Cognition 223 (2022) 105037

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Overrated gaps: Inter-speaker gaps provide limited information about the timing of turns in conversation

Ruth E. Corps [a,*], Birgit Knudsen [a], Antje S. Meyer [b]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] Radboud University, Nijmegen, The Netherlands

ARTICLE INFO

Keywords:
Conversation
Turn-taking
Gaps
Speech production
Corpus analysis

ABSTRACT

Corpus analyses have shown that turn-taking in conversation is much faster than laboratory studies of speech planning would predict. To explain fast turn-taking, Levinson and Torreira (2015) proposed that speakers are highly proactive: They begin to plan a response to their interlocutor's turn as soon as they have understood its gist, and launch this planned response when the turn-end is imminent. Thus, fast turn-taking is possible because speakers use the time while their partner is talking to plan their own utterance. In the present study, we asked how much time upcoming speakers actually have to plan their utterances. Following earlier psycholinguistic work, we used transcripts of spoken conversations in Dutch, German, and English. These transcripts consisted of segments, which are continuous stretches of speech by one speaker. In the psycholinguistic and phonetic literature, such segments have often been used as proxies for turns. We found that in all three corpora, large proportions of the segments comprised of only one or two words, which on our estimate does not give the next speaker enough time to fully plan a response. Further analyses showed that speakers indeed often did not respond to the immediately preceding segment of their partner, but continued an earlier segment of their own. More generally, our findings suggest that speech segments derived from transcribed corpora do not necessarily correspond to turns, and the gaps between speech segments therefore only provide limited information about the planning and timing of turns.

1. Introduction

A hallmark of everyday conversation is the tight temporal coordination between the speakers' utterances: Person A says something, and person B responds almost immediately. As Sacks, Schegloff, and Jefferson (Sacks, Schlegoff, & Jefferson, 1974, p. 700 f.) highlighted in their seminal paper on turn-taking in conversation, "transitions (from one turn to the next) with no gap and no overlap are common. Together with transitions characterized by slight gap or slight overlap, they make up the vast majority of transitions". Quantitative evidence consistent with this statement comes from corpus analyses. In particular, Heldner and Edlund (2010) found that the median inter-speaker gap duration in corpora of conversational speech in Dutch, English, and Swedish ranged between 110 and 130 ms, with modes (most common intervals) around 200 ms. Additionally, Roberts, Torreira, and Levinson (2015) reported a mean gap duration "around 200 ms" (p. 7) in English telephone conversations from the Switchboard Corpus (Godfrey, Hollman, & McDaniel, 1992). Finally, Stivers et al. (2009) found median gaps ranging from 0 to 300 ms in question-answer sequences across ten languages, with modes ranging from 0 to 200 ms. Laboratory studies suggest that planning and launching utterances cannot be accomplished in 300 ms or less (e.g., Ferreira, 1991; Indefrey & Levelt, 2004), and so theories of conversation agree that these gap durations suggest that speakers must begin to plan their utterances while still listening to their interlocutors (e.g., Corps, Gambi, & Pickering, 2018; Garrod & Pickering, 2015; Levinson, 2016; Levinson & Torreira, 2015).

In this paper, we take a close look at the corpus data that have been used to quantify inter-speaker gaps and the timing of turns in conversation. Our analyses show that inter-speaker gaps, defined in phonetic terms (see below for details), often do not occur between turns, as defined in linguistic theory, but rather between speech segments, which are only parts of turns. This means that these gaps do not necessarily provide valid information about the timing of turns, and consequently do not constitute a solid basis for theories about the planning of turns in conversation. Furthermore, we found that many of these segments were, on our estimate, too short to give the upcoming speaker enough time to

* Corresponding author at: Psychology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen 6525 XD, the Netherlands.
E-mail address: Ruth.Corps@mpi.nl (R.E. Corps).

https://doi.org/10.1016/j.cognition.2022.105037
Received 30 July 2021; Received in revised form 18 January 2022; Accepted 20 January 2022
Available online 2 February 2022

respond. Finally, we found that in many instances speakers indeed did not respond to the immediately preceding speech segment produced by their partner, but rather completed an utterance of their own.

Note that the goal of this paper is not to argue that turns or turn-taking do not exist. Nor do we claim that speakers in a conversation are not sensitive to the content and form of their partner's utterance. In fact, there is much evidence from linguistic analyses of conversations (e.g., Clark, 1996; Clark & Schaefer, 1989; Goodwin, 1981; Kendrick & Torreira, 2015; Levinson, 1983; Sacks et al., 1974; Schegloff, 1968; Schegloff, Jefferson, & Sacks, 1977; Schegloff & Sacks, 1973) and from laboratory studies (e.g., Corps, Crossley, Gambi, & Pickering, 2018) to suggest that speakers are sensitive to the timing and content of their partner's utterances. Instead, we argue that evidence about the timing of turns and of the underlying speech planning processes cannot be gleaned from analyses of inter-speaker gap durations. In what follows, we briefly review key studies on the timing of conversational turn-taking before describing our analyses.

Levinson (2016, p. 8) defines a turn as "the unit of conversational communication, expressing a speech act, averaging around 2 seconds in duration but highly variable; in spoken language, typically a phrase or clause grammatically and prosodically complete and pragmatically sufficient" (see also Ford & Thompson, 1996; Skantze, 2021). In Levinson and Torreira's (2015) model of conversational turn-taking, the production system (supporting speaking) and the comprehension system (supporting listening) are simultaneously engaged in conversation. In particular, the next speaker (B) focuses on determining the current speaker's (A) speech act and the gist of their utterance. B begins to plan their response as soon as they have identified the speech act and gist of A's turn, while simultaneously listening to A's turn and waiting for cues that signal that A has almost finished speaking. When there is sufficient evidence that the end of the turn is imminent, B launches the planned response. Thus, the model explains short turn gaps by suggesting that upcoming speakers are proactive and plan their utterances as soon as the response-relevant information has been provided.

Planning in this way would be useful for many highly scripted everyday interactions, such as sales talks (e.g., *Can I get you anything…*?), where there is a clear expectation about what the speaker is likely to say and when they are likely to finish their turn. However, other exchanges, such as informal conversation, are less constrained. The lower predictability of these exchanges may discourage speakers from planning early or predicting the turn end because their partner's utterance may end in unexpected ways, which could make a planned response inappropriate. Furthermore, research suggests that carrying out linguistic tasks in parallel is cognitively demanding (e.g., Barthel & Sauppe, 2019; Fairs, Bögels, & Meyer, 2018; Fargier & Laganaro, 2016), and so speakers may not have sufficient processing capacity to plan early.

Nevertheless, several laboratory studies have shown that upcoming speakers do plan their turns while listening, and that such early planning supports timely responses (e.g., Barthel, Sauppe, Levinson, & Meyer, 2016; Lindsay, Gambi, & Rabagliati, 2019; Sjerps & Meyer, 2015). For example, Bögels, Magyari, and Levinson (2015) found that participants answered questions more quickly when the critical information necessary for answer planning occurred early (e.g., *Which character, also called **007**, appears in the famous movies*; *Mean (M)* = 640 ms) rather than late (e.g., *Which character from the famous movies is also called **007***? $M = 950$ ms), suggesting that participants planned their answer early when it was possible to do so. Comparable results were found by Corps, Crossley, et al. (2018), who had participants answer questions that were either predictable (*Are dogs your favourite animal*?), so that participants could predict the speaker's final word and prepare before this point, or unpredictable (e.g., *Do you enjoy going to the supermarket?*), so that they could not. Participants answered more quickly when questions were predictable ($M = 379$ ms; Experiment 2b) than when they were unpredictable ($M = 536$ ms).

Although these studies support the hypothesis that speech planning and listening occur in parallel, participants' average response times (which are considered lab-equivalents to gap durations) were consistently longer than 200 or 300 ms. These long response times are not particularly noteworthy: In some studies, participants had to answer questions about knowledge or personal experience, which might involve complex memory search processes or decisions between alternatives. What is important, however, is that the average gain in the response latency in the early relative to the late cue condition was substantially less than the time difference between the two cues. For example, participants in Bögels et al.'s (2015) study responded about 300 ms earlier in the early than the late cue condition, but the early cues occurred on average 1700 ms earlier than the late cues. Thus, a full 1400 ms were "lost". A likely explanation for this lost time is that participants did not have sufficient time to fully plan their utterance during the preceding question, and so they had to engage in some planning after question end.

This conclusion is consistent with the results of two studies using experimental paradigms where participants had ample time to plan their complete utterance before responding to a spoken utterance. First, Meyer, Alday, Decuyper, and Knudsen (2018, Experiment 1) had participants answer polar questions (e.g., *Do you have a green sweater*?) about one of four objects displayed on their screen (e.g., a cake, a branch, a sweater, and a barrel). In the early cue condition, all objects had the same colour, and so the participant could begin planning their response as soon as they understood the colour adjective. As the adjective began on average 731 ms before the end of the question, participants had sufficient time to plan a *yes* or *no* response. In the late cue condition, however, the objects had different colours, and so participants had to wait for the object name and check whether it appeared in the colour specified before they could plan their response. In the early cue condition, participants managed to respond with an average latency of 215 ms; in the late cue condition, the average latency was slightly longer, 297 ms, but still close to the average gap durations reported for conversational speech.

In a second study (Brehm & Meyer, 2021; Experiment 2), participants named pairs of pictures together with a pre-recorded confederate. The confederate named the left object on the screen and the participant the right one. In one condition, the participant could see the confederate's object, and in another condition, it was occluded. The participant's object was always in view from trial onset, and so they had plenty of time to prepare their utterance while the confederate was preparing and articulating their own utterance. Participants were quicker to respond when they could see the confederate's picture than when they could not see it (234 ms vs. 247 ms), and when the confederate's picture name was disyllabic (214 ms) rather than monosyllabic (267 ms). Most importantly, response latencies were just above 200 ms in all conditions.

These two studies suggest that response latencies of around 200 ms can be obtained in experimental tasks when participants are given the opportunity to fully plan their utterances before initiating articulation. This conclusion is consistent with Levinson and Torreira's (2015) proposal, which distinguishes between planning a response and launching articulation. The question is, however, whether natural conversations allow speakers to fully prepare a response while listening to their partner. Specifically, are individual utterances long enough to provide the upcoming speaker with enough time to understand the utterance and conceptualise and plan a response that can be articulated with a gap of a few hundred milliseconds? This is the main empirical question that we set out to address.

