The Brain Behind the Response: Insights Into Turn-taking in Conversation From Neuroimaging

Sara Bögels & Stephen C. Levinson

Published online: 29 Dec 2016.

Submit your article to this journal ☐

Article views: 1307

View related articles ☐

View Crossmark data ☐

Citing articles: 15 View citing articles ☐

Routledge
Taylor & Francis Group

The Brain Behind the Response: Insights Into Turn-taking in Conversation From Neuroimaging

Sara Bögels[a] and Stephen C. Levinson[a,b]

[a]Language and Cognition Department, Max Planck Institute for Psycholinguistics, The Netherlands; [b]Donders Institute for Brain, Cognition and Behaviour, The Netherlands

ABSTRACT

This article reviews the prospects for the cross-fertilization of conversation analytic (CA) and neurocognitive studies of conversation, focusing on turn taking. Although conversation is the primary ecological niche for language use, relatively little brain research has focused on interactive language use, partly due to the challenges of using brain-imaging methods that are controlled enough to perform sound experiments but still reflect the rich and spontaneous nature of conversation. Recently, though, brain researchers have started to investigate conversational phenomena—for example, by using "overhearer" or controlled interaction paradigms. We review neuroimaging studies related to turn-taking and sequence organization, phenomena historically described by CA. These studies, for example, show early action recognition and immediate planning of responses midway during an incoming turn. The review discusses studies with an eye to a fruitful interchange between CA and neuroimaging research on conversation and an indication of how these disciplines can benefit from each other.

In Conversation Analysis (CA), phenomena occurring in spontaneous conversation are traditionally studied in a qualitative manner. CA researchers are not primarily concerned with what happens in the minds of conversationalists but rather with the evidenced behavior. As Sacks (1992) put it:

> When people start to analyze social phenomena, if it looks like things occur with the sort of immediacy we find in some of these exchanges, then you figure that they couldn't have thought that fast. I want to suggest that you have to forget that completely. Don't worry about how fast they're thinking. First of all, don't worry about whether they're "thinking." Just try to come to terms with how it is that the thing comes off. (p. 11)

Thus, in the study of conversation, this approach attempts to understand and explain the social processes at work in conversation, using the "off-line" data coming from recordings or transcripts of conversation as sources of information and insight. In contrast, cognitive and neuroimaging approaches use "online" brain measures to study, in real time, the underlying cognitive processes enabling people to understand and produce language. It is our contention that each approach can illuminate the other: Results from off-line data can challenge investigations into the underlying cognitive processes, and the study of cognitive processes can show how the off-line data have the shape they do under the processing constraints inherent in participants.

Despite Sacks's call not to worry about cognitive processes, CA tries to take the participant's perspective, and often this involves a moment-by-moment analysis of what is unfolding—a nice example is Schegloff's analysis of competing first starts in the turn-taking system, where overlap may be resolved on a syllable-by-syllable basis (Schegloff, 2000). The seminal turn-taking paper (Sacks,

CONTACT Sara Bögels ✉ sara.bogels@mpi.nl 🖃 Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, 6525 XD Nijmegen, The Netherlands.

Color versions of one or more of the figures in this article can be found online at www.tandfonline.com/HRLS.

Schegloff, & Jefferson, 1974) also makes the point that the nature of the system obliges any intending speaker to "analyse each utterance across its delivery" (p. 727). It is this kind of analysis that can be pursued in cognitive neuroimaging by online inspection of cognitive processes even where there is no or only slender external behavioral correlate.

In this review article, we will describe and discuss neuroimaging research that bears upon the processes underlying turn-taking in conversation and their timing. The review thus includes research using a neuroimaging method that measures activations in the brain (e.g., EEG, MEG, fMRI). We will sometimes cite or discuss behavioral experimental research if it is a direct predecessor of the neuroimaging research or if it fills a gap in which no neuroimaging research has been done yet, to our knowledge. Furthermore, although we occasionally direct the interested reader to neuroimaging studies on other conversational topics, a lengthy discussion of them is beyond the scope of this review. We will focus here upon the investigation of turn-taking, which was foundational to the birth of the systematic study of conversation (e.g., Sacks et al., 1974) and which poses interesting cognitive processing problems that may be addressed through neuroimaging methods. The goal of this review then is to convey both how neuroimaging research can make use of CA insights in this domain and how CA can be informed by results from neuroimaging research, issues we will return to at the end. Before starting our literature review, we give a brief historical background of psycholinguistic neuroimaging studies to show why it has taken until quite recently for this field to reorient toward conversation.

Psycholinguistics and neuroimaging

Historically, psycholinguists have had little interest in studying conversation or dialogue because they have been focused on individuals' cognitive processes and indeed by tradition have studied language comprehension and production separately, often in different departments. They have attempted to study these processes as they occur step by step—for example, exploring in great detail how in comprehension we get from the acoustics to a word (e.g., McQueen, Norris, & Cutler, 1994; Wise et al., 1991) or in production how we get from an abstract representation of a word to its articulation (e.g., Indefrey & Levelt, 2004; Levelt, 1999). Although the scope of these studies has been highly restricted, the findings remain quite relevant for those trying to understand how language is actually used—they establish boundary conditions for human performance that we can expect to operate in natural conversation just as they do in the lab.

Recently, however, the tide has been turning in the psychology of language. The reasons for this are various. First, there has been increasing appreciation of the relevance of the natural ecology for language for models of language processing (e.g., Brennan, Galati, & Kuhlen, 2010; De Ruiter, Mitterer, & Enfield, 2006; Garrod & Pickering, 2015; Hagoort & Levinson, 2014; Schilbach et al., 2013; Schober & Brennan, 2003; Willems, 2015). Another reason is that, in our opinion, the scholarly tussles between, for example, different models of lexical retrieval are becoming less and less fruitful as the arguments have become long rehearsed and the deciding data more and more wafer thin. And the third reason is probably that CA and neighboring disciplines, including the corpus analysts, have increasingly precise and interesting things to say, which lend themselves to experimental verification. There is thus a general feeling that perhaps now is the time to face the reality of the core uses of language in interaction and its psychological implications.

One inhibiting factor for the psycholinguists is that they have developed many tools for probing cognition in real time as it is being deployed in the mind and the brain, but for the most part, these methods do not extend easily to the study of conversation (except perhaps for the eye-tracking methodology, e.g., Brown-Schmidt & Tanenhaus, 2008; Casillas & Frank, 2017; Holler & Kendrick, 2015). Psycholinguists are actually caught between a rock and a hard place, between the need for experimental control on the one hand and the desire for "ecological validity" (social reality) on the other (see, e.g., Butterworth, 1980 and Danks, 1977 for early descriptions of this problem in language production). Experimental control entails that, in order to be able to draw a causal conclusion about the effect of a

Box 1. Overview of Neuroimaging Methods.

In functional magnetic resonance imaging (fMRI), participants lie in a large machine and perform a certain task while their brain is "scanned." Participants' blood flow to different brain areas is measured (indirectly), which is related to the activation level of these areas while participants perform a task. However, the measure is relatively slow, since the blood flow lags a few seconds behind the actual activation of the area. In contrast, the methods of electroencephalography (EEG) and magnetoencephalography (MEG) show very precisely *when* certain brain activity occurs. However, the exact location of the effects is more difficult to pinpoint with these methods (albeit a bit better with MEG than EEG). Two commonly used methods for analyzing EEG/MEG data are the following. In event-related potentials (ERPs), EEG signals are (as mentioned in the text) averaged over a large number of trials, leading to a number of peaks and troughs that can be compared between conditions. For example, the so-called N400 component (a negative wave that is largest around 400 ms after the start of a word; Kutas & Hillyard, 1980) has been found to index the expectedness of words (see Figure 3 for an example). A second method of analysis concerns the time frequency of oscillatory activity in the brain (here referred to as "frequency analysis"). In this case, one analyzes the way a stimulus modulates the power (amplitude) of ongoing background oscillatory rhythms ("brain waves") at different frequencies in the brains of individuals at a particular moment in time. This is usually presented visually like a spectrogram, representing the power at different frequencies and time points (see Figure 4 for an example). For more information on these methods, see, for example, handbooks edited by Cacioppo, Tassinary, and Berntson (2007) and by Luck and Kappenman (2011).

certain variable, just this variable should change while all other variables should be held constant. Since conversation, or more generally interactive language use, is by definition spontaneous, constraining and controlling its variables inevitably changes the nature of the object of study. Within neuroimaging research, some recent creative compromises for this problem have started to arise (some of them preceded by similar methods in behavioral research). One compromise is to use overhearing paradigms in which participants listen to (parts of) prerecorded conversations, either spontaneous (Bögels, Kendrick, & Levinson, 2015; Magyari, Bastiaansen, De Ruiter, & Levinson, 2014; see De Ruiter et al., 2006 in behavioral research) or reenacted (e.g., Gisladottir, Chwilla, & Levinson, 2015). The assumption here is that the cognitive processes of overhearers are similar to those of participants in a conversation while they are listening (but see Schilbach et al., 2013; Schober & Clark, 1989). Another compromise concerns a live interaction between a participant and someone else, asking questions, for example, which is nevertheless controlled at specific points (Bögels, Barr, Garrod, & Kessler, 2015; Bögels, Magyari, & Levinson, 2015; and, e.g., Metzing & Brennan, 2003 in behavioral research). Although such a format is interactive rather than strictly conversational (e.g., see Heritage & Clayman, 2010, p. 12), it draws on naturally occurring speech exchange systems that have many of the essential properties of conversation.

