

The conversational discourse unit: Identification and its role in conversational turn-taking management

Junfei Hu

JUNFEI.HU@UCLouvain.be

*Institute for Language and Communication
University of Louvain*

Liesbeth Degand

LIESBETH.DEGAND@UCLouvain.be

*Institute for Language and Communication
University of Louvain*

Editor: Junyi Jessy Li

Submitted 08/2022; Accepted 10/2023; Published online 11/2023

Abstract

This study investigates how discourse segmentation and turn-taking interact. Mapping syntactic, prosodic and pragmatic units, five types of conversational discourse units (CDU) were identified. Based on this segmentation, associations were examined between the syntactic, prosodic and pragmatic boundaries and turn-taking, as well as the transition speed after each type of CDU. Results show: 1) The relationships between the three linguistic boundaries and the occurrence of turn-taking were significant, and the association was the strongest for the pragmatic boundaries; it was weaker for prosodic boundaries and the weakest for the syntactic boundaries. 2) The type of CDU influenced the transition speed, with the pragmatic-syntax-bound CDU being fastest. The study highlights the importance of meaning-connection and earlier emergence of the utterance gist in timing turn-taking.

Keywords: Discourse segmentation, discourse unit, corpus analysis, pragmatic unit, turn-taking

1 Introduction

The linear unfolding of (spontaneous) speech results from a combination of different cognitive encoding and decoding processes (e.g. Auer, 2005; Clark, 1996; Pickering & Garrod, 2021). The resulting sequences of segments involve syntactic, prosodic and pragmatic dimensions, which are most often treated in isolation from one another. Yet, we believe it is important to acknowledge the interplay between these three aspects of linguistic encoding when defining the discourse segments that work as building blocks in spoken conversation. In the current study, we propose an operationalisation for segmenting spoken discourse into a sequence of conversational discourse units (CDUs). We see CDUs as a three-fold combination of syntactic, prosodic and pragmatic segments that are mapped onto one another in different configurations. It is our belief that getting

deeper insight into the distribution of CDUs in conversational discourse will improve our understanding of how speakers and hearers co-construct the ongoing conversation.

A key determining aspect of (spontaneous) conversation is that speakership alternates between interlocutors' space (Sacks, Schegloff & Jefferson, 1974). In a comparison of 10 different languages, interlocutors appeared to leave a gap between turns ranging from approximately 7 to 469 ms on average (Stivers et al., 2009), which is far shorter than the time needed to produce a single word (600 to 1,200 ms, depending on its frequency; Indefrey & Levelt, 2004; Levelt, Roelofs, & Meyer, 1999). The mechanism underpinning such smooth transition is knowingly complex and has yielded several competing accounts (De Ruiter, Mitterer & Enfield, 2006; Bögels, 2020; Bögels, Magyari & Levinson, 2015; Bögels & Torreira, 2015; Sjerps & Meyer, 2015). While there is general consensus that syntactic structure, prosodic realisation and pragmatic aspects of the utterance (i.e. the implementation of the communicative action) are involved in the turn-taking operation (Ford & Thompson, 1996; Levinson & Torreira, 2015), studies do not fully agree on the specific weights that syntax, prosody, and pragmatics should be given in such operation (Bögels & Torreira, 2015, 2021; De Ruiter, Mitterer & Enfield, 2006; Riest, Jorschick & De Ruiter, 2015).

Sacks, Schegloff and Jefferson (1974) proposed that conversation develops incrementally through the turn-constructive-unit (TCU), which intrinsically contains a recognisable boundary (i.e. transition-relevance-place, TRP) where transfer of speakership becomes saliently possible in interaction. According to their argument, a transitional gap between interlocutors can largely be attributed to the projectability of the TCU. That is, the boundary of the TCU should be relatively predictable in advance. According to Sacks, Schegloff and Jefferson (ibid: 709), 'sentential constructions are the most interesting of the unit-type... [they] are capable of being analyzed in the course of their production by a party/hearer able to use such analyses (i.e. analysing sentence in terms of its expandability) to project their possible direction and completion loci'. In other words, syntactic structure appears to be the dominant priority in Sacks, Schegloff and Jefferson's framework. Adopting the button-press paradigm, De Ruiter, Mitterer and Enfield (2006) reached a similar conclusion. Note that Sacks, Schegloff and Jefferson (1974) acknowledged that in addition to syntax, other language resources, such as prosody, also play a role in turn-taking management. Bögels and Torreira (2015, 2021) reported that the majority of participants (96%) in an elicited dialogue did not take turns at potential syntactic completion points if prosodic information (*f₀* and *duration*) signalled incompleteness. As a matter of fact, it appeared in a follow-up button-press experiment manipulating the prosodic features of the syntactic completion point that the prosodic information was useful and even more critical than syntactic information in turn-taking. Yet, it is not uncommon for speakers to not take turns at either the syntactic or the prosodic completion point. Speakers may actually take a turn at an incompleteness point of the syntactic unit, as long as that structure is interactionally complete (Li, 2016), or pass on the opportunity of taking a turn at the boundary of the syntactic unit (Houtkoop & Mazeland, 1985). Thus, in addition to the interplay between syntax and prosody, turn-taking should also be interpreted in its pragmatic context (Selting, 2000). The discrepant conclusions reached so far indicate that the three factors hardly create effects independently (Auer, 1996; Selting, 2000) and should be considered together.