When considering this issue, it is important to keep in mind that speakers do not always have to respond to the content of their partner's

immediately preceding utterance (although laboratory studies have typically assumed that they do). Knudsen, Creemers, and Meyer (2020) discussed how the use of backchannels (such as *uh* or *yeah;* called continuers in Conversation Analysis; e.g., Schegloff, 1982) contribute to the flow of conversation, without requiring participants to respond rapidly to the specific content of their partner's utterance. The forms and functions of backchannels have been widely discussed from linguistic and psychological perspectives (e.g., Bangerter & Clark, 2003; Clark & Krych, 2004; Tolins & Fox Tree, 2014). They indicate to the present speaker that they should continue talking either by proceeding in their narrative or elaborating it (e.g., Schegloff, 1982, 2000; Tolins & Fox Tree, 2016). Backchannels are relevant in the current context because they should be easy to plan and ready for launching when the end of the current speaker's turn is imminent (see also Heldner, Edlund, Hjalmarsson, & Laskowski, 2011). Furthermore, backchannels do not require the current speaker to respond to new conceptual content; instead, the current speaker is invited to continue or elaborate their narrative.

Knudsen and colleagues examined how often backchannels occurred in a Dutch and a German corpus of conversational speech, and how they were timed relative to the preceding utterance. They found that backchannels accounted for about 15–20% of the utterances. As expected, the utterances following backchannels were almost always continuations of the utterance the speaker produced before the backchannel. This means that in about 30–40% of the conversation, one person talks, while the other provides backchannels. In these instances, the issue of how to respond quickly to the specific content of the preceding utterance does not arise, either for the current speaker or for the speaker producing the backchannels.

If backchannels and the utterances following them are easier to plan than other utterances because they do not necessitate taking the content of the partner's utterance into account, one might expect the gaps preceding or following them to be shorter than the remaining gaps in a corpus. In Knudsen et al.'s study, this was indeed the case: Backchannels were produced on average 100 ms earlier than other utterances (i.e., utterances that did not begin with a filler or a *yes/no* particle) in the Dutch corpus, and on average 59 ms before other utterances in the German corpus. However, the gaps for the remaining utterances were still close to 0 ms, indicating that even utterances that were not backchannels were planned early and quickly enough to be produced at the offset of the partner's turn. Thus, the question of how speakers find sufficient time to plan them remains.

In the present paper, we took a further step in addressing this question by determining the distribution of utterances of different lengths and the links between them in sections of three corpora of conversational speech: the German Corpus, (GECO; Schweitzer & Lewandowski, 2013), also analysed in Knudsen et al. (2020), the Dutch corpus (Corpus Gesproken Nederlands; https://ivdnt.org/images/stories/producten/documentatie/cgn_website/doc_English/topics/index.htm) and the Santa Barbara Corpus of Spoken American English (Du Bois, Chafe, Meyer, Thompson, & Martey, 2000).

The analyses were based on the following rationale: A key assumption in the Levinson and Torreira model is that upcoming speakers swiftly identify the speech act and gist of their partner's turn, and then use the remaining time during the turn to prepare their utterance. We asked how much time upcoming speakers might need to do this, and whether most utterances in conversations were long enough to give the upcoming speaker sufficient planning time for their utterance. To elaborate, let us assume the upcoming speaker only plans the first word of their utterance before beginning to talk. Laboratory studies have shown that linguistic formulation and articulatory planning for a single word take at least 600 ms (e.g., Indefrey & Levelt, 2004). The words elicited in laboratory studies are typically picture names, and other types of words may be faster to plan. One might, for instance, expect that particles, which often appear at the beginning of utterances, are faster to plan than content words. However, Knudsen and colleagues did not find that utterances beginning with particles were initiated faster than other utterances. As a working hypothesis, we therefore assume that formulating and launching an utterance together take about 600 ms.

The formulation of an utterance must be preceded by the conceptualisation of utterance content. In other words, speakers need additional time to decide what to say. Most word production studies have used picture or definition naming to estimate conceptualisation time, and so the amount of time needed for conceptualisation in conversation is unknown. However, if an utterance is to be a response to the preceding utterance, conceptualisation cannot begin until enough of that utterance has been understood. We do not know how long understanding takes on average. In some cases, the upcoming speaker may have to listen to the entire utterance to understand its gist (as in *What's the name of the woman who served coffee at Bill's last year?*), whereas in other cases (as in *Coffee anyone*?) processing part of the first word of the utterance may suffice. Based on estimates of word recognition and decision times (e.g., Grosjean, 1980; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Marslen-Wilson & Tyler, 1980), it seems unlikely that understanding can be accomplished in less than 300 ms, even when, as in *Coffee anyone?,* only part of a single word needs to be understood and a *yes/no* answer is sufficient. Adding 300 ms to the 600 ms required for formulation and articulatory planning yields a total of 900 ms as the time upcoming speakers might need to respond to an interlocutor's utterance.

Thus, for speakers to achieve inter-speaker gaps of 200 ms, utterances need to be at least 700 ms long. In fact, in the German corpus analysed by Knudsen and colleagues, mean and median gap durations are close to zero ms. As a result, utterances had to be around 900 ms long to provide the upcoming speaker with sufficient planning time. Assuming median syllable durations of 200 to 250 ms, this means that the utterances had to be at least three or four syllables, or two or three words, long to provide the upcoming speaker with sufficient planning time for their utterance (e.g., Arnold & Tomaschek, 2016; Blaauw, 1995; Greenberg, Carvey, Hitchcock, & Chang, 2003a,b; Jacewicz, Fox, & Wei, 2010; Quené, 2008; Verhoeven, De Pauw, & Kloots, 2004; we provide the actual word durations in our corpora below).

Note that these estimates are based on results of laboratory studies of language comprehension and production. Recent psycholinguistic literature has stressed that prediction and priming processes play an important role in language comprehension and production (e.g., Huettig, 2015; Kuperberg & Jaeger, 2015; Pickering & Garrod, 2009, 2013). Prediction may not only be important for rapidly understanding individual words and phrases, but also for recognising the speech acts, which is critical for responding appropriately (e.g., Gisladottir, Bögels, & Levinson, 2018; Gisladottir, Chwilla, & Levinson, 2015). Such processes may have a stronger impact in natural conversation than in the lab, because priming may occur simultaneously on different processing levels, as proposed in Pickering and Garrod's model of conversation, or because interlocutors in a conversation can draw upon shared world knowledge and common ground, which are not available in laboratory settings (e.g., Arnold, Kahn, & Pancani, 2012; Brown-Schmidt, Yoon, & Ryskin, 2015; Galati & Brennan, 2010; Westra & Nagel, 2021).

How much these processes speed up language comprehension and production is currently unknown, but it is possible that speakers can often plan responses much faster than the laboratory estimates suggest. Alternatively, speech comprehension and planning may often be hindered in conversation, for instance by background noise or because the conversation takes place in parallel with another capacity-demanding

activity, such as driving or preparing a meal (e.g., Boiteau, Malone, Peters, & Almor, 2014). Thus the 900-ms/three words estimate can only be seen as a very rough calculation of the time speakers may need to respond to their partner. Regardless of its validity, establishing how much planning time speakers actually have during their partner's utterance is worthwhile because this information can constrain psycholinguistic theories of comprehending and producing speech in conversation.

Thus, our first goal was to obtain a quantitative estimate of how much time speakers have for response planning. Reviewing the phonetic literature, we found studies where the durations of words and utterances must have been measured, because they report how these durations were related to speech rate (Quené, 2008; Yuan, Liberman, & Cieri, 2006), pitch declination (De Looze, Yanushevskaya, Murphy, O'Connor, & Gobl, 2015) or pause duration (Heldner et al., 2011; Marklund, Marklund, Lacerda, & Schwarz, 2015). However, we did not find any studies that reported the median duration of utterances and their distribution in adult conversation. Therefore, obtaining this information for different corpora of conversational speech was of some interest in its own right. We expected that most of the time, upcoming speakers would have ample time to plan their utterances because most turns in adult conversations are probably longer than two words. Consistent with this intuition, Levinson (2016) mentioned that turns are typically around two seconds in duration, which would offer ample planning time. Additionally, Levinson and Torreira report a mean turn duration of 1680 ms, and a median of 1227 ms for the NXT-Switchboard corpus (Calhoun et al., 2010).

An important point to keep in mind is that we analysed same-speaker stretches of speech, which we refer to as segments (following Yuan et al., 2006). These segments do not necessarily correspond to turns or turn-constructional units (i.e., the building blocks of turns as defined in Conversation Analysis, such as a word, a phrase, or a clause; Sacks et al., 1974; Schegloff, 2007). This is because segments are defined purely by reference to time-stamped orthographic transcripts and speaker changes, without reference to their intonation and the action accomplished in the conversation, which are critical for the definition of turns and turn-constructional units (e.g., Schegloff, 2007, p. 3 f.) Nevertheless, segments have been used as proxies for turns in the phonetic and psycholinguistic literature (e.g., Bögels, 2020; Bögels et al., 2015; Bögels & Levinson, 2017; De Looze et al., 2015; Holler et al., 2021; Knudsen et al., 2020; Levinson, 2016; Yuan et al., 2006), and so they provide some insight into the processes involved in turn-taking. Here we follow this practice, but consider its merits in the General Discussion. We also return to the crucial difference between segments and turns in the section on Parallel Talk. We first report the distribution of segment length and duration, since our first goal was to determine whether speakers have sufficient time to plan their utterances.

2. Length and duration of segments in German, Dutch, and English

In this section, we show the distributions of the length (in number of words) and duration (in milliseconds) of segments for parts of three corpora of conversational speech. We chose three corpora that we could readily access and had used in related work: the German GECO Corpus, the Spoken Dutch Corpus (CGN), and the Santa Barbara corpus of American English. The corpora differ not only in the languages, but also in the conversational settings: The speakers in the German corpus were strangers, who were recorded in the lab; the speakers in the Dutch corpus were also recorded in the lab but knew each other; and the speakers in the English corpus knew each other and were recorded in "the wild". Thus, the corpora comprise a variety of settings, such as the lab, at home or at work. By including these three corpora we could examine how similar the distributions of segments were in length across a range of conversations.

2.1. The German corpus

2.1.1. Materials

The German corpus (GECO, Schweitzer & Lewandowski, 2013) consists of 46 two-person dialogues. In 22 of them, the participants could not see each other (unimodal condition), and in the remaining 24 dialogues they were facing each other (multimodal condition). The analyses presented below, as well as those in Knudsen et al. (2020), concern the latter dialogues. The participants were eight women (20 to 30 years old), who did not know each other. Each of them talked to three different partners. A list of potential topics for conversation was provided, but the participants were free to choose other topics as well. Each conversation lasted for approximately 25 min.

Schweitzer and Lewandowksi transcribed and analysed the conversations using Praat software (Boersma, 2001). In the transcripts, speaker changes are traceable through specific participant numbers. The onsets and offsets of the speakers' segments are time-stamped. While working with this corpus, we discovered that occasionally, the speech signal and transcript were misaligned or words were not transcribed. When this occurred, the transcript was corrected accordingly.

