attitude towards the world and others: e.g., lively, relaxed, or submissive. Shoulders shrug, slouch, or are pulled back. Arms and especially hands and fingers – due to their agility and muscular intricacy – provide their owner with another massively expressive instrument. Concluding my downward investigation, pelvis and leg position contribute to posture, while feet are more isolated and agile again – tapping, stamping, shaking – their expressive movements appearing to be more closely connected to the manual or gestural mode. Some theorists say that thought constitutes a mode also. I will leave the issue here, however, because, in a sense, it sums up the subject of this paper and will be discussed sufficiently later. Meanwhile, for Merleau-Ponty (2007), Goodwin (2007), and Andrn (*this volume*), external objects take part in expression in very much a mode-like manner: e.g., the map of the city one is both looking and pointing at.

In face-to-face interaction, meaning is expressed by deploying several modes simultaneously: sometimes more, sometimes fewer, depending on the liveliness of the conversation and the complexity of the subject matter. The investigation of expressive multimodality – i.e., the interaction of different bodily modes in the expression of meaning – is a fairly recent undertaking. *Gesture studies* seems to

¹A widely cited study by Mehrabian & Ferris (1967) argues that the verbal mode conveys only 7% of meaning, the vocal mode (i.e., aspects of voice such as prosody) 38%, and the facial mode 55%. These extreme results must be taken with a grain of salt. The conditions under which the experiment was conducted were artificial; the materials used were impoverished; and the subject matter concerned only the communication of positive or negative emotion. Additionally, one might ask what it means for meaning to be made up of percentages of sub-meanings. Taking away any of the cues is, to some degree, detrimental to the original whole, suggesting that all modalities are equally important; yet, the suppression of some has greater impact than others on the success of conversation.

be the only field at present with a systematic focus on the simultaneous occurrence of two or more modes, including, at minimum, the manual and the verbal. Depending on the investigation's theoretical framework, one might also consider the involvement of and consequences for simultaneous thought (see e.g. McNeill 1992; Goldin-Meadow 2005). The investigation of language, gesture, and thought from a cognitive point of view is the main guide throughout this paper.

Studies of expressive multimodality that do not focus solely on gesture-as-gesticulation but also include modes such as posture, facial expression, and gaze remain largely descriptive or quantitative (see e.g. Goodwin 2003; 2006; 2007; Bavelas & Gerwing 2007; Bavelas *et al.* 2002; Kita 2003). A final category of investigations include descriptions of interaction between expressive modes that have traditionally been less focused on. Usually these modes are not studied on their own but for the purpose of other scientific interests: e.g., Drahota *et al.* (2007) on facial/vocal interactions, Gullberg (2003) on eye movements and gesture, McClave *et al.* (2008) on head nods and speech, and Okrent (2002) on aspects of speech and the full body.

Speech/gesture as the communication of thought

What does the cognitive architecture behind this simultaneity of expressive modes look like? How does the mind organize the use of several modalities at once, making them co-expressive: expressing (different aspects of) the same thought; but non-redundant: the whole cannot be reduced to its parts; while synchronizing their production perfectly, to the millisecond? These questions – potentially a topic for theoretically minded investigators of any linguistic modality – have, so far, been addressed only by gesture theorists: more particularly, by those studying imagistic gestures: i.e., spontaneous, non-conventional, idiosyncratically produced gesticulations of the hands during speech. Any cognitive theory of multimodal language production depends on one's ideas about what gestures are for and how they relate to speech. The field does not even subscribe unanimously to the idea that gestures play a role in face-to-face communication. Rimé and Schiaratura (1991), for example, argue that co-speech gestures are rather a byproduct of verbal articulation: on the way to linguistic encoding, non-verbal means are activated, resulting in what one perceives as gesturing; but the process' actual task is formulation of a lexical-syntactic form. Butterworth (1980) and Butterworth and Hadar (1989) similarly argue against the communicative nature of gestures, though they allow that gestures may play a role in verbal utterance formulation: when, for some reason, the natural flow of speech breaks down due to e.g. interruptions, complexity of content, or confusion, in which case gestures help in getting verbal utterance production back on track. That said, most cognitive theories of multimodal language production *do* ascribe a communicative role to gesture; this has immediate repercussions

on the proposed architectures of the mind: speech and gesture production become two interrelated processes.

Figures 1 – 4 – The inner workings of the mind and brain.

Three different cognitive architectures of multimodal language use take up and amend Levelt's (1989) modular approach to speech production explicitly. According to Levelt's account of the inner workings of a speaker's mind/brain (see Figure 1.), the processing component 'Conceptualizer' comes up with a preverbal message, which, in the 'Formulator', undergoes sequential grammatical encoding, of which a surface structure is the result; and phonological encoding, of which a phonetic plan is the result. The 'Articulator' is then ready to turn this internal speech into overt speech. If one acknowledges a communicative role for gesture, how might the gesture production process be connected to Levelt's speech production?

Krauss and colleagues (2000) contend that gestures can have a communicative function, but that this is not their primary function. Extending the Butterworth picture, they argue that gestures help the formulation of speech by aiding the retrieval of elusive words from lexical memory. Knowledge is stored in memory in several representational formats: e.g., spatial/dynamic or propositional; when a concept is activated in one format, related concepts are activated in other formats. Applied to Levelt's model, where speech is activated post-Conceptualizer, gestures are activated from working memory at a stage pre-Conceptualizer. Speech and gesture are two production systems that operate in concert, the main function of gesture being to facilitate the retrieval of its 'lexical affiliate'. Krauss and colleagues argue that gesturing has an effect on word-form retrieval primarily at the stage of phonological encoding (see Figure 2.).