The starting point of the present study is the surface analysis of conversational discourse (see Degand & Simon, 2009), that is, the syntactic structure, prosodic realisation, and pragmatic unit with their respective boundaries as key elements in deciding where and when a CDU starts and ends. Concretely, a CDU corresponds to a unit with coinciding syntactic, prosodic and pragmatic boundaries. In line with Degand and Simon (2009), we consider that a CDU is not complete as long as one of the three boundary types is open (awaiting completion). Thus, a CDU should not be restricted to the smallest syntactic, prosodic, or pragmatic unit, rather we consider CDUs as the segments resulting from the mapping between these three levels of analysis, giving rise to different types of CDUs (see Section 2). Speakers and hearers rely on this information to construct and interpret the ongoing conversation, including the management of turn-taking.

With the segmentation on the three dimensions and the identification of CDU in hand, the current study aims to explore:

- 1) which boundary—syntactic, prosodic or pragmatic—is the best predictor for the probability of a transition to arise.
- 2) the extent to which the type of CDU influences transition speed between turns.

As discussed above, turn-taking can occur after each type of linguistic boundary (i.e. syntactic, prosodic or pragmatic). This does not mean that every linguistic boundary will lead to a turn-taking in a real conversation. In some cases, interlocutors eschew the opportunity to take the turn when the linguistic boundary is present. Segmentation allows us to calculate how many instances of each type of linguistic boundary have and have not been followed by a turn-taking. Thus, generally, determining the linguistic boundary is the best predictor for the probability of a turn-taking to arise. Meanwhile, the end of the CDU is where the three linguistic boundaries converge. Hence, it is most likely to be followed by a turn-taking (Duncan, 1972; Ford & Thompson, 1996). However, each type of CDU's inner structure is unique in the way in which the syntactic, prosodic, and pragmatic combine. This motivates us to anticipate that the transition time after each type of CDU may be different.

Before we pursue the discussion on CDUs and their role in turn-taking management, we want to acknowledge that human communication is in principle multimodal (Holler & Levinson, 2019). Nonverbal symbols, such as gestures, are posited to coordinate with speech when conveying conceptual information (Kita & Özyürek, 2003; McNeill, 2005) and potentially contribute to timely turn-taking (Holler, Kendrick & Levinson, 2018; Kendrick, Holler & Levinson, 2023). However, because nonverbal semiotics cannot always be observed in language production, they are not the foci of this study.

Section 2 describes how this study defines CDUs. Section 3 introduces the dataset and delineates the methods informing CDU identification on the basis of syntactic, prosodic and pragmatic segmentation and determining turn-taking. Section 4 displays the study's results: significant associations were found between the occurrence of turn-taking and the three linguistic boundaries, which were differentially weighted in managing turn-taking. We also discovered that transition speed was influenced by the type of CDU. Section 5 encompasses a general discussion of the study's outcomes, highlighting that meaning-connections constructed during dialogues underpin the significant relationships found between the three linguistic boundaries and turn-taking management. We also speculate that the transition speed varies across the types of CDU, which may be attributed to the temporal difference in the emergence of the primary idea of the CDU (Figure 5). Section 6 finally presents the conclusions derived from the study.

2 Conversational discourse units

Regarding the role of segmentation in understanding the unfolding of the (linear) discourse structure, two perspectives can be roughly distinguished: (i) the interactional approach focusing on units-of-action (e.g. Ford, 2004; Szczepek Reed & Raymond, 2013), (ii) the descriptive linguistic approach focusing on defining units of linguistic analysis (e.g. Couper-Kuhlen, 2006; Degand & Simon, 2009; Izre'el et al., 2020; Pietrandrea et al., 2014, among others).