As in Knudsen et al. (2020), we excluded 14% of the segments because they included noise or laughter, were unintelligible, or included an interrupted word or a repair. For the present analyses, we additionally excluded backchannels (23% of the segments) and segments consisting only of a filled pause such (such as *ähm, hm,* 1% of the segments). This left 13,481 segments (63% of the selected corpus) for the analyses. To determine the length of each segment, we counted the number of words, excluding filled pauses. To determine the total duration of a segment, we computed the time difference between its offset and onset (with filled pauses included).

2.1.2. Results

As expected, the segments varied greatly in length and duration, from 1 to 167 words, corresponding to 0.03 s (e.g., for a strongly phonetically reduced pronoun or preposition) to 48.14 s. 95% of the segments comprised of up to 24 words or were up to 7.2 s long. These segments are shown in Figs. 1 through 3.

The mean duration of all segments was 2099 ms, corresponding to 6.8 words with an average word duration of 250 ms. However, the medians were much lower: The median duration was 1010 ms, which corresponded to three words. As shown in Figs. 1, 2, and 3, the distributions were extremely skewed both across (Figs. 1 and 2) and within (Fig. 3) dyads, with short segments predominating. Of the 12,772 segments included in the word length analysis, 36% consisted of a single word (the mode), and 48% consisted of one or two words. In terms of duration, 19% of the 12,757 segments included in the duration analysis were shorter than 300 ms, 39% were shorter than 600 ms, and 48% were shorter than 900 ms. Fig. 3 shows the distribution of segment length, in words, for each dyad. As can be seen, the distributions are extremely skewed for all dyads.

As reported by Knudsen et al. (2020), the average and median gap duration in this corpus are close to zero milliseconds, meaning that speakers typically began to talk at the offset of the preceding segment. To do so, they must have begun to plan their utterance at least 900 ms before the offset. But in the corpus, about half the segments were shorter than 900 ms, suggesting that either our estimate of speech planning time is incorrect, or the speakers planned their utterances without responding to the content of the preceding segment. We return to these options

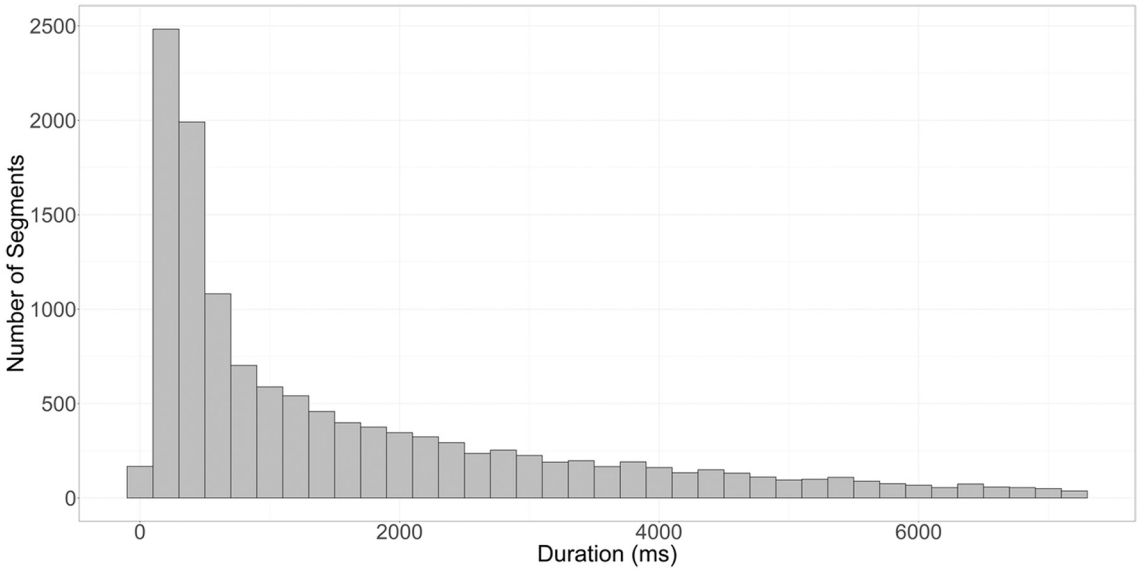

Fig. 1. The distribution of the duration of the segments in the German Corpus (GECO). Duration is placed into 200 ms time bins.

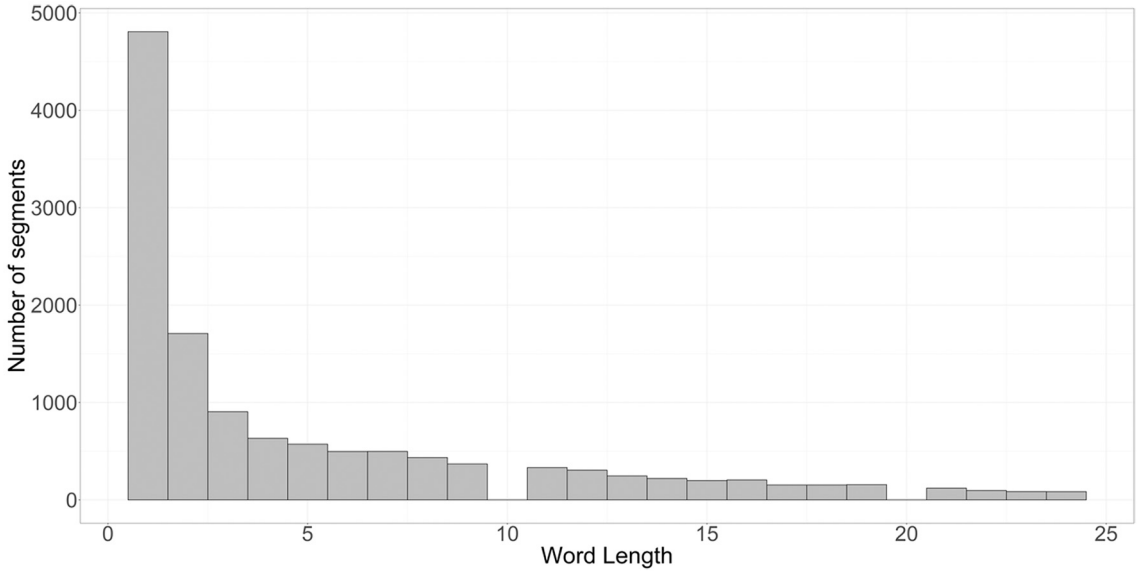

Fig. 2. The distribution of the word length of the segments in the German Corpus (GECO). Length is placed into one word bins.

below, but first report the analyses of the Spoken Dutch corpus and the Santa Barbara corpus of American English.

2.2. The Dutch corpus

2.2.1. Materials

The Dutch conversations analysed here were taken from the Spoken Dutch Corpus (Corpus Gesproken Nederlands, CGN; (https://ivdnt.org/images/stories/producten/documentatie/cgn_website/doc_English/topics/index.htm). We selected the first 18 face-to-face dialogues, where two family members or friends talked to each other without simultaneously performing a joint action (e.g., looking at a photo book, playing Scrabble). The speakers were 33 women and 9 men, who were at least 25 years old. Each conversation lasted for approximately 10–15 min.

For the analyses, we used time-stamped transcripts of the files, which are freely available in The Language Archive of the Max Planck Institute for Psycholinguistics Nijmegen, The Netherlands (https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_0000_0000_0001_53A5_2). The orthographic transcripts (available as ELAN files; https://archive.mpi.nl/tla/elan) provide word onsets, offsets, and durations. As in the German corpus, speaker segments were stretches of speech by one speaker.

We excluded 13% of the segments because they included noise, laughter, interrupted words or were unintelligible. We additionally excluded backchannels (12% of the segments) and segments consisting only of a filled pause (such as *bah, goh, pff, uhu*; 2%). This left 6,505 segments (73% of the selected corpus) for the analyses. Length and duration were computed using the same procedure as in the German corpus.

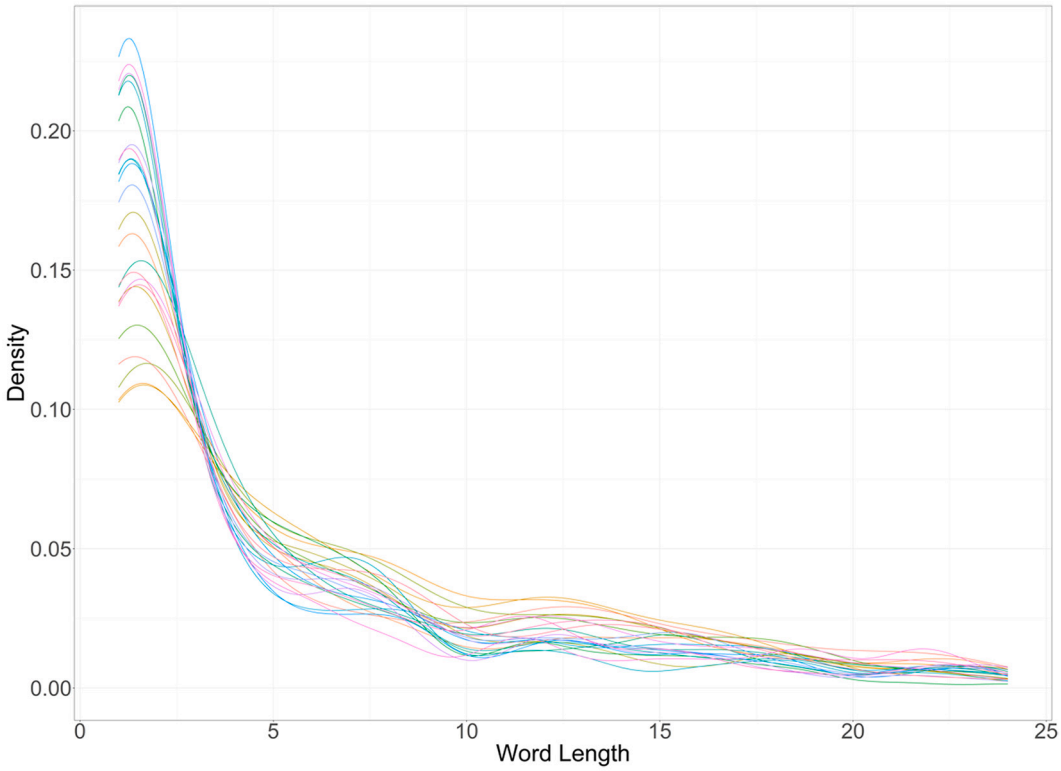

Fig. 3. The distribution of the word length of the segments in the German Corpus (GECO), plotted individually for the 24 conversations. Each coloured line represents a different conversation. Density represents the number of segments.