Another reason why it is difficult to perform experimental and especially neuroimaging research on conversation is that many instances ("trials") of the same phenomenon are needed in one experiment to rule out the possibility that some of the effects are caused by noise or coincidence. Several similar trials that might be relatively rare in spontaneous conversation may render these experiments less ecologically valid. Take the example of EEG (see Box 1). The EEG signal, the weak electrical potential caused by the simultaneous firing of thousands of neurons that are spatially aligned, gives us potentially a very fine-grained temporal resolution for cognitive events and thus recommends itself for the study of the fast pace of interactive language use. But this electrical potential has to be detected against the background of the constant chatter of other neurons. We can do this by locking the EEG of many participants to several similar stimulus events and then letting the unrelated background chatter (which is not time locked and random) cancel itself out. But to do this we need 20 or more participants listening to exactly the same event in exactly the same context, repeated about 30 times for a similar event—hence the need for repetitive testing simply in order to derive a clear signal.

A last problem for neuroimaging research on interactive language use is the fact that most neuroimaging methods do not deal very well with speech production. Any motor movement causes so-called artifacts in the measurements, but these are especially strong for speech since the relevant muscles are close to the brain. Techniques are beginning to appear that "filter out" these artifacts (e.g., Poser, Versluis, Hoogduin, & Norris, 2006; Vos et al., 2010), but it is still difficult to separate brain signal from motion artifacts. An alternative strategy is to analyze only those parts of the data coinciding with listening and/or planning to respond rather than actually speaking (see, e.g., Bögels, Magyari et al., 2015).

Despite these challenges, a number of studies have endeavored to investigate conversational phenomena using neuroimaging methods (see Box 1 for a description of the most commonly used

methods, with their strengths and weaknesses). We now turn to give a brief background to the study of turn-taking, starting with CA work, followed by a description of two cognitive models that could explain the turn-taking system.

Turn-taking background

From the CA literature, most prominently the seminal paper by Sacks et al. (1974), we know that gaps and overlaps between two turns in conversation are minimized, so that most of them are small. Subsequent quantitative studies have indeed shown that the most frequent (or modal) gap duration is about 200 milliseconds (e.g., Heldner & Edlund, 2010; Levinson & Torreira, 2015) and is similarly short in different languages (Stivers et al., 2009). On the other hand, we know from classical psycholinguistic experiments (mostly picture-naming studies), that it takes minimally 600 milliseconds from seeing a picture to starting articulation of its name (e.g., Indefrey & Levelt, 2004). For sentences, this latency quickly rises to about 1,500 milliseconds on average (Griffin & Bock, 2000). These numbers strongly suggest that upcoming speakers do not have enough time to prepare and plan their turn after the previous turn has ended (Levinson, 2016). This raises the question of how upcoming speakers manage these small gaps in turn-taking: What are the underlying cognitive processes at play?

Sacks (1992, quoted previously) suggested that conversation analysts shouldn't primarily worry about the underlying cognitive processing. However, from the point of view of the cognitive sciences, conversational turn-taking is a remarkable demonstration of human multitasking abilities (Levinson, 2016). What is remarkable is that we are able to maintain such small gaps between turns while the language production system is so relatively slow, implying an overlap of comprehension and production toward the end of the incoming turn (see Figure 1). Since this involves multitasking using much of the same underlying language system, it is like juggling with one hand. Elsewhere, Sacks et al. (1974) have in fact pointed out that the turn-taking system must involve prediction or "projection" of the upcoming turn, or strictly the turn-constructional unit (TCU; Clayman, 2013; Sacks et al., 1974).[1] In addition they noted that

> A self-selector … has the problem that his earliest start must begin with a unit-type's beginning—one which, given its projectability will need to reflect some degree of planning … though current turn is still in progress. (p. 719)

Thus from the beginning CA was aware of the cognitive prerequisites for the turn-taking system to work. In the meantime, techniques have developed that allow us to look at these processes directly. This is the promise of neuroimaging—it allows us to glimpse the unseen cognitive processes that lie behind behavior even when there is as yet no overt sign of it.

CA methods have emphasized the insights that come from how recipients treat the prior turn (the proof procedure of Sacks et al., 1974) and thus what in cognitive terms would be thought of as the "comprehension" side of turn-taking (but see, e.g., Goodwin, 1979), but from a processing perspective the striking processing bottleneck is on the production side. Let us therefore take the perspective of the upcoming speaker listening to an incoming turn or TCU. We can distinguish at least three different (but possibly overlapping) types of cognitive processes related to turn-taking. First, the upcoming speaker has to understand the current turn well enough to know what type of next turn is relevant (if one is), which minimally means recognizing or projecting the action that is performed by the incoming turn (e.g., Levinson, 2013). Note that in CA, action recognition (more properly, action ascription; Levinson, 2013) and turn-taking are analytically treated separately, but an approach focused on the underlying cognitive processes is bound to treat them as related (related but not identical: see Schegloff, 1992). Second, the upcoming speaker needs to start planning an appropriate next turn (e.g., of the appropriate action type). Third, she needs to make sure that the articulation of

---

[1]CA treats turns as made up of one or more TCUs, where by definition a TCU is a linguistic unit that can be heard as possibly complete—where an addressee does not take the opportunity to intervene here, the turn may continue, yielding a turn of two or more TCUs. This analysis captures the contingent, collaborative nature of the unit that emerges as a turn. In many of the following experiments the distinction between turn and TCU is experimentally finessed, but it is important to bear in mind.
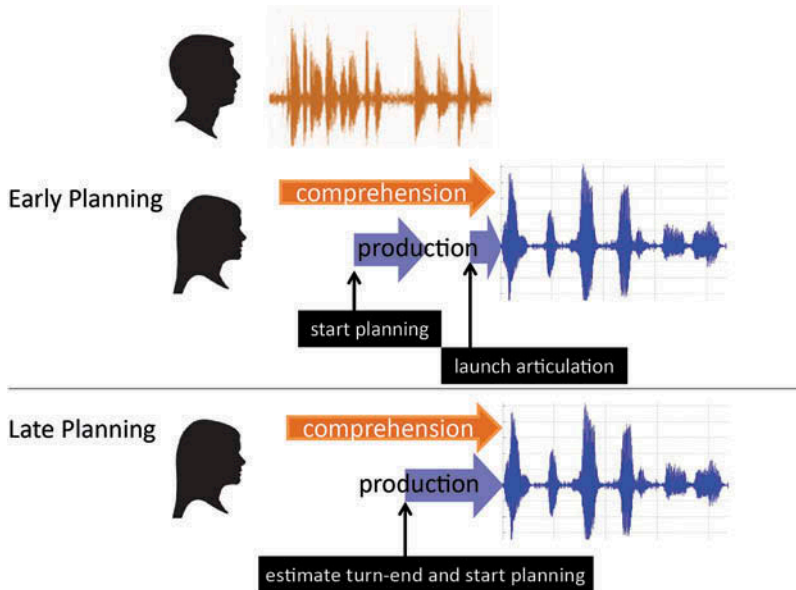
Figure 1. Illustration of the early planning and the late planning model for the timing of cognitive turn-taking processes.

her turn occurs at the right time—that is, without too much overlap or too much of a gap, by either projecting or recognizing the imminent end of a TCU. As will be explained in more detail, different hypotheses have been put forward about the timing of these processes, leading to roughly two, partly opposing, "models" of underlying processes related to turn-taking (see Figure 1). A first option (Early Planning) is that upcoming speakers start planning their turn as soon as possible so that it is ready by the time the incoming possibly complete turn or TCU is projected to end, which also implies some kind of buffer or inhibition of the response plan as long as it is not articulated yet (see Levinson & Torreira, 2015). Alternatively (Late Planning), upcoming speakers might prefer to give their full attention to comprehension of the incoming turn for as long as possible, postponing production planning until the latest moment, based on projection of the completion of the incoming TCU, which still enables them to come in on time (e.g., De Ruiter et al., 2006).