The interactional approach sees discourse as the outcome of a (collaborative) progressive construction of communicative action that can be (verbally) manifested as discourse segments corresponding to units of (collaborative) action. Sacks, Schegloff and Jefferson (1974) view spoken communication through the lens of the turn-taking system where interlocutors undertake initiating and responsive actions in turns. In this context, 'turns' are considered as the basic structuring components of conversation, made up of turn-constructive units (TCUs). Turns themselves are part of larger types of conversational units, such as *adjacency pairs* and *storytelling*, that play a role in the so-called sequence organisation (Schegloff, 2007: 9). The adjacency pair is organised around the actions of first- and second-pair parts (e.g. invitation-acceptance/rejection) and

storytelling is built around expressing a stance toward an event (Stivers, 2013). It is important to acknowledge that the TCU and its related units are inherently not defined in linguistic terms as the boundary of TCUs is defined largely in terms of turn-taking (i.e. a potentially complete turn) (Schegloff, 1996a; Selting, 2000: 478). Thus, such conversational units may be interesting in comprehending how conversation is constructed in terms of the formation and ascription of communicative action, yet they are not really fit as reliable units of linguistic analysis aiming at methodologically sound and descriptively adequate models of conversational structure.

In contrast, the descriptive linguistics approach relies heavily on linguistic cues and rules to determine the (spoken) discourse unit. Previous work has amply studied the linguistic nature of the ‘building blocks’ making up a piece of discourse, more specifically whether these units are based on syntax, prosody, pragmatics or any combination thereof (e.g. Degand & Simon 2009, Izre’el et al., 2020; Pietrandrea et al., 2014; Prévot et al., 2015; Steen, 2005).

One influential descriptive linguistic perspective, known as the coherence approach, views discourse as a cohesive structure that prioritises the (semantic) representation of the discourse structure and its organisation, with each discourse segment being related to another segment to build a coherent whole; for example, through dependent or interdependent relationships (see e.g. Rhetorical Structure Theory [Mann & Thompson, 1988] or the Geneva Discourse Model [Filliettaz & Roulet, 2002]). This perspective also (implicitly) claims that the discourse unit most often corresponds to a syntactic clause. For instance, when attempting to divide the discourse into units for further structural analysis, Mann and Thompson (1988) indicate that the ‘units are essentially clauses, except that clausal subjects and complements and restrictive relative clauses are considered as parts of their host clause units rather than as separate units’ (p. 248).

However, soundless syntactic input is barely imaginable, thus calling for prosodic input for any verbal interaction. As a matter of fact, it was shown that even in silent reading (Fodor, 2002) or in communication between signers (Brentari et al., 2018), prosodic information is used in syntactic decoding. Kiaer (2014) illustrated the interplay between prosody and syntax in determining the grammaticality of discourse units, thus showing that they were not independent of one another. In her example (1), Kiaer showed that in absence of prosodic information ({} stands for the prosodic phase), both sentence a and b were grammatical; however, when the prosodic information was taken into account, sentence b could not be considered grammatical because the intended constituency of ‘Caroline for her family’ was no longer secured due to the improper phrasing. This conclusion is similar to Carnie’s (2021), who highlights the importance of prosody when determining a verbless cluster constituency. In other words, discourse segmentation simply cannot exist without considering prosodic information.

(1a) Alison bakes cakes for tourists and {Caroline for her family}.

(1b) Alison bakes cakes for {tourists and Caroline} for her family. (Kiaer 2014: 8)

Adhering to the principle that ‘syntax and prosody should be considered as two independent but complementary sources for the characterisation of discourse units’ (Degand et al, 2014: 243), Degand and Simon (2009; also see Izre’el et al., 2020) provided an operationalisation for linguistically segmenting spoken discourse into reusable basic discourse units. They proposed to independently segment French speech into intonation units, on the one hand, and syntactic units, on the other. Subsequently, the mapping between these two levels of segmentation yielded several types of basic discourse units.

While their model accounted for both syntax and prosody and the interplay between them, a crucial aspect of human communication was left out, namely pragmatics. Conversation is, by its nature, a vehicle through which interlocutors implement and interpret their interactional plan (Levinson, 2013; Schegloff, 2007). For example, in Extract 1,¹ Speaker 1 is trying to describe the

