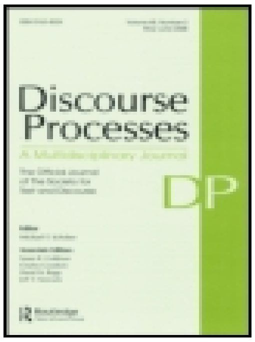

Discourse Processes

ISSN: 0163-853X (Print) 1532-6950 (Online) Journal homepage: http://www.tandfonline.com/loi/hdsp20

Coordinating Utterances During Turn-Taking: The Role of Prediction, Response Preparation, and Articulation

Ruth E. Corps, Chiara Gambi & Martin J. Pickering

To cite this article: Ruth E. Corps, Chiara Gambi & Martin J. Pickering (2017): Coordinating Utterances During Turn-Taking: The Role of Prediction, Response Preparation, and Articulation, Discourse Processes, DOI: 10.1080/0163853X.2017.1330031

To link to this article: http://dx.doi.org/10.1080/0163853X.2017.1330031

Published online: 23 Jun 2017.

DISCOURSE PROCESSES
https://doi.org/10.1080/0163853X.2017.1330031

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Coordinating Utterances During Turn-Taking: The Role of Prediction, Response Preparation, and Articulation

Ruth E. Corps, Chiara Gambi, and Martin J. Pickering

Department of Psychology University of Edinburgh, Edinburgh, UK

ABSTRACT

During conversation, interlocutors rapidly switch between speaker and listener roles and take turns at talk. How do they achieve such fine coordination? Most research has concentrated on the role of prediction, but listeners must also prepare a response in advance (assuming they wish to respond) and articulate this response at the appropriate moment. Such mechanisms may overlap with the processes of comprehending the speaker's incoming turn and predicting its end. However, little is known about the stages of response preparation and production. We discuss three questions pertaining to such stages: (1) Do listeners prepare their own response in advance?, (2) Can listeners buffer their prepared response?, and (3) Does buffering lead to interference with concurrent comprehension? We argue that fine coordination requires more than just an accurate prediction of the interlocutor's incoming turn: Listeners must also simultaneously prepare their own response.

Introduction

Traditional psycholinguistics has often focused on processing isolated words or sentences. Tony Sanford made great contributions in the development of much broader accounts of processing texts (Sanford & Garrod, 1981), including narratives (Sanford & Emmott, 2012). In a similar way, psycholinguistics can be extended to the study of conversational interaction (dialogue), which is arguably the most basic form of language use.

In conversation, interlocutors repeatedly and regularly switch between comprehending their partner's utterance and producing an appropriate and timely response. These processes are so finely coordinated that interlocutors often minimize both overlap and gaps between turns. Indeed, Stivers et al. (2009) found average inter-turn intervals between 0 and 200 ms in a comparison of 10 different languages, with overlap occurring only about 5% of the time (Levinson, 2016).

Duncan (1972, 1974; Duncan & Niederhe, 1974) proposed that interlocutors time their contributions during conversation by *reacting* to the presence of linguistic (e.g., drawl on the final syllable of the utterance) and nonlinguistic (e.g., termination of hand gestures) turn-yielding cues displayed at the end of the speaker's turn. However, turn-taking occurs far too rapidly for listeners to be simply reacting to such signals, given that a single word takes between 600 and 1,200 ms to produce, depending on word frequency (Indefrey & Levelt, 2004; Levelt, Roelofs, & Meyer, 1999) and a complete utterance takes longer still (around 1,500 ms) (Ferreira, 1991; Griffin & Bock, 2000). These timings may be slightly different during conversation but nevertheless suggest that if listeners wish to achieve inter-turn intervals of 200 ms (Stivers et al., 2009), then they must begin preparing their own response at least half a second before the speaker reaches the end of their turn.

CONTACT Ruth E. Corps rcorps@exseed.ed.ac.uk Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom.

Current theories agree that interlocutors achieve such timings using *prediction* (De Ruiter, Mitterer, & Enfield, 2006; Garrod & Pickering, 2015; Levinson, 2016). Research has focused on how the listener can use such predictions to determine when the speaker will reach the end of their turn (e.g., De Ruiter et al., 2006; Magyari, Bastiaansen, De Ruiter, & Levinson, 2014) so they can articulate their response at the appropriate moment. However, if listeners can predict what the speaker will say before the speaker reaches the end of their turn, then the listener can also begin preparing their own response in advance (Garrod & Pickering, 2015; Levinson, 2016). So how are the processes of prediction, response preparation, and articulation coordinated during turn-taking?