Kita and Özyürek (2003) advance a theory about how speech and gesture-production processes are connected, so as to explain their common function in communication. The starting point is the finding that the language one speaks is related to how one gestures: e.g., English has the expression 'to swing across' (e.g., by using a rope), which expresses the manner of the action (*to swing*) as well as the path (*across*); English speakers tend to convey this expression manually by making an arc-like gesture with both hands, holding an invisible rope. Speakers of Japanese do not have a verb meaning 'to swing' and must use two different expressions to convey path and manner: i.e., 'to go across' / 'by holding on to the rope'. Their gesturing reflects the dissociation of verbal properties: one hand holds the rope; the other outlines the path that is followed. From this and other examples, Kita and Özyürek infer that there must be reciprocal links between speech and gesture production. They split Levelt's Conceptualizer into two halves and add a gesture module. One half is turned into the Communication Planner, which generates a communicative intention roughly specifying the information to be expressed while deciding which modalities will be used to communicate this information (see Figure 3.). The output of the Communication Planner is then fed into the Message Generator – which sets in motion speech production – as well as into the Action Generator, which sets in motion gesture production. Via a communication channel between Message Generator and Action Generator, speech and gesture production exchange information: e.g., about producing the path and manner of the action event separately, so that the outcomes will match.

De Ruiter (2007) draws attention to the way part of the information to be communicated is expressed in speech and part in gesture: sometimes overlapping in meaning, sometimes complementary. From this, he infers that speech and gesture are planned together explicitly, to communicate a coherent, multimodal message. De Ruiter stresses this distribution of labor and uses it as the main argument for his thesis that gesture and speech are planned together, at an early stage in utterance production. In de Ruiter's picture, the division of communicative labor between speech and gesture is already resolved at the stage of conceptualization (see Figure 4)².

Figure 5 – Thought precedes its material externalization.

These accounts of how speech and gesture (or multimodal expression in general) are produced simultaneously differ in the cognitive architectures they propose; yet they share a common concept of communication that focuses on the purposiveness – or the transport-like function – of the process. One must ask: given the multimodal nature of communication, what is being communicated? The theories I have discussed so far provide only an implicit answer: what is being communicated is a thought or *communicative intention*³. Note that the role of thought in the communicative process is not merely passive: i.e., it is not only content to be transmitted. A communicative intention is also the instigator of the communicative process: when one wishes to share a thought with someone, one knows that that person cannot directly perceive what one is thinking, so one necessarily makes use of the body's ability to transform perceptible reality: i.e. to produce sound and make a visual display to together

²De Ruiter (2007) repeatedly links his cognitive architecture with Kendon's (2004), but Kendon has never subscribed explicitly to a modular, Levelt-type account of multimodal language processing. Kendon's descriptive, taxonomic study of gesture has been of great importance for renewed scientific interest in the topic from the 1970s onwards. Kendon was one of the first to have at his disposal audiovisual recordings and playback materials for the study of gesture. He came to three main conclusions – at least with respect to my interest in multimodal language production. First, speech and gesture, though very different in nature, are equally important partners in the speaker's integrated speech-gesture ensemble. Second, this multimodal expression is created to meet the demands of communication; this is why it is so tightly orchestrated. The expressive unit has *recipient design*: vocal and manual actions and interactions are orchestrated for the benefit of communication in general and the listener in particular. Third, listeners are able to parse speech/gesture units from the conversational flow and perceive both aspects as contributing equally to the speaker's communicative product. To them, speech/gesture ensembles function as displays: the speaker concocts bodily audiovisual constellations in order for the listener to pick up meaning.

³ Although Kendon is often referred to and his vocabulary of *communicative intention* is often used, nevertheless the cognitive architecture is not Kendon's.

convey one's thoughts. All theories discussed in this section assume that *thought precedes its material externalization* (see Figure 5).

Speech/gesture as thought and communication

The predominant (though not always explicit) way of thinking about the cognitive production of (multimodal) language sees (multimodal) utterances as, somehow, externalizations of preceding thought. A very different way of thinking is to be found in gesture studies⁴. McNeill (1992, 2005) concludes from the close bond between speech and gesture that they constitute a single unit. Attempts at breaking their bonding fail: either speech and gesture remain together, or they break down together⁵.

Figure 6 – The all-at-onceness of thought, speech, and gesture.

For McNeill (1992, 2005), the speech/gesture unit is not a separate entity expressing a thought for the purpose of communication. Rather, thought is a third aspect of the expressive speech/gesture unity (see Figure 6). Seemingly paradoxically, one could say that thought itself is an expressive mode; and that, likewise, speech and gesture are for cognition as well as communication (keeping in mind that these functions do not go beyond speech/gesture production). The different modes cannot be teased apart, and neither can their functions. The minimal unit of McNeill's analysis does not consist of a

⁴ Streeck (2003), for example, presents a Heidegger-style reflection on gesture studies, in which he argues for dichotomy and the 'thought-first-expression-later' view. Through gesturing, the body enacts its intentions. Although Streeck is unsure how his insights should be implemented in gesture analysis, I believe that an interpretation based on McNeill's framework can provide just the kind of theory that is needed.