In the following, we describe how the three processes mentioned previously (action recognition, response planning, and projection of possible completion of TCUs) have been investigated using neuroimaging methods and what the results suggest about the proposed models. Since the timing of the processes involved is crucial to distinguish between the models, studies using EEG, which has an excellent temporal resolution (see Box 1), are described most elaborately.

Recognizing actions

The first process we distinguished is action recognition, which is the first step that is necessary before an upcoming speaker can start designing her next turn. One EEG study looking at action recognition used an overhearer paradigm (Gisladottir et al., 2015). Participants were asked to listen to a number of two-turn interactions that were prerecorded by instructed speakers. This study makes use of the fact that linguistic form is underspecified for action; a turn with the same linguistic form can perform different actions in different contexts (e.g., Levinson, 2013; Schegloff, 1984). Target utterances (e.g., *I have a credit card*) in Dutch were recorded in a list, with unmarked intonation and embedded in three different sequential contexts (extracted from recordings of enactments of conversation), leading the target utterances to perform different actions (see Figure 2, top). The target utterance could perform an answer in response to a question (e.g., *How are you going to pay for the ticket?—I have a credit card*), a pre-offer
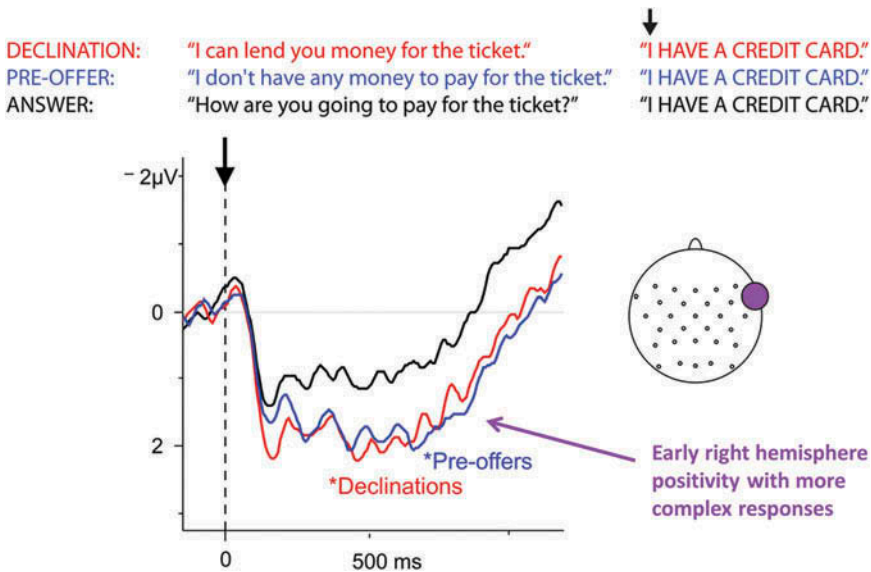
Figure 2. Sample stimuli and main results for Gisladottir et al. (2015).

(Schegloff, 2007) in response to the statement of a current problem (e.g., *I don't have any money to pay for the ticket—I have a credit card*), or a declination in response to an offer (e.g., *I can lend you money for the ticket—I have a credit card*). Each EEG participant heard many (126) such minidialogues in the experiment but heard the same target utterance only once. ERP analyses showed that declinations differed from answers from very early on, already around 200 milliseconds after onset of the target utterance, in the form of a frontal positivity for declinations (see Figure 2, red vs. black line). This early difference indicates that declinations are processed differently from answers from the first word or two, pointing to a very early start of action recognition (as briefly adumbrated in Sacks et al., 1974, p. 719). Moreover, examining specifically the final word or phrase (e.g., *credit card*), declinations and answers showed no differences. Thus, declinations apparently had been recognized fully before the final word. In time-frequency analyses on these same data (Gisladottir, Bögels, & Levinson, 2016) declinations also showed lower beta activity (12–20 Hz) than answers, starting already before onset of the target turn. This suggests that listeners especially attended to declinations, and this attention even started to rise before the declinations occurred. This could be due to the conditional relevance of the offer (Schegloff, 2007) coupled with important implications of the response (e.g., subsequent actions in case of an acceptance of the offer or the potentially negative implications of a dispreferred declination; Pomerantz & Heritage, 2013). In sum, listeners recognize declinations early, before the end of the utterance, and may even anticipate them before their onset. Pre-offers showed a frontal positivity relative to answers as well, which started around 400 milliseconds after target utterance onset (see Figure 2, blue vs. black line). Furthermore, a late posterior negativity was found to the last word of the pre-offer, and no (anticipatory) oscillatory differences were found relative to answers. This pattern of results suggests that listeners are less likely to anticipate pre-offers, that they start recognizing them a bit later than declinations, and that their recognition is still ongoing at the last word. This makes sense given the sequential structure of the dialogues. After a statement such as *I don't have any money to pay for the ticket*, no particular response is conditionally relevant (Kendrick & Drew, 2016), and a pre-offer is not specifically predictable. Thus, judging from this study, listeners can recognize actions early, but the precise speed at which they do so depends on the sequential context.

A recent EEG study also used an overhearer paradigm to study the anticipation of different types of actions, namely preferred and dispreferred responses to prior actions (Bögels, Kendrick

et al., 2015). CA and quantitative research has shown that dispreferred responses are associated with longer gaps (Kendrick & Torreira, 2015; Pomerantz & Heritage, 2013), and earlier behavioral experiments manipulating gap length have shown listeners' sensitivity to the length of the gap (F. Roberts, Francis, & Morgan, 2006; F. Roberts, Margutti, & Takano, 2011). The EEG study built on these results and used recorded stimuli that were extracted from a corpus of spoken Dutch telephone conversations (CGN, Oostdijk, 2000). Participants heard two-turn interactions consisting of an invitation, request, proposal, or offer followed by a preferred or dispreferred response. To ensure experimental control, the actual responses in the corpus were replaced by plain *yes* (preferred) or *no* (dispreferred) responses extracted from a different position in the corpus than the initiating actions. The crucial further manipulation was the duration of the gap between the two turns; participants heard either a short gap of 300 millisenconds or a long gap of 1,000 milliseconds. ERP analyses yielded an N400 effect (related to the expectedness or "surprisal value" of a word, see Box 1; Kutas & Hillyard, 1980) for *no* responses (relative to *yes* responses) after a short gap, showing that *no* responses were less expected. However, after a long gap, the difference between *yes* and *no* responses disappeared, indicating that *yes* responses were no longer more expected than *no* responses (see Figure 3). That is, in general the overhearers expected a preferred response, in line with the preference for agreement (Sacks, 1987), but as the gap became longer, this expectation changed toward a dispreferred response. Such a moment-by-moment change in expectations is likely to occur in the speaker of the initiating action as she awaits response.

Thus, parallel to the anticipatory attention effect found in the beta frequency in the study described previously (Gisladottir et al., 2016), this study provides experimental evidence that interlocutors anticipate actions even before they are delivered (consistent with observations made by, e.g., Sacks et al., 1974). This could be seen as an extremely early form of action projection, at least some early anticipation of the valence of the upcoming response, which could be the basis for planning an appropriate next turn to that upcoming action (see next section).