What the listener can predict depends on the characteristics of the utterance. For example, if the speaker says *Which character from the famous movies is also called 007?* (Bögels, Magyari, & Levinson, 2015), then the listener cannot predict the lexical content of the speaker's utterance until the word *007*. This inability means that the listener cannot predict when the speaker will reach the end of their turn and can only begin response preparation toward the end of the utterance. In these instances the listener likely times their response on the basis of turn-final cues (Beattie, Cutler, & Pearson, 1982; Gravano & Hirschberg, 2011; Hjalmarsson, 2011; Local & Walker, 2012). To achieve inter-turn intervals of 200 ms (Stivers et al., 2009), the processes of response preparation and articulation must overlap: The listener must have to articulate the first syllable of their response while simultaneously preparing subsequent words (i.e., response preparation must be incremental; Ferreira & Swets, 2002).

On the other hand, if the speaker says *Which character, also called 007, appears in the famous movies*?, then the listener may be able to predict the content of the speaker's utterance and could begin preparing their response (*James Bond* in this example) after hearing *007*. If this is the case (as in Bögels et al., 2015), then the processes of preparation and articulation are decoupled, and the listener has to buffer their response until the opportunity to articulate it arises. Buffering this response may in turn interfere with concurrent comprehension and the listener's ability to time their response appropriately.

However, comparatively little research has focused on the role of response preparation and articulation during turn-taking. In the following we review evidence on the role of prediction, response preparation, and articulation to address three unresolved questions: (i) Do listeners prepare their own response in advance?, (ii) Can the listener separate the processes of preparation and articulation and buffer their response?, and (iii) Does buffering this response interfere with concurrent comprehension? We discuss how experimental research has examined each of these questions and highlight issues that have yet to be addressed.

Predicting the speaker's turn

Previous research demonstrates that listeners can predict the semantics (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003), syntax (Staub & Clifton, 2006; Van Berkum, Brown, Zwisterlood, Koojiman, & Hagoort, 2005), and phonology (DeLong, Urbach, & Kutas, 2005; Vissers, Chwilla, & Kolk, 2006) of the speaker's utterance. For example, listeners can use a verb's semantics to predict the meaning of the speaker's forthcoming utterance (e.g., looking toward an edible object more when the speaker says *eat* than when the speaker says *move*; Altmann & Kamide, 1999). Furthermore, Staub and Clifton found that readers can predict the syntax of an utterance: Participants read *or the subway* faster after reading *the team took either the train…* than after *the team took the train…* . Finally, DeLong et al. found a larger N400 effect in response to indefinite articles (*a* or *an*) that mismatched an expected upcoming noun. For example, participants read sentences such as *the day was breezy so the boy went outside to fly…* and displayed an N400 effect when the sentence ended with *an airplane* (unpredictable completion) rather than *a kite* (predictable completion). This effect occurred at *a* or *an*, suggesting participants had predicted the form of the forthcoming utterance. Together, these results suggest participants can use the preceding context of a speaker's utterance to predict the content (the semantics, syntax, and phonology) of the upcoming utterance.

The fact that comprehenders often fail to detect anomalies (e.g., Bohan, Leuthold, Hijikata, & Sanford, 2010; Bohan & Sanford, 2008; Sanford, Leuthold, Bohan, & Sanford, 2010; Sanford & Sturt, 2002) is consistent with the notion of prediction. For example, readers will often not notice anything unusual in the question *When an aircraft crashes, where should the survivors be buried*? (although survivors do not get buried). But they usually detect the anomaly in *When a bicycle accident occurs, where should the survivors be buried?* (Barton & Sanford, 1993). Although burying survivors has a good global fit in the context of an aircraft crash (i.e., readers predict there are often people to be buried), it has a poor fit in the context of a bicycle crash (Sanford & Garrod, 1998). Thus, prediction may be related to some instances of *shallow* semantic processing. A good explanation is that people predict general aspects of meaning (e.g., relating to airplane crashes) rather than specific words or associated concepts (otherwise they would realize that *survivors* would be inappropriate).