⁵ The same is argued by Kendon (1980, 2004). However, for Kendon the speech/gesture unit is produced for the purpose of communicating a thought to the interlocutor. These units are 'displays' which have a *recipient design* and therefore goal-directedness. McNeill (work in progress), argues against Kendon that gestures are not constructed to attain some goal. This aspect of non-goal-directedness goes hand in hand with a theory which views expression to be an element of thought, and not the consequence.

succession of single poles (i.e., separate modes), but an integrated whole of antithetical poles (i.e., a *growth point*). Imagery (gestural) and language (verbal) are two, very different, ways of expressing (aspects of) the same meaning. This is essential to this multimodal entity and not something to be overcome by cognitive processing⁶. In the spirit of Slobin's (1987) *thinking-for-speaking*, McNeill and Duncan (2000) consider differences between the medium of image and the medium of language to be the origin of the speech/gesture/thought dynamic: i.e., the flow of the thinking/speaking/gesturing process. The different modalities used have their own possibilities and limitations, depending on which language one speaks and which gestural culture one belongs to. In this dialectical relation, thinking shapes speaking and gesturing, while speaking and gesturing shape each other and thinking. The modes depend on each other. Thought is no longer something over and above its *material externalization*. Rather, thought coincides with it: being the whole meaning to which the other modes contribute.

Why do people gesture? What are gestures for? The way this oft-asked question (see Krauss 1998 for an overview) is formulated presupposes some reason behind multimodal expressive action: it has a function; one does it to attain some goal. Meanwhile, in my 'all-at-once' picture, gesturing – like speech – *is* (part of) cognition; communication occurs when there is an onlooker who (in a broad sense) speaks the same language and understands the expressive cognition-in-action. Cognition – or thought – does not happen over and above expressive action. Communication does not require, for the listener's benefit, incorporation of dedicated features into the expression⁷. (People do, of course, adapt themselves to their audience if they explicitly choose to do so!). Cognition and communication are not functions of expressive action; they are aspects of it.

The following experimental data and everyday phenomena support my thesis that thought and multimodal expression are mutually dependent:

- There are cases in which the communicative aspect of gesture is removed, with little consequence for expressive action. This suggests that one needs gestures in order to think. Iverson and Goldin-Meadow (1997, 1998) report that people who are blind from birth gesture – even when talking to people they know are also blind. Phantom gesture also exists: Ramachandran and Blakeslee (1998) mention a conversation with Mirabelle, a woman born without arms, who experiences her phantom limbs as gesticulating while she speaks.

⁶ De Ruiter (2007) mistakenly assumes that McNeill follows Levelt's (1989) model when it comes to speech production, so that gestures somehow emerge from thought in an unmediated way. Gesture provides a window into the mind not because it requires no processing, while speech does. For McNeill, both gesture and speech provide a window into the mind, precisely because they are the formulation of thought: the material carrier without which thought would not exist – and not the externalization of it.

⁷ One might argue that e.g. a hand gesture that "hands over" the turn to someone is a counterexample to my claim that communication does not require incorporation of dedicated features into the expression. However, my argument does not take place at the level of the specific content of a speech/gesture unit. Manually handing over a turn is a form of expressing the speaker's intention to regulate the conversation. That the gesture refers to the conversation does not imply it was constructed for communication rather than as a part of cognition – which, in social settings, is more often than not communicative.

- When no one is watching or listening, people still sometimes talk and even gesture to themselves while thinking. This can be extended to ‘external’ devices for verbal expression: e.g., when using a word processor, people sometimes think, type, and talk out loud – all at the same time. Apparently, people also talk when they think they are silent: McGuigan (1966, 1970) has studied subjects while they were thinking, reading in silence, and doing handwriting. Using electromyography, he recorded significant activation of the larynx, lips, chin, and tongue, with associated influence on breathing rate and salivation. Removing the communicative function, people still express their thoughts using more modes than would be necessary if the mental mode were independent from the others⁸.
- Conversely, when the possibility of expressive action is removed, it has consequences for thought. Considering facial rather than verbal or manual expression: subjects suffering from Bell’s palsy, a temporal paralysis of the muscles on one or both sides of the face, complain that they have difficulties formulating clear thoughts (Cole 1998).

So again: why do people gesture? What are gestures for? I have argued that speech and gesture serve both cognitive and communicative functions, without being produced *in order to* serve these functions. Thought coincides with multimodal expression: it is the subject’s formulation of her thoughts to herself; while communication is an external feature, dependent on the listener. One may turn the question around and ask: considering that expression enables thought, and that one possesses a potentially very expressive body, why would one not gesture? Asking the question like this gives better insight into the phenomena of *redundancy* and the synchronization of co-expressive speech and gesture.

The question “why do people gesture?” presumes that the default position of the human body is one of rest, and that any action undertaken – expressive or practical – makes use of the body’s possibilities as economically as possible. This means that people should employ the manual mode only when the verbal/vocal mode fails for some reason.

The experimental literature shows that this is not at all the case. Having observed that everyday human expressive action is massively multimodal, one may consequently let go of the efficiency thesis in exchange for one of an expressive body, whose default mode is multimodal action. I will explore this thesis further.