A few other neuroimaging studies have looked at action recognition, but these were based on speech act theory (Austin, 1975; Searle, 1969) rather than the CA tradition of analysis. These studies contrasted indirect replies or requests (Levinson, 1983), the meaning of which goes beyond the literal meaning of the words, with utterances that should be interpreted literally. An EEG study used written stories ending in an indirect or a literal request with the same linguistic form (Coulson & Lovett, 2010). Participants showed differential ERP responses while they read indirect as opposed to literal requests, from the second word onwards, but also on the fifth word that was crucial for understanding the request. Thus, action recognition appears to start early but is not finished before the crucial word in the utterance. This is consistent with the pre-offer results in the first study described (Gisladottir et al., 2015), since requests are initiating actions and are therefore not very predictable from the previous turn. One other EEG study used an overhearer paradigm to compare responses to requests for information and responses to offers in very controlled one-word turns, leading to differences as
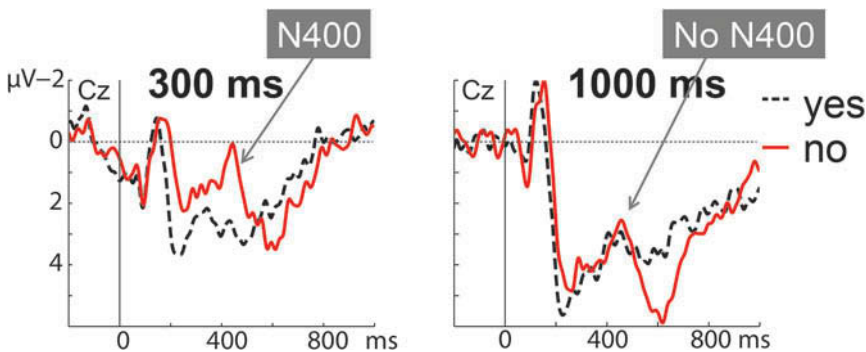
Figure 3. Results from Bögels et al. (2015) for *yes* and *no* responses after a 300-ms gap (left) and a 1,000-ms gap (right).

early as 120 milliseconds after response onset (Egorova, Shtyrov, & Pulvermüller, 2013). An MEG study with the same setup showed even earlier differences, around 50–90 milliseconds (Egorova, Pulvermüller, & Shtyrov, 2014).

In addition to the timing of action recognition, brain areas and circuits activated in this process have also been explored in some fMRI studies, mainly using overhearer paradigms (Bašnáková, van Berkum, Weber, & Hagoort, 2015; Bašnáková, Weber, Petersson, Van Berkum, & Hagoort, 2014; Egorova, Shtyrov, & Pulvermüller, 2016; van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). These studies contrasted direct and indirect actions and found that interpreting indirect actions activated theory of mind areas in the brain (related to thinking about other's intentions) more than direct actions. If the action required a bodily response (e.g., in requests for action), motor areas were activated (Egorova et al., 2016; van Ackeren et al., 2012). Moreover, areas related to empathy and emotion were more strongly activated for interpreting indirect replies in an interactive paradigm than in an overhearer paradigm (Bašnáková et al., 2015). However, since fMRI cannot show the exact speed of speech act recognition, these studies are not very informative about the timing of turn-taking processes as predicted by the two models (see Figure 1).

To sum up the timing results from the EEG research reviewed in this section, the experimental studies showed that listeners can recognize actions early on in the utterance and in some cases even anticipate them (in accordance with, e.g., Sacks et al., 1974). The more constraining the sequential context, the earlier the recognition process appears to start and finish. Note that these results allow for early planning of next turns (at least in the case of certain turns, such as responses) as hypothesized in the Early Planning model (Figure 1), but they still leave open the possibility that planning does not follow action recognition immediately and still starts late (Late Planning, Figure 1). This is the next topic of investigation.

Production planning in interaction

At the time of writing, only one brain-imaging study has directly addressed this problem (Bögels, Magyari et al., 2015). Participants answered questions in an interactive quiz paradigm. The experimenter interacted freely with them both before and during the experiment (e.g., by giving feedback), but unbeknownst to participants, the crucial questions were prerecorded by the experimenter. Thus a controlled but clearly interactive situation was created (see Figure 4, panel A). Crucially, the questions appeared in two different forms for different participants: Either the answer could be determined halfway through the question (e.g., *Which character, also called 007, appears in the famous movies?*), or the participant only had access to the answer after having processed the last word of the question (e.g., *Which character from the famous movies is also called 007?*).

ERP responses showed a large positivity around the point that participants could start retrieving the answer (while listening to *007* in the examples, see Figure 4, panel B, top). This positivity was localized to brain areas that have been related to language production (Indefrey & Levelt, 2004). Moreover, in a control experiment, in which participants heard the same questions but were instructed not to answer them but rather to remember them, the effect was much reduced and localized to different areas. In frequency analyses, a suppression of alpha band activity (see Figure 4, panel B, bottom) was found at the back of the scalp around the same time as the positivity—an effect entirely absent from the control study. This effect was interpreted as reflecting a switch in attention from purely comprehension of the incoming question to production planning processes. On this basis it was concluded that planning of the next turn appears to start as soon as possible.

A recent MEG study using a clearly nonconversational setting (Piai, Roelofs, Rommers, Dahlslätt, & Maris, 2015) also provides some supportive evidence for this idea. When speakers could start planning a nonword some time before they had to say it, in the intermediate interval beta band activity was found in frontal regions, reflecting response inhibition according to the authors. This
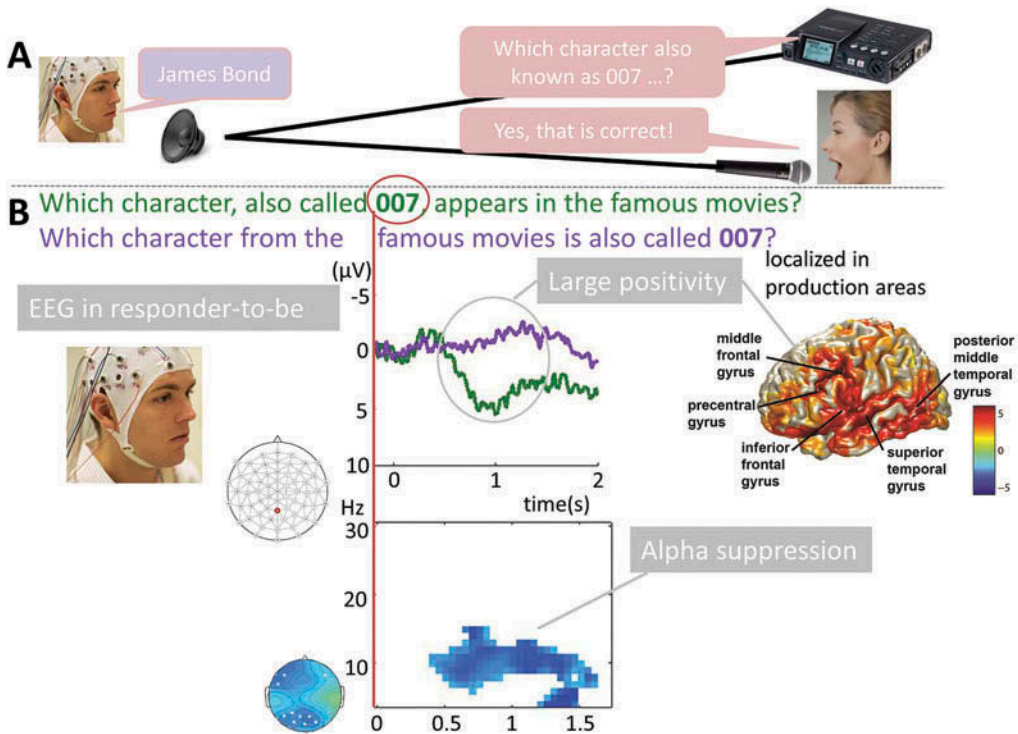
Figure 4. Setup (panel A) and results (panel B) for the ERP (top) and time-frequency domain (bottom) from Bögels, Magyari, and Levinson (2015).

supports the Early Planning model (Figure 1), since early preparation of a response implies that sometimes responses will need to be inhibited until the ongoing speaker seems about to finish.

Whereas no brain-imaging studies to date have challenged the idea that planning starts as soon as possible, it seems appropriate here to also briefly describe two non-neuroimaging studies that have investigated the timing of next-turn planning and reached a different conclusion. These studies used a dual task paradigm that puts stress on language production by requiring a simultaneous irrelevant task (Boiteau, Malone, Peters, & Almor, 2014; Sjerps & Meyer, 2015). Participants in Boiteau et al. (2014) conversed with a friend while tracking a target on a screen. Tracking task results were worst at the ends of the other speaker's turns and during production, thus indicating that production planning was consuming processing recourses toward the end of the incoming turn. In Sjerps and Meyer's (2015) study participants named a series of pictures while performing a finger-tapping task. Participants listened to recorded names of one row of pictures and subsequently named the other row themselves. Finger tapping only deteriorated within the last 500 milliseconds before the end of the recorded turn. Additional eye-tracking results showed that participants only started looking at their own pictures within a similar time frame. These two studies thus suggest that upcoming speakers start planning their next turn only at the very end of the incoming turn or TCU.