However, content predictions alone may not be sufficient for finely coordinated turn- taking, as the listener must also predict when the speaker is likely to reach the end of their utterance (a *turn-end* prediction). Predicting this point enables the listener to articulate their response at the appropriate moment, so the risk of extensive conversational overlap or a long gap is minimized. Most research assessing turn-end prediction has used a button-press paradigm (e.g., De Ruiter et al., 2006; Magyari et al., 2014), in which participants are presented with full conversational turns and predict (indicated via a button-press) when they expect speakers to reach the end of their utterance. Using this paradigm, De Ruiter et al. found that participants could accurately predict when a speaker's turn was about to end. Importantly, they found that the actual words of the utterance (but not the pitch) were necessary for accurate turn-end prediction. Although it is possible other sources of prosodic information may be important (e.g., duration; Bögels & Torreira, 2015), these results nevertheless suggest that listeners predicted the speaker's turn-end by predicting the content of the utterance.

Further evidence for this conclusion comes from a study by Magyari et al. (2014), who manipulated the content predictability of their stimuli. In a gating paradigm, participants were auditorily presented with turns from actual conversations in fragments of increasing duration and completed these turns with the words they expected to follow the preceding context (much like a typical cloze task; Taylor, 1953). Magyari et al. assessed the predictability of these responses using entropy, which measures the consistency of completions across participants. Participants provided more consistent completions in the predictable (e.g., *I live in the same house with four women and another man*) than unpredictable condition (e.g., *She was again alone in the north*) and were more likely to complete these fragments with the words the original speakers had used. A separate group of participants, who completed the button-press task, responded 80 ms before the turn end when the final words were predictable but 139 ms after the turn end when the final words were unpredictable.

Using the same gating method as Magyari et al. (2014), Magyari and De Ruiter (2012) assessed the number of words participants expected to complete sentence fragments. They correlated these results with De Ruiter et al.'s (2006) and found that turns that elicited later button-presses tended to be completed with more words in the gating paradigm. Although such correlational data should be interpreted with some caution, these results suggest listeners may also predict the turn-end by predicting the number of words the speaker will use.

Magyari and De Ruiter (2012) suggested that listeners' ability to predict the number of words in a turn may depend on their ability to predict syntactic structure. However, the role of syntactic information during turn-end prediction is unclear. Riest, Jorschick, and De Ruiter (2015) showed that listeners could still predict the speaker's turn-end when closed class words (which primarily serve a syntactic role; Brown, Hagoort, & Ter Kaus, 1999) were removed from turns using low-pass filtering. However, when open class words (which primarily serve a semantic role) were removed, participants were more likely to respond reactively (after the turn-end), suggesting that semantic information is more important for turn-end prediction than syntactic information.

Consistent with the comprehension literature, these results suggest interlocutors may predict the speaker's unfolding utterance at a number of linguistic levels. Listeners can use these predictions to determine when the speaker will reach the end of their turn and prepare a response in advance.

Response preparation

Before preparing their own response, listeners must determine the speaker's underlying speech act to decide whether a response is normatively required (in some cases a response is not necessary, such as after rhetorical questions; Sacks, Schegloff, & Jefferson, 1974; Schegloff, 1968, 2000) and what type of response is required (in some cases, a simple *yes* or *no* response is sufficient). Gisladottir, Chwilla, and Levinson (2015) explored the time course of speech act recognition in an electroencephalographic (EEG) experiment where participants listened to two-turn dialogues. They manipulated the speech act of the second turn, so it was either an answer (e.g., Speaker A: *How are you going to pay?* Speaker B: *I have a credit card*), a declination (e.g., A: *I can lend you money*. B: *I have a credit card*), or an offer (e.g., A: *I don't have any money*. B: *I have a credit card*). Electrophysiological results showed differential frontal positivities around 200 ms after utterance onset, suggesting listeners are capable of determining the speaker's speech act early, before the speaker reaches the end of their utterance.