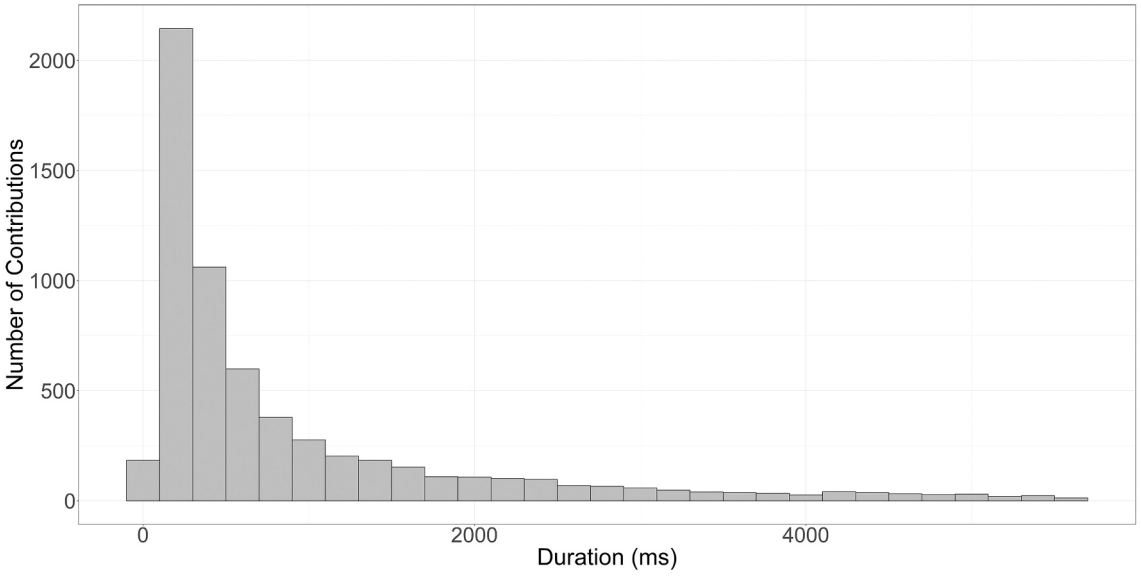

Fig. 4. The distribution of the duration of the segments in the Dutch Corpus (CGN). Duration is placed into 200 ms time bins.

2.2.2. Results

The length and duration of the segments varied from one to 99 words, corresponding to 0.04 s to 26.66 s. 95% of the segments included up to 17 words, or were up to 5.7 s long. These segments are shown in Figs. 4 to 6.

The mean duration of all included segments was 1297 ms, corresponding to 4.7 words with an average word duration of 276 ms. The median duration was 465 ms, corresponding to two words. As the figures show, the distributions were again extremely skewed towards short segments. In particular, 45% consisted of a single word (the mode), and 59% included only one or two words. In terms of duration, 36% of the segments were shorter than 300 ms, 57% were shorter than 600 ms, and 67% shorter than 900 ms. Fig. 6 shows that the distribution of turn length was very similar across pairs of speakers. Thus, much like in the

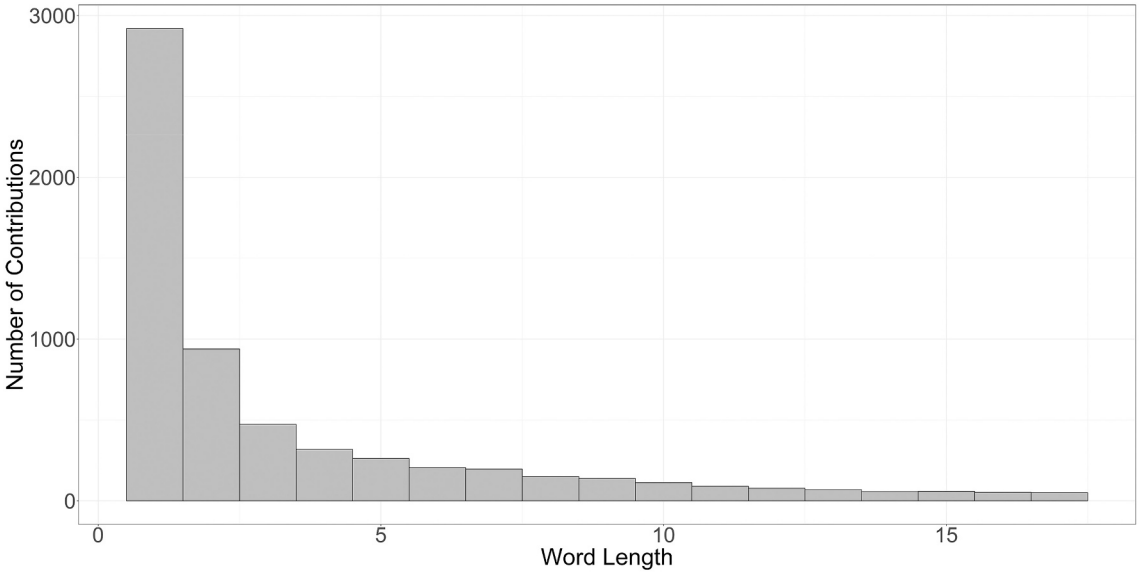

Fig. 5. The distribution of the word length of the segments in the Dutch Corpus (CGN). Length is placed into one word bins.

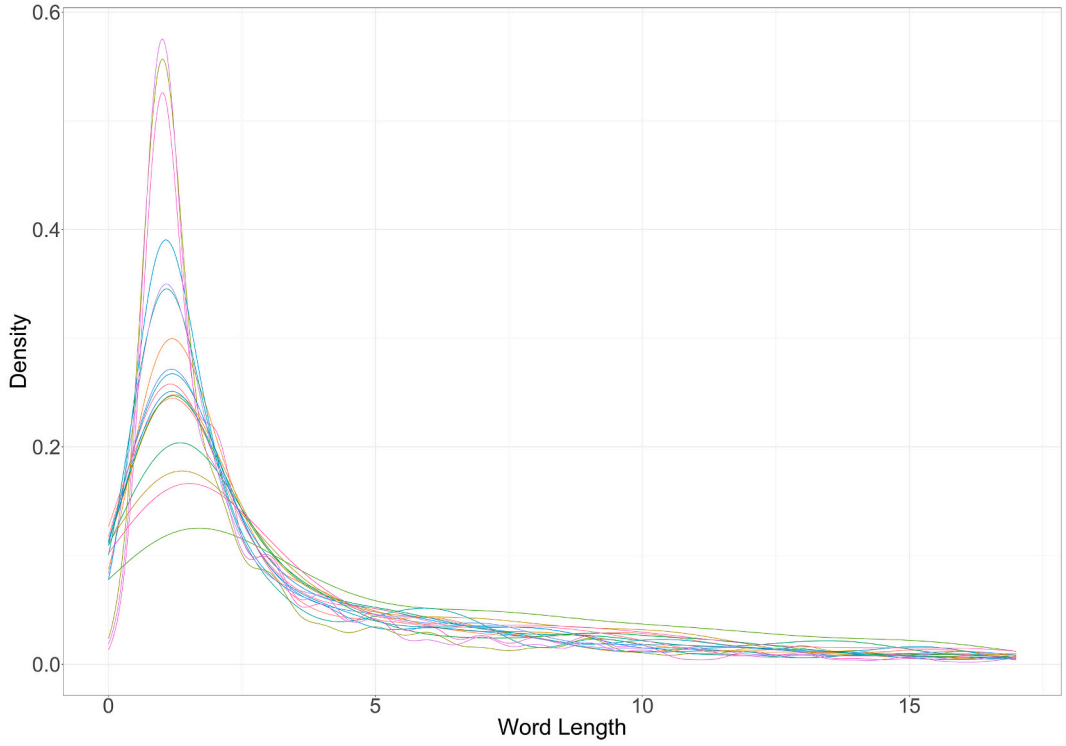

Fig. 6. The distribution of the word length of the segments in the Dutch Corpus (CGN), plotted individually for the 24 conversations. Each coloured line represents a different conversation. Density represents the number of segments.

German corpus, short segments predominated in the Dutch corpus.

2.3. The Santa Barbara Corpus of American English

2.3.1. Materials

The Santa Barbara Corpus of American English was prepared and made publicly available by Du Bois et al. (2000; accessible via https://www.linguistics.ucsb.edu/research/santa-barbara-corpus). The corpus includes 60 recorded and annotated spoken interactions between people from all over the United States. The transcriptions include face-to-face conversations, telephone conversations, task-related conversations (e.g., card games), on-the-job talk (e.g., sales encounters), classroom lectures, sermons, story-telling, town hall meetings, and tour-guide talks.

We were interested in instances where speakers took turns at talk without restrictions on the interaction and without strong expectation of what would be said, and so we focused on face-to-face conversations. For

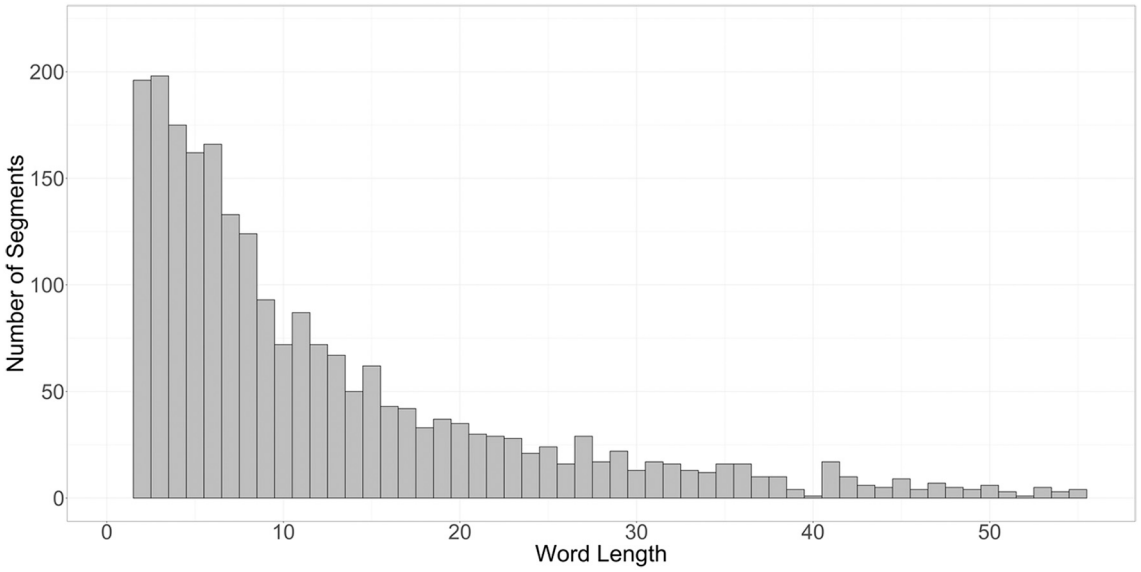

Fig. 7. The distribution of the word length of the segments in the English Corpus (Santa Barbara). Length is placed into one word bins.