⁸ If thought and expressive action are so tightly linked, how is it possible for people to formulate thoughts in silence? Ryle (1978:28) reminds his readers that this faculty is not self-evident: “th[e] trick of talking to oneself in silence is acquired neither quickly nor without effort; and it is a necessary condition of our acquiring it that we should have previously learned to talk intelligently aloud and have heard and understood other people doing so. Keeping our thoughts to ourselves is an accomplishment. It was not until the Middle Ages that people learned to read without reading aloud. Similarly a boy has to learn to read aloud before he learns to read under his breath, and to prattle aloud before he prattles to himself. Yet many theorists have supposed that the silence in which most of us have learned to think is a defining property of thought.... People tend to identify their minds with the ‘place’ where they conduct their secret thoughts. They even come to suppose that there is a special mystery about how we publish our thoughts instead of realizing that we employ a special artifice to keep them to ourselves.”

Why would people *not* gesture? Proponents of the efficiency thesis hold that, whenever some meaning is communicated sufficiently using only the verbal mode, a co-occurrent gesture that conveys the same meaning, without illuminating any aspect of meaning aspect not already conveyed verbally, is redundant (see e.g. Crowder & Newman 1993). The very terminology of ‘redundancy’ suggests that, had these gestures not been produced, nothing would have been lost. This answers my question: when gesturing adds no communicative value to the expression, there is no reason to gesture. That said, it seems that redundancy is spoken of only when gesture is concerned – but never other modalities. When a cry of pain is produced with a loud and high voice – eyes shut tight, with a brisk move away from the cause of pain – one does not think this a redundantly expressive way to convey an intention. In cases of pain or joy, it seems only normal to people that the body finds as many ways as possible to ‘defuse’ itself or ‘channel away’ its intentions. In a framework where thought does not precede its expression but rather is dialectically revealed by it, multimodality makes sense as the default form of linguistic expression: it allows thought to be nuanced and multifaceted in cases of non-‘redundancy’, while placing stress on certain elements in cases of ‘redundancy’.

Another issue is directly resolved by taking multimodality as the default: the synchronization of the co-expressive speech/gesture unit. Models which take speech and gesture to be the externalization of thought must explain how these different modes – along with facial expression, postural organization, etc. – are timed to remain in perfect synchrony, so as to constitute a meaningful and coherent display for the listener’s benefit. Because modes are meant to be separate cognitive channels, synchronization will prove to be a complex processing task. Meanwhile, in McNeill’s non-modular model, synchrony is not a problem at all: “in a dialectic, synchrony is part of the thought itself and arises from the speaker’s conceptual intentions; there are no external signals needed or desired” (McNeill 2005:134). Following Vygotsky’s (1962) distinction between analysis into elements (i.e., the verbal element, the gestural element) and analysis into units (i.e., the speech unit, the gesture unit, the thought unit), he defies the tendency to attribute autonomous status to every element that can be singled out during investigation of the phenomenon. By virtue of this autonomy attribution, any correlations between the behaviors of separate elements – such as speech and gesture synchrony – must be explained by reference to factors external to the elements involved. However, this attribution autonomy is contingent upon the investigator. Bodily expression – including thought – may also be approached as a unit: emerging together, their aspects are naturally correlated and therefore synchronized.

The all-at-onceness of expressive bodily behaviour provides an interesting alternative to the cognitive-processing view in which thought precedes its externalization. However, so far I have only talked about “outward” expressive action and not touched upon the question of how listeners can also understand multimodal utterances. One must step back, to include the behavior of other conversational participants.

THE EMBODIMENT OF EXPRESSIVE INTERSUBJECTIVITY

Multimodal face-to-face interaction

When it comes to expressing oneself to a social ‘other’, what natural means does one have at one’s disposal? Earlier, I reviewed a person’s expressive potential in a head-to-toe analysis of the human body. When people express themselves in linguistic face-to-face interaction with another person, they are – by definition! – never alone. One must take into account the effect of the other’s presence on the now shared expressive possibilities. In communication, one does not only express oneself outwardly, as it were; one’s behavior is also meaningful because it is always related to another person’s behavior.

Now that another person has appeared, all of a sudden one’s bodily behavior derives meaning from how it is spatially located *vis-à-vis* its interlocutor. One’s body and face may be oriented towards or away from the other person. The distance between the bodies (i.e., *proxemics*) is telling, as is whether and how one touches the other while speaking (*haptics*). The timing of linguistic interaction is salient: one may show willingness to wait until one’s conversational partner has finished speaking, or one may speed up one’s speech when giving orders to an inferior (*chronemics*). Mutual gaze is subjected to cultural patterns: e.g., looking at and looking away during conversational turn-taking: see e.g. (Argyle & Cook 1976); so that any deviation from the norm – or, for that matter, lack of deviation from the norm – is expressive of meaning over and above the functional property of gaze during interaction: “I am addressing/attending to you”.

Speech/gesture as the communication of thought

Whereas – as I have shown – some theories address multimodal language production, in particular how gestures are produced in connection to speech – surprisingly little is to be found about how multimodal comprehension works, including how gestures are *understood* in connection to speech. A possible reason for this is the lack of a standard model for speech comprehension, such as Levelt (1989) has provided for speech production: something one can criticize, enhance, or generally build upon. The basic idea of Levelt’s psycholinguistic model – that speech production starts with a thought or communicative intention, which is then processed and externalized – is implicitly present in many theories of speech comprehension: so “the receptive side of communication is often considered as essentially an inverse process where the uttered message is de-composed in an attempt to retrace the sender’s intention” (Loncke *et al.* 2006:171). Supposing one acknowledges that other modes of bodily expression contribute to linguistic communication: i.e., that not only does the speaker use multiple modes to express herself, but the listener’s cognitive system takes these cues into account as channels meant to be perceived⁹: one may ask how these expressive modes are processed in relation to one

⁹ In the gesture literature, it is not taken for granted at all that the bodily signals people produce expressing themselves in social interaction are also taken to be meaningful by the listener (see e.g. Alibali *et al.* 2001).

another. Following the suggestion of Özyürek and colleagues (2007:609), I wish to distinguish between sequential and parallel models of multimodal comprehension.