After determining the speech act and having heard or predicted a sufficient part of the speaker's utterance, listeners can begin preparing their own response. Most theories of language production suggest preparation involves at least three stages: message construction (conceptualization), formulation (lexical selection, structure building, and phonological encoding), and articulation (Bock, 1995; Levelt, 1983, 1989, 1992). But when do listeners begin preparing this response? Torreira, Bögels, and Levinson (2015) examined the time course of listeners' prespeech inbreaths, which have been shown to be related to response preparation (e.g., Fuchs, Petrone, Krivokapić, & Hoole, 2013). They found that listeners took inbreaths after the end of the speaker's utterance, suggesting listeners may have reacted to turn final cues displayed at the end of the speaker's utterance. However, it is not clear whether inbreaths index articulation or earlier stages of response preparation. As a result, we cannot determine how much of their response listeners prepared before they took an inbreath.

Additional studies have used dual-task paradigms, where participants engage in conversation while conducting an unrelated secondary task. Previous research demonstrates that all stages of preparation (such as lemma, word form, and phoneme selection; Cook & Meyer, 2008; Ferreira & Pashler, 2002; Roelofs, 2008; Roelofs & Piai, 2011) are cognitively demanding. For instance, Ferreira and Pashler had participants name pictures while discriminating between tones and found that increasing the time required for lemma selection (by presenting pictures following less constraining sentences) and word form selection (by presenting pictures with lower frequency names) delayed both picture naming and tone discrimination responses. Manipulating the time required for phoneme selection (by presenting pictures with phonologically related distractors) facilitated picture naming but did not affect tone discrimination. However, Cook and Meyer failed to replicate these latter results and instead found that phoneme selection did interfere with dual-task performance. Thus, it is possible all stages of response preparation require central processing capacity.

As a result, dual-task paradigms assessing the time course of response preparation during turn-taking assume that performance on a secondary task should decline when participants begin response preparation. Using this method, Boiteau, Malone, Peters, and Almor (2014) found that listeners' performance on a visuomotor tracking task declined toward the end of their interlocutor's turn, suggesting they began preparing their own response toward the end of the speaker's utterance. Sjerps and Meyer (2015) found similar results (using a finger-tapping task), even when participants knew which row of pictures they would have to describe as soon as the speaker produced their first word and could prepare a response in advance of the turn-end. These results are consistent with findings in monologue: Speakers are slower to categorize tones played toward the end than at the beginning of their clause (Ford & Holmes, 1978) reflecting planning of the upcoming clause.

These studies suggest that response preparation and articulation are tightly interwoven during turn-taking: Listeners only begin response preparation toward the end of the speaker's utterance, when they will soon have the opportunity to articulate this response. As a result, preparation and articulation appear to be separate from content predictions: Listeners do not prepare their response well in advance, even when they can predict the content of the speaker's utterance (as in Sjerps & Meyer, 2015). Whether an accurate turn-end prediction is necessary for preparation is unclear, however. Listeners may delay preparation until they can predict the speaker's turn-end, most likely toward the end of the utterance. Alternatively, listeners may only begin preparation once the speaker displays a turn-final cue, which usually occurs towards the end of the speaker's turn (linguistic or nonlinguistic; Beattie et al., 1982; Gravano & Hirschberg, 2011; Hjalmarsson, 2011; Local & Walker, 2012).

If listeners only begin preparation towards the end of the speaker's utterance, then they may not be able to prepare their whole response before articulation. To avoid long gaps between utterances, they must have to plan their response incrementally at the same time as they articulate this response. There is extensive evidence that language production can be incremental in this way (Ferreira & Swets, 2002; Meyer, 1996), and so listeners could begin articulation very early, perhaps after they have prepared the first syllable of their utterance, while simultaneously planning and preparing subsequent parts of their response.

Although dual-task paradigms shed some light on the processes of response preparation, the secondary tasks (e.g., finger tapping, visuomotor tracking) involved in these paradigms are non-linguistic and often involve processes that are unrelated to the main task. This is not the case in conversation, where interlocutors engage in simultaneous production and comprehension, which are often related: Listeners use production mechanisms to prepare utterances that often complement their comprehension of the speaker's utterance (e.g., question–answer sequences). Furthermore, it is unclear which stages of response preparation the dual-task paradigm taps into. Previous research demonstrates that all aspects of response preparation, including phonological encoding (Roelofs & Piai, 2011) and possibly articulation or speech monitoring (Almor, 2008), are cognitively demanding. As a result, it is possible that dual-task difficulty only arises toward the end of the speaker's utterance because it is more sensitive to later, rather than earlier, stages of response preparation.