The latter two studies thus appear to support the Late Planning model, whereas the two neuroimaging studies that were described first support the Early Planning model (see Figure 1). These different results might be caused by a difference in methods. For example, the studies by Sjerps and Meyer (2015) and Boiteau et al. (2014) used dual tasking to see when the conversational processing was most intensive and thus most disruptive for the second task. The advantage of using neuroimaging methods in this respect is that no additional artificial task is needed to measure the dependent variable, and the absence of a distracting side task better approximates to ordinary talk. It

is also important to strike a balance between ecological validity and experimental control. An important aspect of interactive conversation, as pointed out by the CA literature (e.g., Schegloff, 2007), is that the content of the current turn is often contingent on the previous one, while this was not the case in the study of Sjerps and Meyer (2015). In Bögels, Magyari et al. (2015), contingent responses to questions were used, but the types of turns investigated were restricted to question–answer pairs. Boiteau et al.'s (2014) study had relatively high ecological validity since they used spontaneous conversation (albeit with a dual task). However, this came at a cost of no experimental control over when planning of the next turn could start during the incoming turn. It is then hard to interpret their finding of worse tracking task performance *toward the ends* of the other speaker's turns, since it is unclear when upcoming speakers had enough information to start planning. Another open question in connection with the previous section is how certain the upcoming speaker has to be about the action performed by the incoming turn before she starts planning her next turn. Perhaps anticipating an action with a certain probability is sometimes already enough to start planning a next turn.

Speaking on time

For the turn-taking system to run smoothly with minimal gaps and overlaps, not only will response preparation have to be finished in time, but also some reasonably accurate estimation of the likely end of the incoming TCU is required. The Early and Late Planning models described previously (see Figure 1), lead to different predictions in this respect as well. If the goal is to start production planning as soon as possible, it can be unrelated to an estimation of when the current TCU will end. However, estimation or identification of the possible completion of the current TCU still has to take place independently in order to launch one's turn at the right time (Early Planning model). On the other hand, if production planning starts as late as possible, estimations of how soon the incoming TCU will end and how long the production planning will take have to be matched against each other (Late Planning model).

The latter account has been advocated in some recent papers (De Ruiter et al., 2006; Magyari & De Ruiter, 2012). The idea is that recipients of a turn try to predict the upcoming content of that turn unit or TCU, which allows them, first, to be better able to plan their response and, second, to estimate when the turn unit ends and produce their response around that time. An EEG study that explored prediction of turn ends (Magyari et al., 2014, see below) was based on an earlier behavioral (non-brain-imaging) study using a button-press paradigm (De Ruiter et al., 2006). In the latter study, participants heard turns taken from a corpus of spontaneous conversation and were asked to press a button at the moment they thought the turn would end. Listeners were found to need lexicosyntactic information to do this accurately, but not necessarily pitch information. A follow-up behavioral study (Magyari & De Ruiter, 2012) showed that button presses were more accurate for more predictable turns. In the subsequent EEG study (Magyari et al., 2014) the same button-press paradigm was used while participants' EEG was measured. This study focused on frequency analyses and found that predictable turns yielded reduced beta activity from around 1250 milliseconds before the end of the turn. This was interpreted as an anticipation of the turn end from about midway in the turns. (Note that here as elsewhere in the experimental literature the distinction between turn and TCU has been conflated: The full turns were extracted from corpora, and they may have contained more than one TCU—it is the TCU or hearably complete unit that is actually relevant for anticipatory processing, and results could be fine-tuned by tracking these boundaries.)

A similar conclusion was drawn from two recent EEG studies (Jansen, Wesselmeier, De Ruiter, & Mueller, 2014; Wesselmeier, Jansen, & Müller, 2014) using a known ERP component, the Readiness Potential (RP) or Bereitschaftspotential (Kornhuber & Deecke, 1965), a slow negative-going cortical potential associated with the preparation of voluntary movements. Participants heard materials recorded by an instructed speaker, rather than being extracted from conversational corpus materials as used in the studies cited in the previous paragraph. The first study (Jansen et al., 2014) used both a

button-press and a spoken-response paradigm, in which participants were asked to give a short verbal answer (*yes* or *no*). In both tasks, an RP was found to start almost 1,200 milliseconds before the response. The second study (Wesselmeier et al., 2014) used only the button-press task but included stimuli with semantic and syntactic violations. The RP started already around 1,400 milliseconds before the button press for normal sentences but was disrupted for sentences with semantic or syntactic violations, starting only around 900 milliseconds before the button press. The authors conclude that semantic and syntactic information are both important for anticipating turn ends. In a later study using the same RP component (Wesselmeier & Müller, 2015), participants listened to recorded questions and had to answer more substantially than *yes* or *no*. In this case, response latencies were rather long, and the RP started only after the end of the turn, at a relatively fixed time from the response (irrespective of the cognitive complexity of the answer). The authors suggest the RP is related to speech preparation rather than intention to speak. This shows that results are dependent on the task (response versus button press). Taken together, the studies described in this section suggest that recipients of a turn (often upcoming speakers) attempt to anticipate the upcoming content of that turn and that this might help them to estimate when the turn (or, more accurately, TCU) ends.

The Early Planning model (see Figure 1) argues that recipients make use of possibly turn-final, mainly prosodic, cues to identify the possible completion of a TCU (see also, e.g., Duncan, 1972; Ford & Thompson, 1996; Schegloff, 1998; Wells & Macfarlane, 1998). Arguably, these cues may occur too late to cue the start of next-turn planning (De Ruiter et al., 2006; Magyari & De Ruiter, 2012). However, this could be resolved by starting to plan as soon as possible right up to the last stage before actual launch of articulation, which would be postponed until the TCU final cues can be identified. This last stage may take only a few hundred milliseconds, leading to the common short gap in turn-taking. De Ruiter et al. (2006) seemed to show that prosodic information was not important for anticipating ends of turns or TCUs, but their manipulation consisted of purely flattening the pitch. We know that the speech signal contains other prosodic cues such as final lengthening and intensity (see also Local & Walker, 2012). No neuroimaging studies have addressed this issue yet, but we here briefly describe a recent behavioral study very relevant to the models under discussion. Bögels and Torreira (2015) used a button-press task with manipulated questions. If a short question (*Are you a student?*) was spliced into a longer question (*Are you a student at Radboud University?*), participants were tricked into pressing the button after *student* in one-third of cases (i.e., they heard the embedded part as already potentially complete). On the other hand, when the short question was cut out of the long one, button presses were exceptionally late. This showed that late prosodic cues (lengthening of final syllable and pitch rise) are not only present in the signal (e.g., Local & Walker, 2012) but can actually be used by listeners to determine ends of TCUs. Several recent nonneurological measures similarly show behavior that indicates an upcoming response only around the ends of turns or TCUs, such as breathing for speaking (Torreira, Bögels, & Levinson, 2015), beginning manual movements in sign language (De Vos, Torreira, & Levinson, 2015), and eye gaze changes by a third nonaddressed party (Holler & Kendrick, 2015). Interestingly, in the last study described previously on the Readiness Potential in which participants answered recorded questions (Wesselmeier & Müller, 2015), the authors did not find evidence for an RP before the first syntactic completion point in a question, suggesting that "prosodic characteristics of the question could have signaled that the question was not completed at this point" (p. 151; see also Sacks et al., 1974). In sum, cues signaling the possible ends of TCUs appear to be important for recipients. Unfortunately, no brain-imaging studies have yet been performed to investigate such cues directly. This might be an interesting avenue for future research, since it could enable experimental investigation without an additional artificial task (e.g., button pressing).

Note that anticipation of content and recognition of cues to TCU completion are not mutually exclusive and might both be relevant for identifying the possible completion of a TCU. For example,

the higher the probability that the current TCU is coming to an end (based on anticipation), the more attentional resources could be allocated to identifying cues to TCU completion.