Figure 7 – Sequential processing of a multimodal utterance.

In the sequential model, modes are processed one at a time. This is analogous to the view that semantic interpretation of sentences precedes integration of pragmatic or contextual information (see e.g. Forster 1979). Speech will typically be processed first, the other modes later (see Figure 7). Özyürek and colleagues note that this view is in line with that of gesture theorists such as Krauss and colleagues (1991), who argue that listeners construct the meaning of speaker gestures on the basis of concurrent speech, while defending the (heavily contested) view that gestures add no information to what is already conveyed through speech. For this to be so, listeners must process speech segments first, gesture later. To be clear: the sequential view of multimodal language comprehension is not defended explicitly by anyone. It is unlikely that such a theory will arise any time soon: it is incompatible with the emphasis on distributed processing now popular in cognitive science. The sequential view is only a theoretical option.

Figure 8 -- Parallel processing of a multimodal utterance.

In the parallel model, modes are processed at the same time (see Figure 8). Paivio's (1986) *dual-coding approach* can be seen as a forerunner of this view of multimodal language processing, even though his discussion of the non-verbal channel does not deal explicitly with bodily expression, but rather with imagistic objects, such as pictures. Drawing on extensive experiments showing the impact on speech of pictures presented simultaneously with that speech, Paivio concludes that information from the non-verbal channel is integrated with that of the verbal channel while the verbal channel is being processed. Verbal and non-verbal processing are done by two separate cognitive subsystems in

communication with each other¹⁰. Özyürek and colleagues (2007) defend the parallel view of speech/gesture integration explicitly, supporting it by their finding that neurological activity during speech and gesture comprehension shares latency, amplitude, and topographical distribution in the brain. As cognitive science's aforementioned emphasis on distributed processing suggests, an increasing number of theories on comprehension of contextualized speech – involving semantic meaning and contextual information, including other modalities – depend upon a parallel view of verbal/non-verbal processing (see e.g. Dąbrowska 2004, Wilson & Sperber 2004, Recanati 2004).

In my discussion of theories all of which hold that speech and gesture are produced for the purpose of communicating thought, I concluded that all of them presuppose implicitly that thought precedes its material externalization. Having discussed models of multimodal language comprehension, I conclude that the inverse presupposition also holds: i.e., the material externalization of a speaker's thought – a multimodal expression – precedes the communicative intention (a shared thought) that the listener then reconstructs on its basis (see Figure 9).

Figure 9 – Externalization of a speaker's thought processes precedes reconstruction of that thought by the listener.

Speech/gesture as thought and communication

Any framework according to which speech/gesture is produced for purposes of communicating thought gets into trouble by adhering implicitly to two presuppositions: (i) the material carrier of meaning (i.e., bodily expression) is fundamentally different from the meaning it carries from one mind to another, so that a process of coding and decoding must take place; (ii) because it is possible to analyze expressive modes in isolation, the speaker's and listener's cognitive systems must treat

¹⁰ McNeill (2005) comments that Paivio's separate-but-communicating-channels picture is similar to Kita and Özyürek's (2003) cognitive architecture of speech/gesture production. As the coding-decoding model of communication predicts, Paivio's theory of comprehension is exactly its inverse.

different modes as different channels with separate processing streams (sequentially or in parallel)¹¹. The picture I wish to paint calls these two presuppositions into question, along with the consequent cognitive architecture of multimodal communicative understanding. Elaborating on the views of McNeill (1992, 2005) presented earlier, I argue for a framework in which (i) meaning coincides with its material carrier, while (ii) keeping in mind that this material carrier far exceeds the boundaries of the verbal mode. This leads me to argue that linguistic understanding does not consist of a temporally extended process in which meaning is inferred from a perceived multimodal utterance: rather, it happens immediately and all at once.

I have shown that, in McNeill's framework, speech and gesture are not an external realization of thought; rather, the mental mode emerges simultaneously with multimodal expression. These modes drive as well as are driven by each other dialectically, in a process of 'thinking-for-speaking'. McNeill extends Vygotsky's material carrier from word meaning to the expressive power of the full body: i.e., not only do words and meanings coincide; expressive bodies and their meanings do as well. When one express oneself, one's thoughts are present in the world and readily available to others at the same time that they are unfolding to oneself. How can one make the idea more transparent that one's conversational participants' thoughts are formulated at the same time as one's own are: i.e., how does 'self' all-at-onceness extend to interactional all-at-onceness?