As an alternative approach, Bögels et al. (2015) measured EEG correlates during a question–answering task, where the information needed to prepare a response was available either early (e.g., *which character, also called 007, appears in the famous movies?*) or late (e.g., *which character from the famous movies is also called 007?*) in the utterance. Participants were quicker to answer questions when the critical information (*007*) was available early rather than late, and EEG correlates revealed a positive electrophysiological effect in the middle frontal and precentral gyri, which overlap with brain areas involved in speech production (Indefrey & Levelt, 2004), and a reduced alpha power, which is associated with motor response preparation (Babiloni et al., 1999). Both of these effects occurred around 500 ms after the onset of the critical information (*007*) necessary for response preparation.

These results are inconsistent with the dual-task findings reported by Sjerps and Meyer (2015) and suggest that listeners can prepare their own response as soon as they can predict the content of the speaker's utterance. As a result, they suggest the processes of content prediction and response preparation can be decoupled from turn-end prediction and articulation. After hearing *007*, listeners can predict the content of the speaker's utterance (e.g., that the speaker is asking a question about James Bond) and can prepare a response consistent with this prediction, even though they will not have the opportunity to articulate this response until the speaker reaches the end of their turn.

However, we note that Bögels et al. (2015) used general knowledge questions, and answers likely had to be retrieved from memory. Although previous experimental research has found that the middle frontal and precentral gyri are associated with language production (Indefrey & Levelt, 2004), other studies report that the middle frontal gyrus may also be involved in episodic memory retrieval (Cabeza, 2002; Rajah, Languay, & Grady, 2011; Raz et al., 2005). The effects observed by Bögels et al. may thus reflect the processes of memory retrieval rather than utterance preparation.

Additional research has found that the extent of advance planning is fairly flexible (Konopka, 2012; Swets, Jacovina, & Gerrig, 2013; Van de Velde, Meyer, & Konopka, 2014), which may explain the discrepancies in results of Boiteau et al. (2014), Sjerps and Meyer (2015), and Bögels et al. (2015). For example, Konopka found that increasing the familiarity of sentence structure (through repetition) and lexical items (by manipulating frequency and recent usage) increased speakers' scope of response preparation from one to two words. Although participants did not have to coordinate their utterances with another speaker in this study, interlocutors in dialogue often align their representations and repeat sentence structures and words previously used by their partner (Branigan, Pickering, & Cleland, 2000; Garrod & Anderson, 1987; see Pickering & Garrod, 2004), which may facilitate advance planning.

Buffering and articulating a response

In instances where the listener prepares their response in advance of articulation, they must store this response in a buffer until it can be articulated at the appropriate moment (see Postma, 2000). If listeners could not buffer their prepared response, then they would have to either take their turn as soon as they completed all stages of response preparation, interrupting their partner, or prepare their response only when they could be sure they would be able to produce this response (i.e., response preparation would be tied to turn-end prediction).

Results from immediate and delayed picture-naming studies, where participants name pictures while ignoring distractor words, suggest participants can buffer their utterances at various stages of production (Mädebach, Oppermann, Hantsch, Curda, & Jescheniak, 2011; Piai, Roelofs, & Schriefers, 2011; Piai, Roelofs, & Schriefers, 2014; Schriefers, Meyer, & Levelt, 1990). For instance, Piai et al. (2011) found that participants were slower to name pictures when distractor words were semantically related (known as the *semantic interference effect*) in an immediate but not in a delayed naming condition. In the immediate condition a semantically related distractor word interfered with ongoing lexicalization. However, participants in the delayed condition had already completed the processes of lexical selection and were likely buffering their response at the phonological level until they were given the *go* signal. In the context of turn-taking, the *go* signal corresponds to the moment when the speaker reaches the end of his or her turn and the listener can articulate his or her response.

Furthermore, Piai, Roelofs, Rommers, Dahlslätt, and Maris (2015a) found alpha–beta desynchronization (8–30 Hz) in the occipital cortex and beta synchronization (12–40 Hz) in the middle frontal and superior frontal gyri during delayed but not immediate naming. Alpha–beta desynchronization has been associated with motor aspects of articulation (Piai, Roelofs, Rommers, & Maris, 2015b), whereas beta synchronization has been associated with maintaining the current cognitive state until the response can be articulated (Engel & Fries, 2010; Kilavik, Zaepffel, Brovelli, MacKay, & Riehle, 2013). These findings suggest that when listeners prepare their response in advance of articulation, they buffer and continue to rehearse this response (presumably so they do not forget what they wish to say) until they are given the opportunity to take their turn.