A different type of model regarding the coordination of turn-taking involves a cyclic pattern of probabilities to speak (M. Wilson & Wilson, 2005; T. P. Wilson & Zimmerman, 1986; see also Couper-Kuhlen, 1993), based on the idea of the cycle of speaker selection opportunities proposed in the seminal turn-taking paper (Sacks et al., 1974): first, other-speaker selection, then self-selection by other, then continuation by current speaker. This hypothesis was taken into the neurocognitive domain by proposing that oscillators in the brains of speaker and listener might become entrained, but counterphased with respect to each other, so providing a turn-taking metronome as it were (Garrod & Pickering, 2015; M. Wilson & Wilson, 2005). This entrainment, caused by the timing of syllables in the speech stream, continues for a bit after the speech ceases, accounting for the cyclic periodicity of silences. Some preliminary support for this idea comes from an MEG study showing oscillatory coupling between speech properties and different brain rhythms that is reset by edges in the speech envelope (Gross et al., 2013). However, this hypothesis could be tested more directly by correlating oscillatory brain activity between the brains of two interlocutors while they are having a conversation (see also Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012). A recent development in this direction in the neurosciences in general is so-called hyperscanning in which two people are "scanned" (e.g., with fMRI or EEG) at the same time while interacting with each other (Montague et al., 2002). A variant of this is consecutive measurement of brain activity from a "sender" and a subsequent "receiver," who hears and/or sees the recorded message from the sender at a later point in time, allowing the use of a single machine. These methods have been used to find coordinated activity between participants in different types of joint action such as hand movement imitations and guitar playing (e.g., Dumas, Nadel, Soussignan, Martinerie, & Garnero, 2010; Lindenberger, Li, Gruber, & Müller, 2009; see also Dumas, Chavez, Nadel, & Martinerie, 2012 for a newer modeling approach to hyperscanning), but only a few studies have looked at communication (for a review, see Kuhlen, Allefeld, Anders, & Haynes, 2015). A few fMRI studies have used consecutive scanning—that is, one-way communication—to look at communication by eye movements (Bilek et al., 2015), facial expressions (Anders, Heinzle, Weiskopf, Ethofer, & Haynes, 2011), gestures (Schippers, Roebroeck, Renken, Nanetti, & Keysers, 2010), and the telling of a story (Stephens, Silbert, & Hasson, 2010). Correlations between activation of the same brain networks in sender and receiver were found, often with a certain time lag from sender to receiver. In the storytelling study, areas in the listener were sometimes also activated before those of the speaker. Such anticipatory correlations were related to a better understanding of the story (Stephens et al., 2010).

In an EEG study on communication, also using consecutive scanning, speakers told a story while their EEG was measured (Kuhlen, Allefeld, & Haynes, 2012). Listeners subsequently saw a video recording of two different speakers telling different stories, overlaid on top of each other, but were instructed to attend to only one of the speakers. In this way, communication of content could be separated from effects of low-level sensory input. Listeners' EEG activity correlated with that of the speakers with a time lag of about 12.5 seconds. The authors interpret this as coordination on a very high-level "situation model." While these are interesting findings, they are rather hard to interpret with respect to coordination in turn-taking since coordination is here analyzed over a large episode, and communication is only one-way. Thus, no direct interaction between sender and receiver could take place, and they could not affect each other dynamically, as in conversation. Therefore, these studies do not directly say much about coordination in turn-taking, but they nevertheless make a case for neural entrainment between participants. A last, more directly relevant, study used a different brain-imaging method, namely fNIRS (functional near-infrared spectroscopy). This method is similar to fMRI (although less precise) but more portable and less prone to motion artifacts and thus suitable for production studies. fNIRS was used for hyperscanning two participants at the same time (in contrast to the previously described studies) during face-to-face conversation, back-to-back conversation, and monologue (Jiang et al., 2012). Simultaneous neural synchronization in the left

inferior frontal cortex (usually important for language production) was increased for face-to-face conversation relative to the other language modes. The authors argue that this synchronization was larger when participants performed gestures or turn-taking behavior, but these behaviors were not defined clearly. Thus simultaneous synchronization was found for interactive conversation, possibly related to turn-taking behavior. However, to appropriately test the M. Wilson and Wilson (2005) model, oscillations of EEG or MEG signals should be measured in an interactive hyperscanning paradigm because the model makes specific predictions about coupled oscillators between brains. For example, it might be interesting to look at interbrain synchronization in different turn-taking situations. Coordination might be stronger when turn-taking occurs more smoothly, or for sequences, and less strong when long lags are more common or repair is necessary, for example (see, e.g., Schilbach, 2015 for a similar proposal in social cognition more generally). In this respect, collaborations between CA and neuroimaging researchers might be very fruitful. CA researchers could, for example, retrospectively analyze properties of the conversations recorded during hyperscanning. Then, specific action types and features of the turn-taking practices employed by interlocutors could be related to interbrain coordination between them.

In summary, brain-imaging studies have elucidated many of the cognitive processes involved in turn-taking. They have shown experimentally that listeners often recognize actions early in the incoming turn and in some cases even anticipate them. After having recognized the action (or being able to predict it), listeners appear to start planning their next turn as soon as possible, according to neuroimaging research. Recipients also appear to anticipate the end of the current TCU well in advance and can sometimes precisely estimate the actual forthcoming words (see also Lerner, 2002). No brain-imaging studies have yet investigated the importance of completion cues in identifying ends of TCUs, but other experimental research suggests they play an important role as well. Most of the neuroimaging evidence thus appears to point to the Early Planning model outlined in Figure 1, although more research is needed before a definitive conclusion can be drawn. In addition, future hyperscanning EEG/MEG studies on conversation can potentially investigate claims about cyclic properties of interturn silences.

What can CA learn from neuroimaging studies?

The procedures of conversation analysis are designed to reveal "practices"—that is, the conventional use of certain linguistic or bodily resources to do certain interactional jobs, often of the form "An X normatively expects a Y" (Sidnell & Stivers, 2013). Such practices are often observed through the study of auditory or audiovisual recordings or of transcripts recording in redacted form a communicational exchange. These procedures have yielded bountiful insights into the organization of conversation (see next section). But they have their limitations: They offer only indirect insight into the mental processes that drive communication. In contrast, as we have reviewed, brain-imaging studies use online measures to focus more directly on the underlying cognitive processes, and we were able to show how they might adjudicate between the two models of mental processes underlying turn-taking sketched in Figure 1. The two models make different predictions—for example, early planning may require inhibition before release of articulation, relative lack of attention to later elements of the incoming turn, and difficulty recovering from mispredictions, while late planning risks come in too late. Such models directly address a key issue in turn-taking—namely, how is the amazing speed of turn transition actually achieved?

More generally, the study of online cognitive processes offers fresh insights that may be obscured by a CA approach. For example, it's now canonical in CA to analyze conversation as the outcome of different subsystems or modules (e.g., Schegloff, 2006): turn taking (Sacks et al., 1974), sequence organization (Schegloff, 2007), turn design and action recognition (Levinson, 2013), repair (Schegloff, Jefferson, & Sacks, 1977), and so on. But viewing these systems in terms of the underlying cognitive processes, one must perforce integrate them into a single online procedure: Action projection may leap ahead of unfolding turn design and be utilized to predict possible upcoming

completion of the current TCU. The great contribution that neuroimaging can make is to give us millisecond-by-millisecond insight into the underlying mental processes and how they are called upon to implement "practices." Neuroimaging also offers its own analysis of subsystems—for example, the use of the particular "mentalizing" circuitry involved in recognition of actions versus the specialized motor system circuitry involved in language output. By combining temporal precision with the activation of distinct mental subsystems, neuroimaging promises to deepen our understanding of the remarkable skills involved in carrying off conversation. In due course it may also offer important insights into systematic failures of the interactive system in autism, schizophrenia, brain injury, and the like (cf., e.g., Arie, Tartaro, & Cassell, 2008; Goodwin, 2003; O'Reilly, Lester, & Muskett, 2015). That said, it is interesting to note that the neuroimaging research reviewed here has by and large confirmed earlier CA observations, providing converging evidence.

How can future neuroimaging studies benefit from CA?

As psycholinguists generally become more interested in tackling questions about interactive language use, they will find the insights of CA increasingly useful. We have here reviewed just one of the subsystems normally treated in CA—turn-taking. And we have done so with numerous limitations. Readers may have noted that most such research has exploited adjacency-pair organization and focused on projectable responses within such pairs, rather than on the less predictable initiating actions. However, the EEG study by Gisladottir et al. (2015) on action recognition showed that listeners may recognize actions more quickly when they are second pair-parts than initiating actions like pre-offers, thus predicting faster responses within adjacency pairs (see also S. G. Roberts, Torreira, & Levinson, 2015). There are many further aspects of the turn-taking system already noted by CA that would repay the attention of neuroimaging research. How are multimodal signals pertaining to turn taking (gesture, gaze, speech) integrated, and which modalities are preferentially and more speedily processed? Are the planning processes involved in dyadic conversation of a different kind compared to multiparty conversation, where next-speaker selection is more of an issue and competition for the floor more likely?