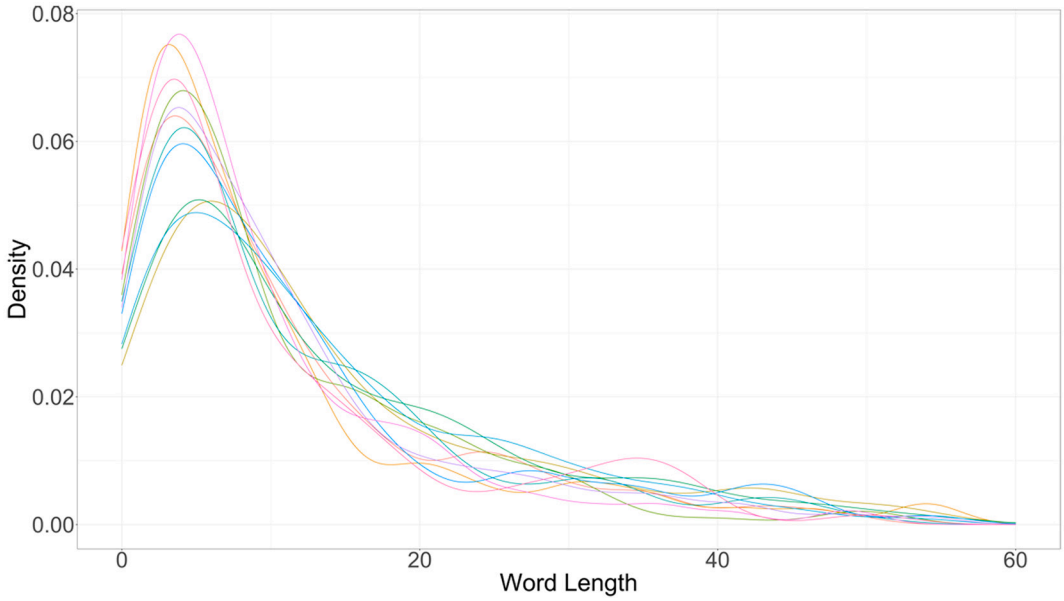

Fig. 8. The distribution of the word length of the segments in the English Corpus (Santa Barbara), plotted individually for the 11 conversations. Each coloured line represents a different conversation. Density represents the number of segments.

comparability with the German and Dutch corpora, we analysed conversations involving only two speakers. These conversations (11 in total) consisted of exchanges of dyads who were family members, friends, or colleagues. The participants talked about topics of their choice while they were recorded on Digital Audio Tape using small microphones. Each conversation lasted for 15–30 min (with an average length of 22 min).

We used the transcripts of the conversations, rather than the audio files. Each row in the transcript represents an intonation unit, which is "a stretch of speech uttered under a coherent intonation contour" (Du Bois, Scheutze-Coburn, Paolino, & Cummings, 1992, p. 17). Time stamps are included for the onsets and offsets of individual intonation units. But onset times could include pauses at the beginning of the unit, and so they do not correspond to the actual onset of speech. Thus, we could not calculate the precise durations of the segments or gaps, and only report the results in terms of number of words. A total of 12,185 intonation units were coded. We excluded 998 intonation units (8.19%) because they consisted exclusively of noise or laughter (coded as @) or were not understood by the transcriber (transcribed as "xxx").

We converted intonation units into segments by collapsing all intonation units produced in succession by one speaker in one segment. This yielded 3,190 segments. 17% of the segments were backchannels and were excluded from the length analysis, as were segments that consisted only of a filled pause, such as (such as *ah, hm*, *aw*, *uh*, *huh*, and *uhhuh*, 0.68% of the segments). Segments that consisted of both words and filled pauses (as in the utterance *And uh we had a cab driver*) were included in the analyses, but the filled pauses were not included in the word count. We analysed 2,635 segments in total.

2.3.2. Results

The segments were between 1 and, in an extreme case where a speaker conducted a monologue, 766 words long. The mean length of

the segments was 16 words and the median was eight words. 95% of the segments were 55 words or less. Figs. 7 and 8 show the lengths of these segments. As in the German and Dutch corpus, very short segments were the most frequent. In particular, 9% of the corpus consisted of a single word (the mode), 16% of one or two words, and 24% of up to three words. Fig. 8 shows that the 11 conversations were very similar in the distribution of segments of different length.

3. Discussion

The results of all three corpus analyses confirm the earlier claims that utterances in conversation vary greatly in length (Levinson, 2016; Sacks et al., 1974). In placing these results in the context of earlier linguistic work, it is important to bear in mind that we analysed segments, relying on time-stamped transcripts of the conversations, rather than turns. In the General Discussion, we return to the distinction between segments and turns. For now, the main point to note is that short segments predominated in all three corpora and in all conversations within the corpora.

Comparing the three corpora, we observe that the segments were shortest in the Dutch corpus. The median length was two words, corresponding to a spoken duration of 465 ms. The German corpus was similar, with a median of three words, corresponding to a spoken duration of 1010 ms. The difference in the spoken durations suggests that the German speakers spoke at a slower rate than the Dutch speakers. In the Santa Barbara Corpus, the median length of the segments was much higher (at eight words), suggesting that the speakers in this corpus delivered their messages in larger increments than those in the other corpora. Alternatively, the difference in segment length may be due to differences in the annotation and parsing of the corpora into segments. A potentially important difference is that for the Santa Barbara corpus, prosodic information was used to determine intonational units, which were subsequently concatenated into segments where appropriate. By contrast, the segments in the Dutch and German corpus were defined exclusively on the basis of temporal information, i.e., inter-speaker gaps. Importantly however, one-word segments were most frequent in all three corpora. Recall that backchannels and segments consisting only of filled pauses were not included in the analyses. If they were, the proportion of very short segments would be even higher.

In the Introduction, we reasoned that the duration of the current speaker's segment may constrain the utterance planning time for the next speaker. Thus, we might expect short segments to be followed by longer gaps than longer segments, which afford more planning time. We assessed this prediction for the German and the Dutch corpus, for which gap durations were available. In the German corpus, the average gap duration was −54 ms (SD = 305; N = 7396, excluding segments following backchannels) and there was a significant positive correlation between average gap duration and the duration of the preceding segment (r = 0.17, $p < .001$). In the Dutch corpus, the average gap duration was 66 ms (SD = 500 ms N = 4843), again excluding segments following backchannels) and the correlation was again positive (r = 0.26, $p < .001$). Thus, contrary to the prediction, longer segments were followed by longer inter-speaker gaps.

The results are, however, consistent with findings reported by Roberts et al. (2015), who examined determinants of gap durations in the Switchboard corpus. They found that shorter turns tended to be followed by shorter gaps, with the exception of very short turns (less than 700 ms), which were followed by relatively long gaps. Three quarters of these very short turns were backchannels, which we excluded from the analyses. Thus, the results of both studies are broadly consistent in showing a positive relationship between turn (or segment, in our case) duration and the following gap duration. As Roberts and colleagues discuss, turns of different length are bound to differ in many ways, including, for instance, syntactic complexity and the length and complexity of the response. Indeed, they showed that both of these variables impacted on gap durations. In short, while longer segments offer longer preparation times, they may also be harder to comprehend and respond to, resulting overall in a negative correlation between segment and gap duration. Further work, taking into account the content and structure of the segments, is needed to substantiate these suggestions. We focus on utterance content in the next section.

To return to our main argument, we suggested in the Introduction that segments shorter than about 900 ms, or three words, were unlikely to give the upcoming speaker enough time to respond. This estimate should hold even though short utterances are likely to be easy to comprehend and perhaps often require short answers. Our analyses suggest that this situation arises frequently: Roughly half of the segments in the German and Dutch corpus, and 17% of the segments in the English corpus included only one or two words. Our 900-ms/three word estimate was derived from results of laboratory studies, and may overestimate the time people need to understand and respond to utterances in conversation. But responding to one- or two-word utterances within 300 ms seems very taxing, even when comprehending and understanding utterances are facilitated through predictive or priming processes (e.g., Sjerps & Meyer, 2015). Nevertheless, the speakers did respond, and the analyses of the gap durations for the German and Dutch corpus showed that they did so very close to the utterance offset. How is this possible? We suggest that many segments were not responses to the immediately preceding segments, but were planned earlier and independently of the content of that segment. We motivate and assess this hypothesis in the next section.

4. Parallel talk

When coding the conversations, we noted that they often involved parallel talk, where each individual develops their turn in parallel with the other individual over several segments. An example is (1) from the Santa Barbara corpus, where Phil formulates a lunch invitation, while Brad talks about a third party, Pat, referred to as *her* (note that square brackets indicate overlap). Excerpt (2) is an example from the German corpus. Note that the numbers in the square brackets indicate the length of the overlap between speakers. Speaker 31 describes where they live (*und ähm…das kleine Dorf da neben Ehningen, da wohnen wir; and uhm… the small village next to Ehningen, there we live)* while Speaker 32 develops a question (*Und du fährtst eine dreiviertel Stunde? And you travel three-quarters of an hour?)*. For such stretches of parallel talk, the riddle how speakers manage to respond to segments that are too short to be responded to has a simple solution: Speakers do not actually respond to the immediately preceding segment, but instead produce a new segment of their own turn.

(1) PHIL: .. W- .. w- .. why don't you call me at least a little bit later [maybe, BRAD: [Yeah].
PHIL: and] we can [go do that].
BRAD: [Can I] do that? Cause I .. she'll be .. Uh..
PHIL: [Ji- .. Jim and I are gonna] have lunch,
BRAD: Uh .. I don't want to get her uh ..]
PHIL: I don't know if you have plans or not. But, we're gonna have lunch later at noon.

(2) 1 SPEAKER 32: Ja, (Yes,) [0.11].
2 SPEAKER 31: Also, (Well,) [−0.01].
3 SPEAKER 32: klar. (of course.) [−0.13].
4 SPEAKER 31: Kreis Böblingen und (district Böblingen and) [−0.2].
5 SPEAKER 32: Mhm. (Uhm) [−0.35].
6 SPEAKER 31: ähm…das kleine Dorf daneben Ehningen…da (uhm…the small village next to Ehningen…there) [0.08].
7 SPEAKER 32: Und (And) [−0.13].
8 SPEAKER 31: wohnen wir. (we live.) [−0.19]
9 SPEAKER 32: Du fährst eine dreiviertel Stunde? (you travel three-quarters of an hour?) [−0.12]
10 SPEAKER 31: Ja. (Yes.) [−0.02].

Such parallel talk has been described in linguistic studies of conversation, typically in discussions of turn taking rules and opportunities (e.g., Drew, 2009; Jefferson, 1986, 2004; Vatanen, 2018) and in the phonetic literature (e.g., Kurtić, Brown, & Wells, 2013; Kurtić & Gorisch, 2018). Here, we focus on how parallel talk is planned and how its occurrence may contribute to explaining the short durations of inter-speaker gaps in conversation. Informal inspection of parallel talk in our corpora suggests that the partners' utterances are often related to a common theme, but that successive segments are not adjacency pairs, with one segment being a response to the other. For instance, in the example from the German corpus, both speakers talk about Speaker 31's home town, but the utterance by Speaker 31 in line eight (*wohnen wir*) is not a response to Speaker 32's *und* in line seven. Instead, it is a continuation of Speaker 31's utterance in line six (*ähm…das kleine Dorf daneben Ehningen, da*).