In the context of turn-taking, two questions follow from Piai et al.'s (2011, 2015a) results. First, does preparing a response and holding it in an articulatory buffer interfere with the listeners' ability to comprehend their interlocutor's incoming turn? If so, does this interference depend on how much of their response the listener has prepared and buffered? In cases where listeners prepare their response in advance, they have to represent both their prepared response (using production mechanisms) and their interlocutor's unfolding utterance (using comprehension mechanisms). Production and comprehension recruit overlapping neural circuits (Menenti, Gerhan, Segaert, & Hagoort, 2011; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012; Silbert, Honey, Simony, Poeppel, & Hasson, 2014; Watkins, Strafella, & Paus, 2003; Wilson, Saygin, Sereno, & Iacoboni, 2004) and hence most likely share processes.

As a result, preparing, buffering, and rehearsing a response using production processes could be detrimental to concurrent comprehension. When the listener has to prepare a longer response, then the resources allocated to buffering and rehearsing this response may be larger, which may lead to greater interference with comprehension mechanisms. However, it could also be the case that activating representations during comprehension makes it easier to activate these representations during subsequent production, which may make it easier for the listener to prepare their own response. Future research could examine these issues by assessing the listener's comprehension of the speaker's utterance in instances where the listener can prepare their response well before they are given the opportunity to articulate (as in Bögels et al., 2015) and by comparing instances where the listener has to prepare and buffer a shorter response (e.g., a simple *yes* or *no* answer) to instances where a much longer, multiword response is required. Although preparing a response in advance presumably enables the listener to achieve shorter inter-turn intervals (e.g., Stivers et al., 2009), addressing these issues will determine whether there is any cognitive cost to having to buffer this response when articulation cannot begin immediately.

In addition, buffering a response may interfere with the listener's ability to produce an utterance at the appropriate moment. Regardless of whether listeners launch articulation of their response on the basis of turn-final cues (Beattie et al., 1982; Gravano & Hirschberg, 2011; Hjalmarsson, 2011; Local & Walker, 2012) or a prediction of the speaker's turn-end (De Ruiter et al., 2006; Magyari et al., 2014), failure to fully comprehend the speaker's utterance because of preparing, buffering, and rehearsing a response may impact the listener's ability to respond at the appropriate moment, which may lead to more overlaps or gaps between utterances. On the other hand, the speaker's utterance generally becomes increasingly predictable as the turn unfolds. For example, if the speaker says *Dogs are…*, then the listener cannot accurately predict the speaker's next words. However, as the speaker produces the words *my favorite*, the listener can predict that the most likely next words will be *animal.* As a result, it may matter less that the listener fully comprehends the end of the speaker's turn toward the end of it because they can better rely on their predictions. Future research could address these issues by examining whether buffering a prepared response for a long time leads to longer overlaps or gaps between utterances. If it does, then it would suggest that early preparation interferes with the mechanisms of turn-end prediction.

Finally, once listeners know they can produce their response, is there a *target* (or an ideal moment) for articulating this response? Schegloff (1968, 2000; Jefferson, 1984; see also Garrod & Pickering, 2015) claimed that listeners aim to respond to the speaker's utterance on the beat, which probably corresponds to a syllable. Moreover, Wilson and Wilson (2005) argued that turn-taking involves the automatic entrainment of oscillators (which are pools of neurons that serve timing-related functions) between speaker and listener. These oscillators entrain to syllable rate, so the listener's readiness to speak is at a maximum when the speaker's readiness is at a minimum (i.e., when the speaker is mid-syllable) and vice versa. Previous research supports the involvement of entrainment during conversation (Cummins, 2002, 2003, 2009; Jungers, Palmer, & Speer, 2002; Street, 1984; Zion Golumbic et al., 2013). For example, Jungers and Hupp (2009) found that listeners were more likely to produce a response at a fast rate after hearing a speaker produce an utterance at this rate, suggesting listeners entrained to the timing of their interlocutor's speech rate, which in turn influenced their own production. Consequently, listeners could use entrainment to attempt to articulate their own response on the next beat after the end of the speaker's utterance.