In my discussion of the all-at-onceness of expressive subjectivity, I introduced Vygotsky's (1962) distinction between analysis-in-elements and analysis-in-units. To presuppose that, in the language user's mind, different linguistic modes are realized as different cognitive channels is to take the results of the investigator's analysis-in-elements as faithful representation of reality. A problem immediately arises after such an analysis: what is the explanation of existing correlations between the elements? Theoretically, the synchronized nature of multimodal expressive action elicits a huge amount of additional cognitive processing. An analysis in units regards the full expressive body as a basic entity, driven by one and the same process of meaning constitution – thereby removing the need for an external synchronization task. The distinction between analysis-in-elements and analysis-in-units can be extrapolated to the behavior of participants in linguistic interaction. Assume that the interactants have been brought up in roughly the same circles and that they speak the same language (broadly conceived): i.e., they share a means of thinking and expressing that is natural to them. If one considers these conversational participants to be autonomous elements, then correlations between the elements – such as sharing thoughts or expressive behavior – present themselves as non-default states that need to be reached somehow. Interactional understanding and adaptation are the result of a cognitive task of mental and behavioral synchronization.

¹¹ Even Goodwin & Duranti (1992) – though they defend broadening the units of linguistic analysis – divide up the perceptual field into a focal event and a context. The focal event need not necessarily be the verbal utterance; but, when e.g. a gesture is the focal event, the verbal moves to context – thus leaving the 'separate channels' problem untouched.

If, however, one considers conversational participants to be aspects of an interactional unit – i.e., a dyad or group rather than a collection of autonomous agents – one may see how their whole takes on a form of autonomy that cannot be explained by merely referring to their parts (see e.g. De Jaegher & Di Paolo 2007). Consider a speaker's arm swing and pointing gesture: in the production of this action, the speaker drags the listener's gaze towards the object or in the direction pointed at, as if speaker and listener were playing a game of pulling-the-rope, in which the 'rope' is a shared meaning. This exemplifies well Merleau-Ponty's (2007) phenomenological concept of *intercorporeity*, in which, in social or linguistic interaction, the participants' bodies become aspects of a shared dynamic system: the participants share a 'body', because they have been formed the same way: i.e., they speak the same verbal and non-verbal language; and because they find themselves in a concrete situation together: i.e., they have a shared framework to which to make reference. Instead of both the speaker's and listener's cognitive systems needing to put in a separate effort, rather it is the interaction that is the (non-preceding) cause of both productive and receptive meaning constitution. People need not exchange the conversational roles of speaker and listener each time they take a conversational turn – as an investigation of the verbal medium on its own would suggest. A full multimodal analysis shows that, when people are engaged in conversation, they are, at all times, immediately attuned to each other's meaningful behavior. It is better to say that they think, speak, and listen along. Scientific and everyday findings again offer support:

- The covert linguistic behavior I mentioned in my discussion of expressive subjectivity – i.e., that even when people think they are thinking in silence, their vocal organs are active – extends to expressive intersubjectivity, so that this 'sub-vocalization' is also present when people listen to someone else speaking (McGuigan 1979). This suggests that one's understanding of another person's words emerges from 'speaking' along non-vocally with the other person.
- Everyone is acquainted with the phenomena whereby someone finishes another person's sentence or helps them look for a word they cannot find. Because, during listening, one thinks along with the other instead of decoding the cues one is sent, one knows what direction thoughts and sentences are headed in even before they have been expressed.
- The phenomenon of interactional *entrainment* is widely documented. Entrainment is the convergence of communicative behavior across speakers. It occurs either in overlap: both parties engage in some behavior simultaneously; or sequentially: one party repeats an action previously produced by the other (Parrill & Kimbara 2006)¹². Verbal entrainment includes mutual adaptation of accent and intonation, or adoption of word choice and

¹² 'Mimicry', 'mirroring', and 'synchrony' are terms also – and even more frequently – used. I prefer 'entrainment' because it evades the image of one conversational participant mimicking, mirroring, or 'getting in sync' with the other, and suggests instead that both participants are carried away by the linguistic event: "motor mimicry is not an individual phenomenon. It depends on, is shaped by, and in turn influences a particular moment in dialogue, illustrating the fine-tuned reciprocity that is the essence of the micro-social context" (Bavelas 2007:136).

sentence structure. Non-verbal entrainment extends to facial expression, gesture, and posture. Beyond what is perceptible simply by sight or hearing, entrainment e.g. includes adaptation of the listener's breathing rate to that of the speaker's (McFarland 2001). All these examples of entrainment – many more can be found in e.g. (Ellis & Beattie 1986) or, under the term 'appropriation', in (McNeill 2005) – take a form in which the onlooker is unsure who is leading and who is following. They are difficult to explain in a cognitive-processing framework. If one extends the term to subsume non-identical behavior matched by a common intention, then the phenomenon becomes ubiquitous – as the interactional framework predicts. Consider e.g. speaker and listener facing each other, so that both their heads are turned in a different direction relative to the body: one looking left, the other looking right; yet, in absolute terms, they are turned in the same direction.

The concept of intercorporeity has drastic consequences for a model of linguistic interaction. If the action/reaction, active/passive, input/output, sender/receiver, coding/decoding picture of language production and understanding can be represented by an image of throwing a ball back and forth, then the dialectic model I propose is represented by an image of playing musical instruments together or dancing together to a rhythm: one does not wait for the beat or for a sign from the other to move one's body this way or that; self and other form a dynamic system that is led by the music. One's dance steps are formed by anticipating where the music will go next and one's partner along with it. Likewise in conversation, one is fully submerged and attuned to the other's behavior and the contents of linguistic interaction. This is why, in the interpersonal realm, linguistic expression has an all-at-onceness. All-at-onceness does not mean that one's understanding of the other's expression happens at one point in time: a finger snap and then it is past; rather, self and other move through time together, as a dyadic unity; during this ongoing interaction, expressions and thoughts unfold simultaneously. The speaker's thoughts are revealed to her at the same time they are revealed to the listener (see Figure 10)^{13,14}.