In these instances of parallel talk, successive segments do not necessarily refer to each other and so the duration of one speaker's segment does not limit the next speaker's planning time for their segment. For instance, planning of the utterance in line eight most likely began well before the onset of the partner's utterance in line seven, and was not intended to be a response to that utterance. In other words, when speakers talk at the same time, it makes little sense to treat inter-speaker gaps as if they were gaps between turns. We return to the issue of inter-speaker gaps in the General Discussion, but we first examine how often such parallel talk arises.

To determine the occurrence of parallel talk, we conducted a second set of analyses where we coded whether or not each segment was a continuation of an earlier segment produced by the same speaker. We used a restrictive coding scheme that considered only certain lexical and syntactic properties of the segments. Same-speaker continuations are of central interest to our argument because they are, by definition, not a response to the partner's immediately preceding segment, and the issue of how speakers respond with the observed short gaps does not arise.

4.1. Method

All segments were manually coded as a same-speaker continuation (continuation hereafter) or as belonging to one of the other categories described below. Author BK coded the German and Dutch corpus, and RC the Santa Barbara corpus. Coding criteria were discussed and agreed in the team.

In conversations, there may often be links between adjacent segments that are obvious to the participants because they share common ground, but are not obvious to a coder of the transcript. To achieve optimal objectivity and reproducibility of the results, we defined continuations solely in lexical-semantic and grammatical terms. A segment was counted as a continuation if it contributed to completing a syntactically incomplete earlier segment. For instance, in (2), segments in lines four, six, and eight were coded as continuations because word meanings and grammatical structure clearly indicated that they belonged to the turn developed by Speaker 31. In addition, we coded segments as continuations when they were unambiguously linked by a pronoun or conjunction to a complete sentence of the same speaker. The use of these coding criteria means that we provide conservative estimates of the prevalence of continuations in our corpora.

For completeness, we describe how the remaining segments (those segments that were not counted as continuations) were categorised. First, there were **proceeds**, where Speaker A continued their narrative, as in continuations, but their first segment was grammatically complete and there was no pronoun or conjunction unambiguously linking it to their second segment (e.g., Speaker A: *Yes, I think it is incredible how big the differences are across grades,* Speaker B: *Yes, unbelievable*. Speaker A: *Somehow, some are almost full-grown whereas others are two heads smaller;* all examples are translations from utterances in the German corpus.). On a more lenient count, these segments would also be considered continuations.

Second, there were **direct responses,** which were answers to an interlocutor's question (as in line ten of (2)), expressions of (dis-) agreement (e.g., *That's right indeed* or *No, that was before my time*), literal repetitions of parts of the partner's utterances (e.g., Speaker A: *…in a boarding school.* Speaker B: *In a boarding school!*), segments referring directly back to the partner's preceding segment, for instance with a pronoun (e.g., Speaker A: *I don't have the ambition to speak flawless French one day.* Speaker B: *Which actually is almost impossible.*), or elaborations and associations (e.g., Speaker A: *my boyfriend's brother had a neighbor who used to cut his lawn meticulously,* Speaker B: *With nail scissors.*). Third, there were **questions** referring to the partner's segment asking for clarification or additional information (e.g., Speaker A:… *What do you do?*, Speaker B: *Eh, which sport?*); most of them would probably be categorised as repairs in Conversation Analysis (e.g., Albert & de Ruiter, 2018; Schegloff et al., 1977). Fourth, there were **content responses**, which were segments that were conceptually linked to the partner's segment, but were not obviously linked through pronouns or word repetitions as was the case for direct responses (e.g., Speaker A: *…and, if I want to teach there myself, I don't know.* Speaker B: *Well, what I thought was nice was that the groups were considerably smaller as compared to public schools.*). Finally, there were segments that introduced a **new topic** (e.g., Speaker A:*…I think people always have something to complain about.* Speaker B: *This sheet of glass is really weird.*). We list these categories in order to provide the reader with an impression of the nature of the segments that we did *not* consider to be continuations. The coding scheme is, of course, not meant to replace the far more detailed schemes developed in the linguistic literature, in particular in Conversation Analysis (e.g., Roberts et al., 2015; Schegloff, 2007).

To assess the reliability of the coding, three new coders, who were native speakers of the respective languages, independently coded the corpora. The Dutch and US corpora were recoded completely, while 11 of the 22 conversations were recoded in the much larger German corpus. Each coder was only asked to establish whether or not a segment was a continuation, using the criteria described above. We focused on continuations because the distinction between continuations and any response to the partner is the central to the purpose of this paper. The Cohen's kappa for the two coders was 0.84, 0.76, and 0.79 for the German, Dutch, and English corpora respectively. Thus, there was excellent agreement between coders (e.g., Landis & Koch, 1977).

4.2. Results and discussion

Table 1 shows the frequencies of each type of segment for each of the three corpora. As mentioned, 14% of the segments in the German corpus were excluded from analysis, and 23% were backchannels. We additionally excluded segments that were interrupted or not finished by the speaker and therefore could not be placed in one of the categories (1%). 43% of the segments were continuations, either following backchannels

Table 1
The number (*n*) and percentage (%) of segments in each of the coding categories in the German GECO Corpus, the Spoken Dutch Corpus (CGN), and the English Santa Barbara Corpus.

Category	GECO		CGN		Santa Barbara	
	n	%	n	%	n	%
Exclusions	3267	15%	1349	15%	187	6%
Backchannels	4956	23%	1045	12%	533	17%
Continuations-after-backchannels	4095	19%	832	9%	515	16%
Same-speaker continuations	5191	24%	3428	39%	430	14%
Proceed	501	2%	291	3%	129	4%
Direct responses	1942	9%	783	9%	749	24%
Content responses	802	4%	746	8%	179	6%
Repairs	789	4%	389	4%	386	12%
New topic	14	0%	14	0%	81	3%

(19%) or other segments by the other speaker (24%). In addition, 2% of the segments were proceeds, which one may want to also classify as continuations. 17% of the segments were responses to the other conversational partner.

In the Dutch corpus, 15% of the segments were excluded from the analysis. The results were similar as for the German corpus: 12% of the segments were backchannels and 48% were continuations, either following backchannels (9%) or other utterances by the other speaker (39%). 3% were proceeds and 21% of the segments were responses to the other speakers.

In the Santa Barbara corpus, 6% of the segments were excluded, either because they consisted exclusively of a filled pause, they were incomplete and so we could not identify the content of the segment, or they were produced at the start of the conversation and so there was no context in which to situate the segment. 17% were backchannels and 30% were continuations, either following backchannels (16%) or other utterances by the other speaker (14%). 4% of the segments were proceeds. In this corpus, the proportion of direct responses to the partner was 42% – higher than in the Dutch and German corpus.

In sum, the three corpora featured similar rates of backchannels and, consequently, continuations following backchannels. The German and Dutch corpus were also very similar in the rates of continuations that did not follow backchannels. Continuations occurred less frequently in the Santa Barbara corpus of American English. This corpus also featured, on average, longer segments than the other two corpora. Given that the three corpora were relatively small and differed in many ways, including the language, speaker characteristics, the settings in which they were generated, and the coding schemes used in the transcription, it is unclear how these differences arose. But nevertheless, the corpora do demonstrate that parallel talk regularly occurs in different languages and conversational settings. Speakers often build up their own utterances over two or more segments rather than responding directly to a segment produced by their partner.

5. General discussion

A core assumption in models of conversation is that speakers are proactive and plan their turns while listening to their interlocutor. Without such proactive behavior, speakers cannot achieve the well-attested short inter-speaker gaps (e.g., Levinson & Torreira, 2015). In the Introduction, we argued that speakers need to have *completely* planned the first part of their utterance before the end of the interlocutor's turn in order to respond promptly, within 300 ms, after the end of the turn. We suggested that it may take them about 900 ms (corresponding to two or three words in the partner's utterance) to do so. This estimate is derived from laboratory studies, and so is only a rough indication of planning time. Nevertheless, determining how much time people actually have to plan their responses is useful because this information can constrain theories of speech planning in conversation. To address this question, we analysed three corpora of conversational speech, in Dutch, German, and English, respectively. As highlighted in the Introduction, these corpora do not offer parsing of conversations into turns, as defined in linguistic theory, but into phonetically defined segments, which may or may not correspond to turns. We return to this point later in the General Discussion.

Consistent with earlier studies (e.g., Levinson & Torreira, 2015), we found that the segments varied greatly in length, ranging from one word to 766 words. In addition, we showed that the distributions were strongly skewed in favour of short segments in all corpora, and for each conversation within the corpora (see Figs. 1 to 8). This skew was most obvious in the German and Dutch corpora. In the German corpus, 53% of the segments were shorter than 900 ms, which, at an average word length of 294 ms, corresponded to just over three words. In the Dutch corpus, 67% of the segments were shorter than 900 ms, also corresponding to just over three words. In both corpora more than half of the segments included only one or two words. The Santa Barbara Corpus of American English featured longer segments. But even here, 17% of the segments included only one or two words, and 24% up to three words. Thus, regardless of whether the length or duration of the segments is considered, many segments gave the upcoming speaker only scant time to plan their response.

How, then, can we explain the short gaps between the segments? One possible answer is that many segments in the three corpora were not responses to the immediately preceding segment produced by the other speaker. Instead, they were linked to segments produced earlier by the same speaker. In other words, sometimes the speakers developed their turns in parallel over several segments, without immediately responding to intervening segments produced by the partner. This finding means that the length of the current speaker's segment does not limit the utterance planning time for the next speaker.

With respect to speech planning processes, such parallel talk is in some ways similar to using backchannels. Backchannels can shape conversations by encouraging speakers to continue unfolding their utterance plan or to elaborate on what they said before (e.g., Schegloff, 1982, 2000; Tolins & Fox Tree, 2016). In this regard, backchannels are different from sequences of parallel talk, where the speakers do not immediately respond to each other. However, just like the segments in parallel talk, backchannels do not provide novel conceptual content to be considered in planning the following segment. Combined, the rates of continuations after backchannels and of continuations in parallel talk added up to 51%, 48%, and 30% of the segments in the German, Dutch, and English corpora, respectively. In all of these cases, speakers continued their own utterances, rather than directly responding to the immediately preceding segment produced by the other speaker, and so their speech planning was not constrained by the content of this segment. As a result, the question of how speakers manage to respond to each other's utterances with near zero-gaps does not arise.