CA has been a prodigal provider of insights into the subsystems involved in conversational organization, and all of these would repay reexamination from a cognitive processing point of view, using the millisecond-by-millisecond scalpel of EEG/MEG. We have mentioned action formation and recognition, which must be cognitively highly complex areas involving context, format, and epistemics (see Heritage, 2012)—how are conflicting factors here resolved? The CA distinction between the turn—the collaborative outcome of the turn-taking system—and the TCU—the turn's possible subunits—has yet to be properly exploited in experimental studies: As we noted, it is the TCU that is the relevant unit for understanding projection. Consider also sequence organization and repair. Sequences offer potential distant look-ahead, as when a pre-action projects a future course of action: Can we detect the forward processing here, and how are the potential ambiguities mentally navigated? Equally, can we find online evidence for the CA analysis of the successive loci for possible repair (Schegloff et al., 1977)? How do participants judge the proper timing for other-initiation of repair or to what extent is this a natural outcome of attempted resolution (Kendrick, 2015), and how do they choose between the formats for doing so (Dingemanse et al., 2015)? Another area inviting cognitive processing exploration is preference organization: There are some preliminary indications that disaffiliative actions trigger early and distinctive brain activity (as in the dispreferred response in Bögels et al., 2015 or the declining of offers in Gisladottir et al., 2015). This special sensitivity may be a further indication of the central role that "face" considerations may play in social life (e.g., Brown & Levinson, 1987).

An important general lesson from CA is that conversation is the core niche for both language use and social life and that it has special properties not found in many other kinds of language use. As we have emphasized, there are many difficulties surrounding the neuroimaging of spontaneous interaction, and the compromises that have been made are often unsatisfactory. For example, the use

of reenacted stimuli or read speech inevitably changes prosody (e.g., Blaauw, 1994), and overhearing recorded conversations is distinct from being a participant in a conversation (see Schilbach et al., 2013; Schober & Clark, 1989), which is reflected in different brain activation in both production (Sassa et al., 2007; Willems et al., 2010) and in comprehension (Rice & Redcay, 2016). A way forward is to create interactive situations in the lab (see Bašnáková et al., 2015; Bögels, Barr et al., 2015, Bögels, Magyari et al., 2015), or to use stimuli from actual conversational corpora (see Bögels, Kendrick et al., 2015; Magyari et al., 2014), or even a combination of the two.

In conclusion, brain-imaging research into turn taking and into conversation more generally is still in its infancy, but it promises a different and important perspective on the underlying iceberg of cognitive processes that support our wondrous conversational capacities.

Funding

References

Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., & Haynes, J.-D. (2011). Flow of affective information between communicating brains. *NeuroImage*, *54*(1), 439–446. doi:10.1016/j.neuroimage.2010.07.004

Arie, M., Tartaro, A., & Cassell, J. (2008, May). *Conversational turn-taking in children with autism: Deconstructing reciprocity into specific turn-taking behaviors*. Paper presented at the International Meeting for Autism Research, London, England.

Austin, J. L. (1975). *How to do things with words* (Vol. 1955). Oxford, England: Oxford University Press.

Bašnáková, J., van Berkum, J., Weber, K., & Hagoort, P. (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers? *Neuropsychologia*, *76*, 79–91. doi:10.1016/j.neuropsychologia.2015.03.030

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578. doi:10.1093/cercor/bht112

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V. D., Schmahl, C., . . . Meyer-Lindenberg, A. (2015). Information flow between interacting human brains: Identification, validation, and relationship to social expertise. *Proceedings of the National Academy of Sciences*, *112*(16), 5207–5212. doi:10.1073/pnas.1421831112

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, *14*(4), 359–375. doi:10.1016/0167-6393(94)90028-0

Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2015). Conversational interaction in the scanner: Mentalizing during language processing as revealed by MEG. *Cerebral Cortex*, *25*(9), 3219–3234. doi:10.1093/cercor/bhu116

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no . . . How the brain interprets the pregnant pause in conversation. *Plos One*, *10*(12), e0145474. doi:10.1371/journal.pone.0145474

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*, 12881. doi:10.1038/srep12881

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57. doi:10.1016/j.wocn.2015.04.004

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General*, *143*(1), 295–311. doi:10.1037/a0031858

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation*, *53*, 301–344.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge, England: Cambridge University Press.

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*(4), 643–684. doi:10.1080/03640210802066816

Butterworth, B. (Ed.). (1980). *Language production* (Vol. 1). London, England: Academic Press.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (2007). *Handbook of psychophysiology*. Cambridge, England: Cambridge University Press.

Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language, 92*, 234–253. doi:10.1016/j.jml.2016.06.013

Clayman, S. E. (2013). Turn-constructional units and the transition-relevance place. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 151–166). Malden, MA: Blackwell.

Coulson, S., & Lovett, C. (2010). Comprehension of non-conventional indirect requests: An event-related brain potential study. *Italian Journal of Linguistics*, *22*(1), 107–124.

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction* (Vol. 25). Amsterdam, The Netherlands: John Benjamins.

Danks, J. H. (1977). Producing ideas and sentences. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory*. Hillsdale, NJ: Lawrence Erlbaum.

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82, 515–535. doi:10.1353/lan.2006.0130

De Vos, C., Torreira, F., & Levinson, S. C. (2015). Turn-timing in signed conversations: Coordinating stroke-to-stroke turn boundaries. *Frontiers in Psychology*, 6, 268. doi:10.3389/fpsyg.2015.00268

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., … Kotz, S. (2015). Universal principles in the repair of communication problems. *Plos ONE*, 10(9), e0136100. doi:10.1371/journal.pone.0136100

Dumas, G., Chavez, M., Nadel, J., & Martinerie, J. (2012). Anatomical connectivity influences both intra- and inter-brain synchronizations. *Plos One*, 7(5), e36414. doi:10.1371/journal.pone.0036414

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *Plos ONE*, 5(8), e12166. doi:10.1371/journal.pone.0012166

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2), 283–292. doi:10.1037/h0033031

Egorova, N., Pulvermüller, F., & Shtyrov, Y. (2014). Neural dynamics of speech act comprehension: An MEG study of naming and requesting. *Brain Topography*, 27(3), 375–392. doi:10.1007/s10548-013-0329-3

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2013). Early and parallel processing of pragmatic and semantic information in speech acts: Neurophysiological evidence. *Frontiers in Human Neuroscience*, 7, 86. doi:10.3389/fnhum.2013.00086

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2016). Brain basis of communicative actions in language. *NeuroImage*, 125, 857–867. doi:10.1016/j.neuroimage.2015.10.055

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. *Studies in Interactional Sociolinguistics*, 13, 134–184.

Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, 6, 751. doi:10.3389/fpsyg.2015.00751

Gisladottir, R. S., Bögels, S., & Levinson, S. C. (2016). Oscillatory brain responses reflect anticipation during comprehension of speech acts in spoken dialogue. Manuscript submitted for publication.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *Plos ONE*, 10(3), e0120068. doi:10.1371/journal.pone.0120068

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York, NY: Irvington.

Goodwin, C. (2003). *Conversation and brain damage*. Oxford, England: Oxford University Press.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279. doi:10.1111/1467-9280.00255

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., … Poeppel, D. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *Plos Biology*, 11(12), e1001752. doi:10.1371/journal.pbio.1001752

Hagoort, P., & Levinson, S. C. (2014). Neuropragmatics. In M. S. Gazzaniga & G. R. Mangun (Eds.), *The cognitive neurosciences* (pp. 667–674). Cambridge, MA: MIT Press.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, 16(2), 114–121. doi:10.1016/j.tics.2011.12.007

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555–568. doi:10.1016/j.wocn.2010.08.002

Heritage, J. (2012). The epistemic engine: Sequence organization and territories of knowledge. *Research on Language & Social Interaction*, 45(1), 30–52. doi:10.1080/08351813.2012.646685

Heritage, J., & Clayman, S. (2010). Conversation analysis: Some theoretical background. In J. Heritage & S. Clayman (Eds.), *Talk in action: Interactions, identities, and institutions* (pp. 5–19). Chichester, England: Wiley-Blackwell.

