

THE ART OF CONVERSATION

Why it's harder than you might think

By Antje S. Meyer and Svetlana Gerakaki, Max Planck Institute for Psycholinguistics and Radboud University, Nijmegen, The Netherlands

Most people like to chat. It's pleasant to talk to your family over breakfast, and at work, you might go to the coffee room or water cooler mainly because you hope to bump into someone and have a little chat. These observations are consistent with scientific findings: As far as we know, conversation exists in all cultures (Levinson & Torreira, 2015). It is the most common form of using language and it is, of course, where children acquire their language.

What are conversations? A defining feature is that they consist of turns. As Levinson et al. put it, speakers adhere to a "one-at-a-time" principle: Speaker A says something and then B, then A again, or perhaps C, and so on. As the order of the speakers and the length of the turns are not fixed, these sequences cannot be pre-planned, but appear to evolve quite naturally. Importantly, the turns are tightly coordinated in time. Analyses of corpora of natural conversations in many different languages have shown that the most common gap length (the period of silence) between turns is around 300 ms. Thus, a turn by speaker A is usually followed by a turn from B or C within a third of a second. Occasionally, longer gaps are observed or speakers may talk simultaneously, but by and large, speakers tightly coordinate their utterances in time. This may contribute to the pleasant feeling of having a fluent, effortless conversation.

How do we manage to create this fluent succession of turns? Planning a single word, for instance the word *dog* to name a picture of a dog takes roughly a second. Approximately half of this time is needed to identify the picture, and remaining time, half a second or so, to retrieve the word from the mental lexicon (the speaker's store of the words they know; Indefrey & Levelt, 2000). Planning a full sentence, such as *the dog is chasing the boy* can easily take several seconds. Given these planning times, how can conversation be so fluent? In part, this is because we often do not say full sentences but use particles such *hm* or *oh* or just nod to signal consent or interest. Such back-channelling requires little or no linguistic planning but contributes much to the perceived fluency of conversations.

However, often a verbal answer is required. A waiter asking *What can I get you to drink?* needs more than a *hm...* How do we respond quickly enough? Levinson and Torreira propose that we are highly pro-active. We often need little verbal information to guess the intention of a speaker and even the content of their utterance. Seasoned restaurant goers know that a waiter starting a question with *What...* will probably ask about their wishes

(rather than, for instance, their holiday plans) and might anticipate that the question, posed at the beginning of the meal will concern drinks rather than, say, desert. Levinson and Torreira propose that we use the utterance and context to begin to plan a response as early as possible. As soon as we have understood the speech act (whether it is a question, statement, etc.) and the gist (the broad content) of an utterance, we begin to plan our response. Sometimes, we may even have fully planned what to say before the end of the preceding turn. In such cases, we store the plan in working memory until we feel that the end of turn is imminent and then launch it. This proactive planning allows interlocutors to minimize the gaps between their turns.

Experimental Support

This proposal is based not only on casual observation and analyses of corpora of utterances, but also on experimental findings. For instance, in a study by Boegels, Magyari, and Levinson (2015), participants heard sentences such as *Which character, also called 007, appeared in the famous movies?* or *Which character from the famous movies is also called 007?* The critical information needed to answer the question (007) appeared about 1.5 second earlier in the first than in the second question. If participants begin to plan their response as soon as the relevant information is available, the gap after the end of the question should be much shorter after early-cue than late-cue questions. This result was indeed obtained. Responses were faster by nearly 300 ms when the cue appeared early than when it appeared late. (Real quiz masters know about early planning, as they always formulate their questions in such a way that the clue to the answer appears at the very end of the question). Thus, the participants indeed began to plan their utterance as soon as they could. In this study, the participants' brain activity was recorded while they were listening to the answers. The recordings suggested that they not only began to think about the answer but actually retrieved the words as soon as they could. Other research has confirmed the general conclusion from this study: Speakers often begin to plan their utterance while still listening to the other person (Sjerps & Meyer, 2015).

Speaking and driving

These conclusions are in-line with our intuitions about conversation. However, we know from many studies that both listening to speech and planning speech require attention. Thus, carrying out the two tasks simultaneously should be quite difficult. For instance, studies using driving simulators have shown that producing simple utterances, such as route descriptions, interferes with indicators of driving performance, such as lane keeping and braking. Listening to such descriptions has similar, though sometimes less pronounced, effects. This shows that some of the attention required for optimal driving is absorbed by the linguistic tasks. Other studies have shown that people who differ in their attention skills (being more or less able to concentrate on the task) differ in their performance in simple linguistic tasks, such naming pictures or identifying words in noise (Jongman, Meyer, & Roelofs, 2015). Thus, speaking and listening both require attention. Importantly, it has

been shown that attention is not only required for thinking about the content of utterances, but also for the processing of individual words and grammatical structures.

Planning while listening

In these studies, speaking and listening tasks were combined with non-verbal tasks, such as driving a car. What happens when two linguistic tasks, listening and speaking, are combined? As both tasks require attention, one would expect them to interfere with each other. In fact, similar tasks interfere more with each other than dissimilar ones. As listening and speaking are similar in many ways, they should strongly interfere with each other (Meyer & Huettig, 2016). And this is indeed the case. For instance, speakers are slower to name pictures when they simultaneously hear words compared to hearing stretches of noise. This interference effect increases when the name of the picture and the heard word are related in meaning (as in *cat – dog*) rather than unrelated (as in *spoon – dog*; Schriefers, Meyer, & Levelt, 1990). These findings show that the spoken word competes with the word the speaker plans to say.