Note again that these considerations apply to segments, not turns. Turns are thought to be grammatically and prosodically complete and pragmatically sufficient (Levinson, 2016). But the transcripts of corpora we used did not include prosodic information, and whether or not a segment is pragmatically sufficient for the speakers can only be determined by considering the context in which it appears (e.g., Kendrick, 2015; Roberts et al., 2015). Given our interest in the planning of turns, this is an obvious limitation of our research. However, the results are informative, especially since segments have been used as proxies for turns in the phonetic and psycholinguistic literature (e.g., Bögels et al., 2015; Bögels & Levinson, 2017; De Looze et al., 2015; Holler et al., 2021; Knudsen et al., 2020; Levinson, 2016; Yuan et al., 2006).

Although informal inspection of the corpora suggests that successive segments in parallel talk may be unrelated, the turns developed by the two speakers often are related. In particular the speakers usually refer to a common theme, as illustrated in (1), where both speakers talk about Speaker 31's home town. Scholars working in the framework of Conversational Analysis have proposed that parallel talk (referred to as overlap or simultaneous talk; e.g., Schegloff, 2000) may arise from early turn-taking. As Drew (2009; see also Jefferson, 1986, 2004) discusses, upcoming speakers sometimes anticipate an end of turn and time their response accordingly, when, in fact, the current speaker is not yet ready to yield the floor. This leads to overlap in the speakers' speech. As Drew points out, such periods of overlap are frequent, but typically short as one of the speakers often 'drops out' when they realise that they are talking at the same time as their partner (e.g., Schegloff, 2000, for an extensive discussion of how overlap is resolved).

To substantiate this view, and more generally, to understand how speakers generate and comprehend parallel talk, more extensive analyses are needed than undertaken here. In our study, we determined the duration and length of the segments and determined whether or not they were same-speaker continuations. To assess whether or not the turns developed in parallel are thematically related, detailed semantic and functional analyses of the turns are required. Such analyses would also provide valuable information about the nature of the segments. From

the transcripts it is not clear whether participants simply speak in parallel, regardless of the other person's speech, or aim to produce their segments in alternation. Thus, one option is that both speakers talk simultaneously, happen to pause at the same time, and that the procedure used to segment the corpus detects a speaker switch. An alternative is that the segments correspond to planning units, which the speakers produce in alternation. In other words, though adjacent segments might not be directly linked in content, they might still be temporally coordinated. Further analyses might reveal that at least some of the segments correspond to turn-constructional units as defined by Sacks et al. (1974, p. 720). Based on informal inspection of the corpora, we expect that both of these scenarios occur some of the time.

Moving on from continuations and parallel talk to conversations more generally, our findings offer a more far-reaching answer to the question of how to reconcile the short gap durations in corpora of conversation with the likely longer speech planning times. We found that the transcribed and time-stamped corpora of spoken conversation yield information about speech segments, which may or may not correspond to turns. As segments are not necessarily turns, gaps between them only provide limited information about the planning of turns. To put this differently, analyses of gap durations only allow us to draw inferences about the planning of turns when the gaps actually occur between turns, not between stretches of speech that are parts of turns. This may be an entirely obvious point to linguists working with corpora of speech, but it may have been overlooked in part of the psycholinguistic literature on conversation. For instance, two studies are often cited as demonstrating short gaps between turns (e.g., Bögels, 2020; Bögels et al., 2015; Bögels & Levinson, 2017; Holler et al., 2021; Knudsen et al., 2020; Levinson, 2016): Stivers et al. (2009) and Heldner and Edlund (2010). Stivers et al. (2009) specifically investigated question-answer sequences in different languages, and so provided insight into turns. By contrast, the corpora generated by Heldner and Edlund (2010) include transcripts of conversations and temporal information about segments, but do not provide information about turns and therefore should not be cited as providing evidence for short gap durations between turns.

To understand how turns, rather than segments, are planned and timed, we need more comprehensive analyses of conversation. An obvious first step is to use corpora where the structure of the conversations (i.e., the parsing into turns and the types of turns) is annotated (e.g., Heeman & Lunsford, 2017; Kendrick & Torreira, 2015; Mertens & de Ruiter, 2021; Roberts et al., 2015; Skantze, 2021). Given the multimodal nature of the cues relevant to turn-taking, audio-visual corpora will be most informative (e.g., Holler, Kendrick, & Levinson, 2018; Holler & Levinson, 2019). Thus, a pressing issue is to find ways of objectively and efficiently defining the beginnings and ends of turns. Without such information, little can be said about the coordination of turns and their planning.

Other work essential for generating processing models of speaking and listening in conversation, and ultimately understanding the dynamics of conversation, concerns the conceptual, semantic, and pragmatic content of turns and the links between them. There is a rich linguistic literature on these issues, much of it in the framework of Conversation Analyses (e.g., Goodwin, 1981; Sacks et al., 1974; Schegloff, 1968; Schegloff et al., 1977; Schegloff & Sacks, 1973), but it is often difficult to bridge between linguistic theories of conversation and experimental psycholinguistics (for further discussion see also Healey, Mills, Eshghi, & Howes, 2018; Horton, 2017; De Ruiter & Albert, 2017; De Ruiter, Mitterer, & Enfield, 2006; Stivers, 2015). To illustrate, a central claim of Levinson and Torreira's (2015) model is that upcoming speakers begin to plan their utterance as soon as they have understood the gist of the present speaker's turn. Testing this hypothesis requires defining the gist of utterances. This is challenging because in casual conversation it is often not obvious, at least from a transcript, what the gist of a turn is. This problem was demonstrated empirically in a study by Bögels (2020), in which participants answered spontaneous interview questions. As in previous scripted studies (e.g., Bögels et al., 2015), participants answered earlier when the critical information necessary for preparation was available early rather than late. But importantly, two independent coders had difficulty identifying *when* the critical information that would enable answer preparation actually occurred (e.g., *007* in the question *Which character, also called 007, appears in the famous movies?*). In the first set of ratings, the two coders agreed on this answer word only 57% of the time. After discussion, the agreement reached 78%. These results demonstrate that it is not easy to identify the gist of utterances, even in question-answer exchanges where there should be a close relation between turns. Thus, the claim that upcoming speakers begin to plan a response as soon as they have identified the gist of the partner's turn is, at present, difficult to assess. In order to assess hypotheses about the time course of understanding the gist of utterances, further theoretical and empirical work is needed to clarify the notion of gist and provide clear criteria for its identification (e.g., Griffiths, Steyvers, & Tenenbaum, 2007, for a stimulating starting point).

Further research into the gist of turns is not only important to assess the specific claim that speakers plan their utterances as soon as they have identified the gist of their partner's turn, but is also essential for understanding how each person's speech affects their partner's speech planning. Turns in casual conversation can often be responded to in many different ways, and responses may be more inspired by the upcoming speaker's associations than by the partner's utterance. For example, a turn such as *We had great pizza last night* can be followed by responses such as *Where did you go?, Did Alice come?*, *That's so unfair. I had to work,* and so forth. This means that an upcoming speaker can begin to plan their response very early during the current turn, or even before its onset. This freedom in deciding what to say and when to plan it may greatly facilitate smooth turn-taking because the content and form of responses are not tightly constrained. However, work on mutual alignment in conversation has often proposed the opposite, namely that tight conceptual and linguistic links between turns facilitate speaking in conversation (Garrod & Pickering, 2004; Pickering & Garrod, 2009). Thus, for designing processing models of speaking in conversation it would be important to quantify how tightly linked turns in conversation actually are conceptually and linguistically (building, for instance, on work by Xu & Reitter, 2018), and to map out in detail whether and how conceptual and linguistic links affect utterance planning and turn taking.

To sum up, we have shown that substantial proportions of the segments in three corpora of conversational speech, held in Dutch, German, and American English, included only one or two words, giving upcoming speakers little time to respond. Further analyses showed that speakers often did not respond to the immediately preceding segment of their partner, but instead continued an earlier segment of their own. For these same-speaker continuations, the issue of how to respond on time does not arise. In such parallel talk, interlocutors' utterance segments alternate but they are not directly linked in content. Thus, conversation is not like ping pong in these cases. Instead, interlocutors develop their utterances in parallel. More generally, speech segments derived from transcribed corpora of conversation may not always be good proxies for turns and therefore the gaps between them may only provide limited information about the planning of turns.

Acknowledgements

We thank Natalia Cervantes for acting as second coder for the English corpus, Caitlin Decuyper for acting as second coder for the Dutch corpus, and Franziska Schulz for assisting with transcription of the German corpus.

Appendix A. List of backchannels in the German (GECO), Dutch (CGN), and English (Santa Barbara) corpora

Table A1
Backchannels in the German Corpus (GECO).

Backchannels
Generic
Aaah, Aah, Ach, Ah, Äh, Ähm, Aha, Ahahah, Ahh, Ai, Au
Buah
Eh
Ha, Hä?, Häh, Hah, Haha, He, Hm, Hm?, Hmh, Hmhm, Hmhmhm, Hmm, Hmmh, Ho, Hoho
Mh, Mh?, Mhh, Mhhh, Mhm, M-hm, Mhmh, Mm, Mmh, Mmm, Muah
Na, Ne, Nö
Och, Oh, Öh, Oha, Oho, Oooh, Ooooh
Pff, Psch
Uh, Ui
Woah, wow
Yeah
Single words
Achso
Cool
Doch
Eben, Echt? Ehrlich?
Geil, Gell?, Genau, Gott, Gut
Hurra
Ja, Ja? Jawohl
Klar, Klaro, Krass
Mega, Mja
Nagut, Naja, Ne, Ne?, Nee, Nee?, Nein, Nicht
Ok, Okay, Okay!
Schon, Schön, Stimmt, Super
Toll
Wahnsinn, Was?, Wirklich?
Yes
Combinations
Ach cool, Ach ja, Ach jee, Ach krass, Ach ok, Ach was, Achso cool, Achso ok, Ah cool
Ah ja, Ah gut, Ah klar, Ah nein, Ah ok, Ah schade, Ah super, Ahja ja, Ahja klar,
Ahja mhm, Äh nett, Alles klar
Cool interessant, Cool wow
Genau genau, Genau ja, Genau eben
Ha cool, Ha ja, Hach cool, Hm ja
Ja absolut, Ja cool, Ja doch, Ja eben, Ja fast, Ja genau, Ja guck, Ja gut, Ja hach, Ja ja,
Ja klar, Ja krass, Ja leider, Ja mega, Ja natürlich, Ja nee, Ja oah, Ja? Ok, Ja ok, Ja pf,
Ja schon, Ja sicher, Ja stimmt, Ja total, Ja übel, Ja vielleicht, Ja voll, Ja Wahnsinn,
Ja wahrscheinlich, Ja mhm, Ja naja, Ja wuah
Krass ja, Krass ok
Mh ja, Mh ok, Mh nee, Mh mhm, Mhm cool, Mhm genau, Mhm ja, Mhm klar, Mhm ok,
Mhm Mh, Mhm mhm, Mm cool, Mm ok, Mmh krass, Mmm ja
Na schade, Naja doch, Na toll, Nee nee
O krass, Och nein, Och ok, O Gott, O je, O Wahnsinn, Oh blöd, Oh Gott, Oh ja, Oh je,
Oh nein, Oh ok, Oh witzig, O ha, Oh süß, Ok cool, Ok hm, Ok ja, Ok krass, Ok oh, Ok gut,
Uh ja
Voll cool, Voll gut, Voll Schön, Voll witzig, Wie cool, Wow ok

Table A2
Backchannels in the Dutch Spoken Corpus (CGN).