Holler, J., & Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: Optimizing recipiency. *Frontiers in Psychology*, 6, 98. doi:10.3389/fpsyg.2015.00098

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1), 101–144. doi:10.1016/j.cognition.2002.06.001

Jansen, S., Wesselmeier, H., De Ruiter, J. P., & Mueller, H. M. (2014). Using the readiness potential of button-press and verbal response within spoken language processing. *Journal of Neuroscience Methods*, 232, 24–29. doi:10.1016/j.jneumeth.2014.04.030

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *The Journal of Neuroscience*, 32(45), 16064–16069. doi:10.1523/JNEUROSCI.2926-12.2012

Kendrick, K. H. (2015). The intersection of turn-taking and repair: The timing of other-initiations of repair in conversation. *Frontiers in Psychology*, 6, 250. doi:10.3389/fpsyg.2015.00250

Kendrick, K. H., & Drew, P. (2016). Recruitment: Offers, requests, and the organization of assistance in interaction. *Research on Language and Social Interaction*, *49*(1), 1–19. doi:10.1080/08351813.2016.1126436

Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*(4), 255–289. doi:10.1080/0163853X.2014.955997

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale [Brain potential changes at arbitrary movements and passive movements of people: Readiness potential and re-afferent potentials]. *Pflüger's Archiv für die Gesamte Physiologie des Menschen und der Tiere*, *284*(1), 1–17. doi:10.1007/BF00412364

Kuhlen, A. K., Allefeld, C., Anders, S., & Haynes, J.-D. (2015). Towards a multi-brain perspective on communication in dialogue. In R. M. Willems (Ed.), *Cognitive neuroscience of natural language use* (pp. 182–200). Cambridge, England: Cambridge University Press.

Kuhlen, A. K., Allefeld, C., & Haynes, J.-D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, *6*, 266. doi:10.3389/fnhum.2012.00266

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. doi:10.1126/science.7350657

Lerner, G. H. (2002). Turn-sharing: The choral co-production of talk-in-interaction. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), *The language of turn and sequence*. New York, NY: Oxford University Press.

Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232. doi:10.1016/S1364-6613(99)01319-4

Levinson, S. C. (1983). *Pragmatics*. Cambridge, England: Cambridge University Press.

Levinson, S. C. (2013). Action formation and ascription. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 101–130). Chichester, England: Wiley-Blackwell.

Levinson, S. C. (2016). Turn-taking in human communication—Origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. doi:10.1016/j.tics.2015.10.010

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731. doi:10.3389/fpsyg.2015.00731

Lindenberger, U., Li, S.-C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neuroscience*, *10*(1), 22. doi:10.1186/1471-2202-10-22

Local, J., & Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association*, *42*(3), 255–280. doi:10.1017/S0025100312000187

Luck, S. J., & Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components*. Oxford, England: Oxford University Press.

Magyari, L., Bastiaansen, M. C., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*, 2530–2539. doi:10.1162/jocn_a_00673

Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, *3*, 376. doi:10.3389/fpsyg.2012.00376

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 621.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213. doi:10.1016/S0749-596X(03)00028-7

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., … Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage*, *16*(4), 1159–1164. doi:10.1006/nimg.2002.1150

Oostdijk, N. (2000). Het Corpus Gesproken Nederlands [The corpus of spoken Dutch]. *Nederlandse Taalkunde*, *5*, 280–284.

O'Reilly, M., Lester, J. N., & Muskett, T. (2015). Discourse/conversation analysis and autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*(2), 355–359. doi:10.1007/s10803-015-2665-5

Piai, V., Roelofs, A., Rommers, J., Dahlslätt, K., & Maris, E. (2015). Withholding planned speech is reflected in synchronized beta-band oscillations. *Frontiers in Human Neuroscience*, *9*, 549. doi:10.3389/fnhum.2015.00549

Pomerantz, A., & Heritage, J. (2013). Preference. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 210–228). Malden, MA: Blackwell.

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227–1235. doi:10.1002/(ISSN)1522-2594

Rice, K., & Redcay, E. (2016). Interaction matters: A perceived social partner alters the neural processing of human speech. *NeuroImage*, *129*, 480–488. doi:10.1016/j.neuroimage.2015.11.041

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, *48*(9), 1079–1093. doi:10.1016/j.specom.2006.02.001

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes*, *48*(5), 331–354. doi:10.1080/0163853X.2011.558002

Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: A corpus study. *Frontiers in Psychology*, 6, 509. doi:10.3389/fpsyg.2015.00509

Sacks, H. (1987). On the preferences for agreement and contiguity in sequences in conversation. In G. Button & J. R. E. Lee (Eds.), *Talk and social organization* (pp. 54–69). Clevedon, England: Multilingual Matters.

Sacks, H. (1992). Lecture 1. Rules of conversational sequence. In H. Sacks & G. Jefferson (Eds.), *Lectures on conversation* (Vol. I, pp. 3–11). Oxford, England: Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735. doi:10.1353/lan.1974.0010

Sassa, Y., Sugiura, M., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2007). Cortical mechanism of communicative speech production. *NeuroImage*, 37(3), 985–992. doi:10.1016/j.neuroimage.2007.05.059

Schegloff, E. A. (1984). On some questions and ambiguities in conversation. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action*. Cambridge, England: Cambridge University Press.

Schegloff, E. A. (1992). To Searle on conversation: A note in return. In H. Parret & J. Verschueren (Eds.), *(On) Searle on conversation* (pp. 113–128). Amsterdam, The Netherlands: John Benjamins.

Schegloff, E. A. (1998). Reflections on studying prosody in talk-in-interaction. *Language and Speech*, 41(3–4), 235–263.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1), 1–63. doi:10.1017/S0047404500001019

Schegloff, E. A. (2006). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 70–96). New York, NY: Berg.

Schegloff, E. A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis* (Vol. 1). Cambridge, England: Cambridge University Press.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361–382. doi:10.1353/lan.1977.0041

Schilbach, L. (2015). The neural correlates of social cognition and social interaction. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference* (pp. 159–164). Amsterdam, The Netherlands: Elsevier.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(4), 393–414. doi:10.1017/S0140525X12000660

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., & Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences*, 107(20), 9388–9393. doi:10.1073/pnas.1001791107

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Ed.), *Handbook of discourse processes* (pp. 123–164). Mahwah, NJ: Lawrence Erlbaum.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2), 211–232. doi:10.1016/0010-0285(89)90008-X

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge, England: Cambridge University Press.

Sidnell, J., & Stivers, T. (2013). *The handbook of conversation analysis* (Vol. 121). Chichester, England: Wiley-Blackwell.

Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, 136, 304–324. doi:10.1016/j.cognition.2014.10.008

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425–14430. doi:10.1073/pnas.1008662107

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592. doi:10.1073/pnas.0903616106

Torreira, F., Bögels, S., & Levinson, S. C. (2015). Breathing for answering: The time course of response planning in conversation. *Frontiers in Psychology*, 6, 284. doi:10.3389/fpsyg.2015.00284

van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S.-A. (2012). Pragmatics in action: Indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience*, 24(11), 2237–2247. doi:10.1162/jocn_a_00274

Vos, D. M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.-X., Huffel, V. S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics*, 8(2), 135–150. doi:10.1007/s12021-010-9071-0

Wells, B., & Macfarlane, S. (1998). Prosody as an interactional resource: Turn-projection and overlap. *Language and Speech*, 41(3–4), 265–294.

Wesselmeier, H., Jansen, S., & Müller, H. M. (2014). Influences of semantic and syntactic incongruence on readiness potential in turn-end anticipation. *Frontiers in Human Neuroscience*, 8, 296. doi:10.3389/fnhum.2014.00296

Wesselmeier, H., & Müller, H. M. (2015). Turn-taking: From perception to speech preparation. *Neuroscience Letters*, 609, 147–151. doi:10.1016/j.neulet.2015.10.033

Willems, R. M. (Ed.). (2015). *Cognitive neuroscience of natural language use*. Cambridge, England: Cambridge University Press.

Willems, R. M., De Boer, M., De Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychological Science*, *21*(1), 8–14. doi:10.1177/0956797609355563

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957–968. doi:10.3758/BF03206432

Wilson, T. P., & Zimmerman, D. H. (1986). The structure of silence between turns in two-party conversation. *Discourse Processes*, *9*(4), 375–390. doi:10.1080/01638538609544649

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*, *114*(4), 1803–1817. doi:10.1093/brain/114.4.1803