Similarly, in the quiz study by Boegels et al. speakers responded earlier when the cue (*ooʔ*) appeared early in the sentence than when it appeared at its very end. But the saving in response time was only 300 ms, whereas the time interval between the appearance of the cues in the early and late condition was much longer, 1.7 seconds on average. Clearly utterance planning after the end of the question, in silence, was far more efficient than planning during the question.

Listening while planning

Thus, speech planning is hindered by concurrent listening. The reverse also holds: Listening is hindered by concurrent speech planning. To illustrate, in a recent study in our lab, participants named pictures while hearing distractor words, which they were told to ignore. In a control condition, they only listened to the distractor words, without planning speech. After an intervening task, participants were unexpectedly tested for their memory of the distractor words. They heard a mixture of “old” distractor words and new words and had to decide whether they had heard each word before. Overall, the participants did not perform very well on this task. But importantly, the performance in the no-planning condition was above chance, while the performance in the planning condition was no better than chance. In other words, the participants did not know whether or not they had heard these words before. This pattern was replicated in a second study where participants were warned about the memory test. Again, performance was much worse for the planning condition than from no-planning trials. In short, planning to speak hampers memory for what is heard while planning.

Another study in our lab has shown that the mental processing of utterances is also affected by speech planning. For instance, listeners hearing a sentence such as *She spread her sandwich with...* expect words such as *jam* or *butter*. They are surprised when they hear

socks, and this can be seen in recordings of their brain activity¹. But when participants hear these odd sentences while at the same time preparing to say something, the surprise signal is much reduced. This demonstrates that listening is disrupted by concurrent speech planning. In sum, we can plan utterances while listening to others, but this comes at a price: Both speech planning and listening are less efficient than they are when done by themselves.

Planning in L2

These studies were carried out with adults using their native language. Speaking and listening in a second language are more effortful and require more time. This is true even for highly proficient L2 speakers. For instance, a fluent bilingual speaker of Dutch and English may name a picture of a cat in 800 ms in their L1, but require 1200 ms to name the same picture in their L2 (van Assche, Duyck, & Gollan, 2016). Understanding words also takes more time in L2: A native listener may be able to decide within 500 ms that the spoken word *cat* is a English word, but a fluent L2 speaker of English may require an additional 50 ms to do so. Combining listening and speech planning should therefore be even harder for L2 than for L1 speakers. Indeed, many L2 speakers will probably confirm that holding a conversation with L1 speakers is hard work. Understanding the other person is hard; formulating a contribution to the conversation is hard as well, and doing both things together is very hard indeed. In fact, they might not be prepared to speak soon enough as the typical 300 ms gap comes and goes, and somebody else may start speaking before they can.

Implications

Nevertheless, being able to hold a conversation is, of course, an important goal of many L2 learners. How might we support them in moving towards this goal? First, as L2 proficiency increases, as learners get better at understanding the second language and as their ability to express themselves improves, the ability to combine listening and speaking will improve as well. In other words, L2 learners need to practice listening to conversational speech in the second language (without participating in the conversation) and they need to practice producing everyday utterances, initially without the pressure of having to respond as fast as native speakers do in everyday conversations.

Second, it would be helpful to raise everyone's awareness of the facts outlined here: that speaking and listening are effortful and take time, more so in L2 than in L1, and that combining them is cognitively challenging, even for highly proficient speakers of a language. Students might be encouraged to take their time in a conversation, and to separate listening and speech planning as much as possible. More importantly perhaps, teachers and employers

1 Brain activity is measured with EEG. Surprise can be seen in a very specific change in the EEG. It is a negative signal, peaking about 400 ms after a stimulus. For this reason, it's known as an N400.

should know that native speakers of a language often plan utterances while listening, but that expecting the same from L2 speakers might just be too much to ask. Teachers should get used to leaving uncomfortably-long silent gaps after asking questions to the class and guard against jumping in to rephrase or answer their own questions. Finally, we should all keep in mind that speech planning takes real time, in addition to thinking, and that a slow response to a question probably does not mean that the speaker is “a bit slow”, but that they need a little extra time to find the right words to express their thoughts.

References

- 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.
- Indefrey, P. & Levelt, W. J. M. (2000) The neural correlates of language production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 845–865). Cambridge, MA: MIT Press.
- Jongman, S. R., Meyer, A. S., & Roelofs, A. (2015). The role of sustained attention in the production of conjoined noun phrases: An individual differences study. *PLoS One*, 10(9): e0137557. doi:10.1371/journal.pone.0137557.
- Levinson, C. S. & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, Article 731, 10–26.
- Meyer, A. S. & Huettig, F. (Eds.). (2016). Speaking and Listening: Relationships Between Language Production and Comprehension [Special Issue]. *Journal of Memory and Language*, 89.
- 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.
- 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.
- van Assche, E. Duyck, W., & Gollan, T. H. (2016). Linking recognition and production: Cross-modal transfer effects between picture naming and lexical decision during first and second language processing in bilinguals. *Journal of Memory and Language*, 89, 37–54.

Author Bios



Antje S. Meyer (PhD, 1988, Radboud University) is a professor at Radboud University and director at the Max Planck Institute for Psycholinguistics in Nijmegen, where she heads the Psychology of Language Department. Before taking up her appointments in Nijmegen (2010), she was a professor of psycholinguistics at the University of Birmingham UK (from 2000). She has worked on various aspects of psycholinguistics, and currently focusses on speaking and listening in dialogue and individual differences in language skills. Her work has been funded by the DFG, British Academy, ESRC, BBSCR, Nuffield Foundation, and NWO. She has supervised more than 30 PhD students and is the author or a co-author of more than 100 articles in peer-reviewed journals. She has co-edited five books.



Svetlana Gerakaki is a graduate student in her department.