Backchannels
Generic
Ach, Ah
Euha
Goh
Ha, Hè, Hu, Hum
Mmm
Oh, Oho
Pff
Uh, Uhm, Uhu
Single words
Getverdemme, God, Goed
Ja, Jawel, Jesus, Juist
Leuk
Mja
Nee
Oké
Precies
Tja, Tjee, Tjees, Tsja
Combinations
Ach bah, Ach jee, Ah ja, Ah joh
Hum ja, Hum oh
Ja goed, Ja hè, Ja hum, Ja ja, Ja joh, Ja nja, Ja nou, Ja precies, Ja uhm, Ja uhu, Ja zeker
M ja, Mm-hu, Mmm ja, Mmm jammer
Nee inderdaad, Nee joh, Nee nee, Nou ja, Nou nou
Oh God
Oh hum, Oh ja, Oh jee, Oh lekker, Oh mmm
Uh ja, Uh nee, Uh oké, Uh uh
Zo ja

Table A3
Backchannels in the English corpus (Santa Barbara).

Backchannels
Generic
Ah, Aw
Hm
Huh
Mhm
Oh, Oo
Uhhunh, Uhoh
Wa
Single words
Cool
Dorks
Exactly
Gee, Geez, God
Man
Okay, Mmkay
Really
Right
Wow
Yeah, Yep, Yes
Combinations
For sure
Hm yeah
I see
Oh gee, Oh God, Oh my God, Oh I see, Oh really, Oh shit, Oh well, Oh wow, Oh yeah
Poor Lisabeth, Poor Mom
That's cute, That's right
Yeah unhunh
You're kidding

References

Albert, S., & de Ruiter, J. P. (2018). Repair: The interface between interaction and cognition. *Topics in Cognitive Science, 10*, 279–313.

Arnold, D., & Tomaschek, F. (2016). The Karl Eberhards corpus of spontaneously spoken southern German in dialogues-audio and articulatory recordings. In C. Draxler, & F. Kleber (Eds.), *12. Tagung Phonetik und Phonologie im deutschsprachigen Raum. [12th meeting phonetics and phonology in German-speaking region.]* (pp. 9–11). Institut für Phonetik und Sprachverarbeitung, Universität München, [Institute for phonetics and language processing University Munich].

Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review, 19*, 505–512.

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science, 27*, 195–225.

Barthel, M., & Sauppe, S. (2019). Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science, 43*. https://doi.org/10.1111/cogs.12768

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01858

Blaauw, E. (1995). *On the perceptual classification of spontaneous and read speech*. Utrecht: LED.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341–345.

Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition, 203*. https://doi.org/10.1016/j.cognition.2020.104347

Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction, 50*, 71–89.

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

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.

Brehm, L., & Meyer, A. S. (2021). Planning when to say: Dissociating cue use in utteranc initiation using cross-validation. *Journal of Experimental Psychology: General, 150*, 1772–1799.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. *Psychology of Language and Motivation, 12*, 59–99.

Calhoun, S., Carletta, J., Brenier, J. M., Mayo, N., Jurafsky, D., Steedman, M., & Beaver, D. (2010). The NXT-format switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics, and prosody of dialogue. *Language Resources and Evaluation, 44*, 387–419.

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

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62–81.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259–294.

Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition, 175*, 77–95.

Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes, 55*, 230–240.

De Looze, C., Yanushevskaya, I., Murphy, A., O'Connor, E., & Gobl, C. (2015). Pitch declination and reset as a function of utterance duration in conversational speech data. In *INTERSPEECH 2015, Dresden, Germany* (pp. 3071–3075).

De Ruiter, J. P., & Albert, S. (2017). An appeal for a methodological fusing of conversation analysis and experimental psychology. *Research on Language and Social Interaction, 50*, 90–107.

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.

Drew, P. (2009). Quit talking while I'm interrupting: A comparison between positions of overlap onset in conversation. In M. Haakana, M. Laakso, & J. Lindström (Eds.), *Talk in interaction: Comparative dimensions* (pp. 70–93). Helsinki: Finnish Literature Society.

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., & Martey, N. (2000). *Santa Barbara corpus of spoken American English, parts 1–4*. Philadelphia: Linguistic Data Consortium.

Du Bois, J. W., Scheutze-Coburn, S., Paolino, D., & Cummings, S. (1992). *Disource transcription (Santa Barbara papers in linguistics)* (Vol. 4). Santa Barbara: University of California, Santa Barbara, Department of Linguistics.

Fairs, A., Bögels, S., & Meyer, A. S. (2018). Dual-tasking with simple linguistic tasks: Evidence for serial processing. *Acta Psychologica, 191*, 131–148.

Fargier, R., & Laganaro, M. (2016). Neurophysiological modulations of non-verbal and verbal dual-tasks interference during word planning. *PLoS One, 11*. https://doi.org/10.1371/journal.pone.0168358

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

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational and pragmatic resources for the projection of turn completion. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 135–184). Cambridge: Cambridge University Press.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35–51.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*, 8–11.

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

Gisladottir, R. S., Bögels, S., & Levinson, S. (2018). Oscillatory brain responses reflect anticipation during comprehension of speech acts in spoken dialog. *Frontiers in Human Neuroscience, 12*. https://doi.org/10.3389/fnhum.2018.00034

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

Godfrey, J. J., Hollman, E. C., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *, 1. 1992 IEEE international conference on acoustics, speech, and signal processing, 1992. ICASSP-92* (pp. 517–520). San Francicso, CA.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003a). Temporal properties of spontaneous speech—A syllable-centric perspective. *Journal of Phonetics, 31*, 465–485.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003b). Temporal properties of spontaneous speech – A syllable-centric perspective. *Journal of Phonetics, 31*, 465–485.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*, 211–244.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*, 267–283.

Healey, P. G. T., Mills, G. J., Eshghi, A., & Howes, C. (2018). Running repairs: Coordinating meaning in dialogue. *Topics in Cognitive Science, 10*, 367–388.

Heeman, P. A., & Lunsford, B. (2017). Turn-taking offsets and dialogue context. In *INTERSPEECH 2017, Stockholm, Sweden* (pp. 1671–1675).

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics, 38*, 555–568.

Heldner, M., Edlund, J., Hjalmarsson, A., & Laskowski, K. (2011). Very short utterances and timing in turn-taking. In *INTERSPEECH 2011, Florence, Italy* (pp. 2837–2840).

Holler, J., Alday, P. M., Decuyper, C., Geiger, M., Kendrick, K. H., & Meyer, A. S. (2021). Competition reduces response times in multiparty conversation. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.693124

Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review, 25*, 1900–1908.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Science, 23*, 639–652.

Horton, W. S. (2017). Theories and approaches to the study of conversation and interactive discourse. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (2nd ed., pp. 22–68). Routledge Press.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118–135.

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

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America, 128*, 839–850.

Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Human Studies*, 153–183.

Jefferson, G. (2004). A sketch of some orderly aspects of overlap in natural conversation. In G. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 43–59). Amsterdam, NL: John Benjamins.

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

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

Knudsen, B., Creemers, A., & Meyer, A. S. (2020). Forgotten little words: How backchannels and particles may facilitate speech planning in conversation? *Frontiers, 11*. https://doi.org/10.3389/fpsyg.2020.593671

Kuperberg, G. R., & Jaeger, T. F. (2015). What do we mean by prediction in language comprehension? *Language, Cognition, and Neuroscience, 31*, 32–59.

Kurtić, E., Brown, G. J., & Wells, B. (2013). Resources for turn competition in overlapping talk. *Speech Communication, 55*, 721–743.

Kurtić, E., & Gorisch, J. (2018). F0 accommodation and turn competition in overlapping talk. *Journal of Phonetics, 71*, 376–394.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics, 33*, 363–374.

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

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

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

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological Science, 30*, 504–515.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 133–156.

Marklund, U., Marklund, E., Lacerda, F., & Schwarz, I. C. (2015). Pause and utterance duration in child-directed speech in relation to child vocabulary size. *Journal of Child Language, 42*, 1158–1171.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1–71.

Mertens, J., & de Ruiter, J. P. (2021). Cognitive and social delays in the initiation of conversational repair. *Dialogue and Discourse, 12*, 21–44.

Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working together: Contributions of corpus analyses and experimental psycholinguistics to understanding conversation. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00525

Pickering, M. J., & Garrod, S. (2009). Prediction and embodiment in dialogue. *European Journal of Social Psychology, 39*, 1162–1168.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences, 36*, 329–347.
Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America, 123*, 1104–1113.
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*. https://doi.org/10.3389/fpsyg.2015.00509
Sacks, H. J., Schlegoff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–735.
Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist, 70*, 1075–1095.
Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Analyzing text and talk* (pp. 71–93). Georgetown: Georgetown University Press.
Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society, 29*, 1–63.
Schegloff, E. A. (2007). *Sequence organization in interaction. A primer in conversation analysis. 1*. Cambridge: 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*, 361–382.
Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*, 289–327.
Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. In *INTERSPEECH 2013, Lyon, France* (pp. 525–529).
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.
Skantze, G. (2021). Turn-taking in conversational systems and human-robot interaction: Review: A review. *Computer Speech & Language, 67*. https://doi.org/10.1016/j.csl.2020.101178
Stivers, T. (2015). Coding social interaction: A heretical approach in conversation analysis? *Research on Language and Social Interaction, 48*, 1–19.
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*, 10587–10592.
Tolins, J., & Fox Tree, J. E. (2014). Addressee backchannels steer narrative development. *Journal of Pragmatics, 70*, 152–164.
Tolins, J., & Fox Tree, J. E. (2016). Overhearers use addressee backchannels in dialog comprehension. *Cognitive Science, 40*, 1412–1434.
Vatanen, A. (2018). Responding in early overlap: Recognitional onsets in assertion sequences. *Research on Language and Social Interaction, 51*, 107–126.
Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech, 47*, 297–308.
Westra, E., & Nagel, J. (2021). Mindreading in conversation. *Cognition, 210*. https://doi.org/10.1016/j.cognition.2021.104618
Xu, Y., & Reitter, D. (2018). Information density converges in dialogue: Towards an information-theoretic model. *Cognition, 170*, 147–163.
Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *INTERSPEECH 2006, Pittsburgh, PA* (pp. 1–4).