However, whether syllables necessarily underlie such entrainment is unclear. Research on speech segmentation suggests that listeners are sensitive to syllabic boundaries in syllable-based languages (such as French; Mehler, Dommergue, Fraunfelder, & Segui, 1981) but not in stress-based languages (such as English; Cutler, Mehler, Norris, & Segui, 1983, 1986; Cutler & Norris, 1988). If entrainment draws on similar information as speech segmentation, then the mechanisms used to determine the beat (or the target for articulation) may be different in different languages: The target may be determined using syllabic information in syllable-based languages but may be identified using different information in stress-based languages. As a result, further research is needed to determine whether the listener does indeed work toward a *target* when producing a response, whether this target corresponds to the next syllable (as argued by Schegloff, 2000), and whether the listener's native language influences his or her ability to determine the target.

Conclusion

By reviewing the literature on turn-taking, we have identified three mechanisms that are involved in fine coordination during dialogue: the prediction of the interlocutor's turn, the preparation of an appropriate response, and the articulation of this response at the correct moment. Although we have a good understanding of how listeners predict their interlocutor's turn, we do not yet know how listeners may use these predictions to prepare and articulate a response. Previous research suggests the scope of preparation is fairly flexible: Listeners delay planning until the end of their interlocutor's turn in some instances but can plan in advance and buffer this response in others. However, research has yet to examine what factors influence this flexibility and whether buffering interferes with the concurrent comprehension of the speaker's turn. But regardless of when listeners prepare their response, they will need to articulate it at the correct moment. Current theories argue listeners may identify this moment on the basis of syllabic entrainment. The syllable does indeed play important role in the segmentation of syllable-based languages (such as French) but not stress-based languages (such as English). As a result, it is unclear whether syllables do indeed underlie entrainment and the listener's accurate articulation of his or her own response. Given these open issues, we argue that future research should not only focus on the role of prediction but should also consider response preparation and articulation and the interplay of these three processes during turn-taking.

Funding

Ruth Corps is supported by the Economic and Social Research Council (grant number ES/J500136/1). Chiara Gambi is supported by a Leverhulme Research Project Grant (RPG- 2014-253) to Martin Pickering and Hugh Rabagliati.

References

Almor, A. (2008). Why does language interfere with vision-based tasks? *Experimental Psychology*, *55*, 260–268.

Altmann, G. T. M., & Kamide. Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Babiloni, C., Carducci, F., Cincotti, F., Rossini, P. M., Neuper, C., Pfurtscheller, G., & Babiloni, F. (1999). Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study. *Neuroimage*, *10*, 658–665.

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*, 477–487.

Beattie, G. W., Cutler, A., & Pearson, M. (1982). Why is Mrs Thatcher interrupted so often? *Nature*, *300*, 744–747.

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition* (pp. 181–216). Orlando, FL: Academic Press.

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.

Bohan, J., Leuthold, H., Hijikata, Y., & Sanford, A. J. (2012). The processing of good-fit semantic anomalies: An ERP investigation. *Neuropsychologia*, *50*, 3174–3184.

Bohan, J., & Sanford, A. (2008). Semantic anomalies at the borderline of consciousness: An eye-tracking investigation. *Quarterly Journal of Experimental Psychology*, *61*, 232–239.

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*, 295–311.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*, B13–B25.

Brown, C. M., Hagoort, P., & Ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing. Open- and closed-class words. *Journal of Cognitive Neuroscience*, *11*, 261–281.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and Aging*, *17*, 85–100.

Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 886–899.

Cummins, F. (2002). On synchronous speech. *Acoustic Research Letters Online*, *3*, 7–11.

Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, *31*, 139–148.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, *37*, 16–28.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language specific comprehension strategy. *Nature*, *304*, 159–160.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in segmentation of French and English. *Journal of Memory and Language*, *25*, 385–400.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121.

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.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology*, *23*, 283–292.

Duncan, S. (1974). On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, *3*, 161–180.

Duncan, S., & Niederehe, G. (1974). On signaling that it's your turn to speak. *Journal of Experimental Social Psychology*, *10*, 234–247.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations—Signalling the status quo? *Current Opinion in Neurobiology*, *20*, 156–165.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*, 210–233.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1187–1199.

Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition*, *6*, 35–53.

Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, *41*, 29–47.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*, 181–218.

Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, *6*, 1–12. doi: 10.3389/fpsyg.2015.00751.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS One*, *10*, 1–24.

Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech and Language*, *25*, 601–634.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279.

Hjalmarsson, A. (2011). The additive effect of turn-taking cues in human and synthetic voice. *Speech Communication*, *53*, 23–35.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144

Jefferson, G. (1984). Notes on some orderliness of overlap onset. *Discourse Analysis and Natural Rhetoric*, *400*, 11–38.

Jungers, M. K., & Hupp, J. M. (2009). Speech priming: Evidence for rate persistence in unscripted speech. *Language and Cognitive Processes*, *24*, 611–624.

Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, *2*, 21–35.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156.

Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., & Riehle, A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology*, *245*, 15–26.

Konopka, A. E. (2012). Planning ahead: How recent experiences with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*, 143–162.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes, and representations. *Cognition*, *42*, 1–22.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in Cognitive Sciences*, *20*, 6–14.

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

Mädebach, A., Oppermann, F., Hantsch, A., Curda, C., & Jescheniak, J. D. (2011). Is there semantic interference in delayed naming? *Journal of Experimental Psychology: Learning, Memory, and Language*, *37*, 522–538.

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

Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, *3*, 1–9. doi:10.3389/fpsyg.2012.00376

Mehler, J., Dommergues, S., Fraunfelder, U. H., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behaviour*, *20*, 298–305.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*, *22*, 1173–1182.

Meyer, A. (1996). Lexical access in phrase and sentence production: Results from picture- word interference experiments *Journal of Memory and Language*, *35*, 477–496.

Piai, V., Roelofs, A., Rommers, J., Dahlslätt, K., & Maris, E. (2015a). Withholding planned speech is reflected in synchronized beta-band oscillations. *Frontiers in Human Neuroscience*, *9*, 1–10. doi: 10.3389/fnhum.2015.00549

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015b). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping*, *36*, 2767–2780.

Piai, V., Roelofs, A., & Schiefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language*, *64*, 404–423.

Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Language*, *40*, 147–165.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–225.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*, 97–131.

Rajah, M. N., Languay, R., & Grady, C. L. (2011). Age-related changes in right middle frontal gyrus volume correlate with altered episodic retrieval activity. *Journal of Neuroscience*, *31*, 17941–17954.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K., M., Head, D., Williamson A, &Acker, J. D.,(2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cerebral Cortex*, *15*, 1676–1689.

Riest, C., Jorschick, A. B., & De Ruiter, J. P. (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology*, *6*, 1–14. doi: 10.3389/fpsyg.2015.00089

Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1580–1598.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology*, *2*, 307. doi: 10.3389/fpsyg.2011.00307

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

Sanford, A. J., & Emmott, C. (2012). *Mind, brain, and narrative*. Cambridge, UK: Cambridge University Press.

Sanford, A. J., & Garrod, S. (1981). *Understanding written language: Explorations of comprehension beyond the sentence*. Chichester, UK: John Wiley & Sons.

Sanford, A. J., & Garrod, S. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, *26*, 159–190.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2010). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, *23*, 514–523.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends of Cognitive Science*, *6*, 382–386.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, *70*, 1075–1095.

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

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102.

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension—An fMRI study. *Cerebral Cortex*, *22*, 1662–1670.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*, E4687–E4696.

Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation speech planning in turn-taking. *Cognition*, *136*, 304–324.

Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from either…or. *Journal Experimental Psychology: Learning, Memory, and Cognition*, *32*, 425–436.

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

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interview. *Human Communication Research*, *11*, 139–169.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes*, *50*, 23–51.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, *30*, 415–433.

Torreira, F., Bögels, S., & Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Frontiers in Psychology*, *6*, 10–3389. doi: 10.3389/fpsyg.2015.00284

Van Berkum, J. J., Brown, C. M., Zwisterlood, P., Koojiman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443–467.

Van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*, 124–144.

Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, *1106*, 150–163.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropscyhologia*, *41*, 989–994.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701–702.

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

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P. , Schevon, C. A., McKhann, G. M. … & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron*, *77*, 980–991.