¹³ Does my claim that expressive subjectivity and intersubjectivity are characterized by all-at-onceness remain only at the personal level, or does it have implications for a functional, subpersonal explanation of face-to-face interaction in terms of neural activity? This is a complicated matter. Contemporary phenomenology – which provides the broader framework for my investigation – as a discipline mainly analyzes phenomena from a first-person point of view, at the same time calling into question the very possibility of the neural, functional explanation that cognitive science searches for. While one cannot deny the importance of brain activity for experience, the argument is that talk of experience is meaningful only at the level of the embodied, situated perceiver. Neural activity has a supportive function, but the constitution of meaning happens through interaction. See Gallagher and Zahavi (2008) for an elaborate explanation of how present-day phenomenology relates to cognitive science.

¹⁴ At this point, it should be clear why I stated at the beginning of the paper that the distinction between expressive subjectivity and intersubjectivity is artificial. Listening involves expressing oneself and *vice versa*. The connections between intra- and interpersonal realms have been explored by Vygotsky (1962); Scheler (2008); Merleau-Ponty (2007); and, more recently, by Gallagher (2005) and Gallagher and Zahavi (2008).

Figure 10 – The cognitive-processing model versus the all-at-onceness of linguistic interaction: is a conversation more like throwing and catching a ball back and forth (1) or more like playing musical instruments together (2)? (S1 = Speaker 1, S2 = Speaker 2).

CONCLUDING REMARKS

I may summarize the paper by asking two questions: (i) how does thought relate to multimodal linguistic behavior in the act of expression and understanding, and (ii) how do these expressive modes cognitively interrelate? I first looked at theories in which thought precedes its material externalization: the purpose of expressing oneself is the communication of thought. In these theories, speech and gesture occupy two cognitive channels or constitute two production processes, whose organization is either carefully planned in advance or reached along the way by communication between channels. In McNeill's alternative framework, thought is formulated at the same time it is expressed, in a process of 'thinking-for-speaking'. Multimodal expressions are not there for the purpose of communication – although this may be a result; rather they *are* thought in action. Speech and gesture form an expressive unit together, cognitively emerging and unfolding at the same time as thought. I have called this the all-at-onceness of multimodal expression. Turning the attention to "inward" understanding in multimodal interaction, I first looked at theories in which thought is distinct from any multimodal utterance from which it is inferred. Different expressive modalities occupy different cognitive channels; yet their information needs to be integrated into a meaningful whole. This can be achieved

by processing modes sequentially or in parallel. My alternative view – elaborating on McNeill’s framework – extends the all-at-onceness of multimodal expression to multimodal interaction. The meaning/thought that coincided with embodied expression now coincides with shared embodiment: i.e., Merleau-Ponty’s *intercorporeity*. Because self and other speak the same language (in a broad sense) and share a concrete situation, their bodies move as one: speaking, listening, and thinking along with each other.

REFERENCES

- Alibali, M., Heath D. & Myers H. (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.
- Andrn, M. (*this volume*). Gesture and object manipulation. *Journal of Cognitive Semiotics*, **4**(1): 136-162.
- Argyle, M. & Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.
- Bavelas, J. (2007) Face-to-face dialogue as a micro-social context: The example of motor mimicry. In Duncan, S., Cassell, J. & Levy, E. (eds.). *Gesture and the Dynamic Dimension of Language*. Amsterdam: John Benjamins, 127-146.
- Bavelas, J. B., Coates, L. & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, **52**: 566-580.
- Bavelas, J. B. & Gerwing, J. (2007). Conversational hand gestures and facial displays in face-to-face dialogue. In Fiedler, K. (ed.), *Social Communication*. New York: Psychology Press, 283-308.
- Butterworth, B. (1980). Evidence from pauses in speech. In Butterworth, B. (ed.), *Speech and Talk*. London: Academic Press, 155-176.
- Butterworth, B. & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, **96**(1): 168-174.
- Cole, Jonathan (1998). *About Face*. Cambridge, MA.: MIT Press.
- Crowder, E.M. & Newman, D. (1993). Telling what they know: The role of gestures and language in children’s science explanations. *Pragmatics & Cognition*, **1**: 339-374.
- Dbrowska, E. (2004). *Language, Mind and Brain: Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh: Edinburgh University Press.
- De Jaegher, H. & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, **6**: 485-507.
- De Ruiter, J. P. (2007). Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture* **7**(1): 21-38.
- Drahota, A.; Costall, A. & Reddy, V. (2007). The vocal communication of different kinds of smile. *Speech Communication*, **50**(4): 278-287.
- Ellis, A. & Beattie, G. (1986). *The Psychology of Language and Communication*. London: Weidenfeld & Nicholson.

- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In Cooper, W.E. & Walker, C.T. (eds.), *Sentence Processing: Psycholinguistic Essays Presented to Merrill Garrett* (27-85). Hillsdale, NJ: Erlbaum.
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Clarendon.
- Gallagher, S. & Zahavi, D. (2008). *The Phenomenological Mind*. London: Routledge.
- Goldin-Meadow, S. (2005). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Harvard University Press.
- Goodwin, C. (2003). The body in action. In Coupland, J. & Gwyn, R. (eds.), *Discourse, the Body and Identity* (19–42). New York: Palgrave/Macmillan.
- Goodwin, C. (2006). Human sociality as mutual orientation in a rich interactive environment: Multimodal utterances and pointing in aphasia. In Enfield, N. & Levinson, S.C. (eds.), *Roots of Human Sociality* (96–125). London: Berg Press.
- Goodwin, C. (2007). Participation, stance, and affect in the organization of activities. *Discourse and Society*, **18**(1): 53–73.
- Goodwin, C. & Duranti, A. (1992). Rethinking context: An introduction. In Duranti, A. & Goodwin, C. (eds.) *Rethinking Context: Language as an Interactive Phenomenon* (1-42). Cambridge, UK: Cambridge University Press.
- Gullberg, M. (2003). Eye movements and gesture in human face-to-face interaction. In Hyönä, J., Radach, R. & Deubel, H. (eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (685-703). Amsterdam: Elsevier.
- Iverson, J. & Goldin-Meadow, S. (1997). What's communication got to do with it?: Gesture in congenitally blind children. *Developmental Psychology*, **33**: 453-467.
- Iverson, J. & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, **396**: 228.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In Key, M.R. (ed.), *The Relationship Between Verbal and Nonverbal Communication*. The Hague: Mouton, 207-227.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge, UK: Cambridge University Press.
- Kita, S. (2003). Interplay of gaze, hand, torso orientation, and language in pointing. In Kita, S. (ed.), *Pointing: Where Language, Culture, and Cognition Meet*. Mahwah, NJ: Erlbaum, 307-328.
- 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**: 16-32.
- Krauss, R.M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, **7**(2): 54-60.
- Krauss, R.M., Morrel-Samuels, P. & Colasante, C. (1991) Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, **61**: 743-754.

- Krauss, R.M., Chen, Y. & Gottesman, R.F. (2000). Lexical gestures and lexical access: A process model. In McNeill, D. (ed.), *Language and Gesture* (261-283). Cambridge, UK: Cambridge University Press.
- Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Loncke, F.T.; Campbell, J.; England, A.M. & Haley, T. (2006). Multimodality: A basis for augmentative and alternative communication – psycholinguistic, cognitive, and clinical/educational aspects. *Disability and Rehabilitation*, **28**(3): 169-174.
- McClave, E., Kim, H., Tamer, R. & Mileff, M. (2008). Linguistic movements of the head in Arabic, Bulgarian, Korean, and African American vernacular English. *Gesture* **7**(3): 343-390.
- McFarland, D.H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, **44**: 128-143.
- McGuigan, F.J. (1966). *Thinking: Studies of Covert Language Processes*. New York: Appleton-Century-Crofts.
- McGuigan, F.J. (1970). Covert oral behavior as a function of quality of handwriting. *The American Journal of Psychology*, **83**: 377-388.
- McGuigan, F.J. (1979). Electrical measurement of internal information processing during silent reading and listening. *Integrative Psychological and Behavioral Science*, **14**(4): 218-225.
- McNeill, D. (in preparation). *Notes on the Origin of Language: What Evolved, and How*.
- McNeill, D. (1992). *Hand and Mind: What Gesture Reveals About Thought*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.
- McNeill, D. & Duncan, S. (2000). Growth points in thinking-for-speaking. In McNeill, D. (ed.), *Language and Gesture* (141-161). Cambridge, UK: Cambridge University Press.
- Mehrabian, A. & Ferris, S.R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, **31**(3): 248-252.
- Merleau-Ponty, M. (2007). *Phenomenology of Perception*. London: Routledge.
- Okrent, A. (2002). A modality-free notion of gesture and how it can help us with the morpheme vs. gesture question in sign language linguistics. In Meier, R.P., Cormier, K. & Quinto-Pozos, D. (eds.), *Modality and Structure in Signed and Spoken Language* (175-198). Cambridge, UK: Cambridge University Press.
- 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): 605-616.
- Paivio, A. (1986). *Mental Representations. A Dual Coding Approach*. New York: Oxford University Press.
- Parrill, F. & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on event representation. *Journal of Nonverbal Behavior*, **30**(4): 157-166.

- Ramachandran, V.S. & Blakeslee, S. (1998). *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: William Morrow.
- Recanati, F. (2004). *Literal Meaning*. Cambridge, UK: Cambridge University Press.
- Rimé, B. & Schiaratura, L. (1991). Gesture and speech. In Feldman, R. & Rimé, B. (eds.), *Fundamentals of Nonverbal Behavior*. Cambridge, UK: Cambridge University Press.
- Ryle, G. (1978). *The Concept of Mind*. Harmondsworth: Penguin.
- Scheler, M. (2008). *The Nature of Sympathy*. New Brunswick, NJ: Transaction.
- Slobin, D. (1987). Thinking for speaking. In Aske, J., Beery, N., Michaelis, L. & Filip, H. (eds.), *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society, 435-445.
- Streeck, J. (2003). The body taken for granted: Lingering dualism in research on social interaction. In Glenn, P.J.; LeBaron, C.D. & Mandelbaum, J. (eds.), *Studies in Language and Social Interaction*. Mahwah, NJ: Erlbaum.
- Vygotsky, L. (1962). *Thought and Language*. London: Chapman & Hall.
- Wilson, D. & Sperber, D. (2004). Relevance Theory. In Horn, L.R. & Ward, G. (eds.), *The Handbook of Pragmatics*. Malden, MA: Blackwell.