¹ Each line represents a single intonation unit (IU). And in this study, we used the terms prosodic unit and intonation unit interchangeably. Each IU has been categorised as one of the four types according to its acoustic characteristics (Du Bois et al., 1993). The *period* is used to indicate a marked fall in pitch at the end of the IU. The *question mark* is used to

two attachments in the red oval in Figure 1 to Speaker 2, who needs to make a drawing of them.² Speaker 1 uses a simple sentence (l.2) with one intonation unit that has a syntactic structure similar to Speaker 2's interrogative utterance (l.1) to confirm Speaker 2's inquiry. It is common for speakers to show a confirmation by (partially) repeating their dialogic partner's utterance (Raymond, 2003; Schegloff, 1996b), and Speaker 1's response is thus, to a certain extent, driven by the aim of her communication—namely, confirming Speaker 2's inquiry. Speaker 2 then uses a simple sentence starting with 'and' to confirm the exact number of attachments and maintain the connection between her upcoming inquiry and the information already provided (l.6). Speaker 1 then produces a confirmative reply in one intonation unit (l.7) that is followed by a metaphor to make it easier for Speaker 2 to visualise the attachments (l.9). The most relevant data about the attachments given by Speaker 1 in lines 2, 7, and 9 are structured in three sentences (<> stands for one sentence) and three intonation units ({} stands for one intonation unit). However, Speaker 2 integrates all this information in one utterance that is realised in one intonation unit to display her understanding (l.10). We believe that the difference in the utterance-design in terms of their syntactic and prosodic realisation is (moderately) the result of the different communicative purpose held by these two speakers: Speaker 1 tries providing the information step by step, while Speaker 2 attempts to summarise, in one step, the information collected to finalise her drawing. The extract shows that syntactic (Tao & Hu, 2019; Thompson & Hopper, 2001) and prosodic realisations (Szczepek Reed, 2010, 2011) contribute toward realising the speaker's communicative intention (e.g. the purpose of the communication). In other words, the pragmatic aspect of spoken discourse should not be ignored. Indeed, several studies have indirectly noted that the understanding of discourse units should consider the pragmatic component of the utterance, arguing that 'units in the conversation must be understood as usable for construction of joint activity, not merely as packages of information to be parsed' (Ford, Fox & Thompson, 1996: 427) and that discourse units 'consist of an illocutionary act, a proposition, a clause, and an intonation unit' (Steen, 2005: 283).

Extract 1

- | | | |
|-----------------------|------------|---|
| 1 | Speaker 2 | <{And they're going to the left?}> |
| 2 | Speaker 1: | <{They're going to the left}>. |
| (3 lines are omitted) | | |
| 6 | Speaker 2: | <{And there are two of them}>. |
| 7 | Speaker 1: | <{Yeah}>. |
| 8 | | <{It's-} |
| 9 | | {Like he's sticking out two arms}>. |
| 10. | Speaker 2: | <{Sticking out two arms but on the same side}>. |

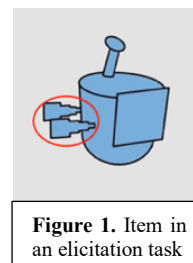


Figure 1. Item in an elicitation task

It should be noted that some discourse theories do emphasise that discourse analysis should account for the pragmatic aspects of the utterance, such as the speaker's intention and speech act. For instance, Segmented Discourse Representation Theory (Asher & Lascarides, 2003), which analyses the interactions between discourse interpretation and discourse coherence, proposes that the discourse segments are rhetorically related and that each rhetorical relation can be considered a type of speech act, like narration or elaboration. Lexical cues (e.g. referring phrases, [Grosz & Sidner, 1986]) and syntactic knowledge (e.g. clause, appositions and adverbials, [Afanenos et al., 2012]) can be used to detect the boundary of a discourse segment. Yet, a clearly elaborated operationalisation for determining the discourse segment is largely absent.

indicate a high rise in pitch at the end of the IU. The *comma* type is used to indicate (1) a pitch that rises slightly at the end of the IU, which typically begins with a pitch at the low or mid-level, or (2) a terminal pitch that either remains level or falls slightly but not far enough to be considered final. The *dash* refers to an IU that breaks off in mid-utterance. The *period* and *question-mark* IUs belong to the *final* type of IU. The *comma* and *dash* IU are labelled as to the *non-final* type (Ford & Thompson, 1996).

² Further information about the specific experimental dataset containing this extract will be provided in Section 3.1.

Reconciling the perspectives respectively prioritised by the aforementioned two schools, we extend upon Degand and Simon's (2009) operationalisation proposing that spoken discourse should be segmented into a set of conversational discourse units (CDUs) that integrate syntactic, prosodic and pragmatic units.

In addition to its linguistic analytical components, the concept of CDU also has a cognitive foundation that echoes Pickering and Garrod's (2021) model of language production in dialogue. Roughly speaking, they propose that to communicate an idea in dialogue, the speaker needs to go through the planning and implementing phases. The planning phase ('dialogue planner') involves the situation and game models. The situation model corresponds to the content of what the interlocutors are discussing. The game model represents the form of the contribution the interlocutor seeks to make (e.g. raising a question to seek information from your partner). Furthermore, 'each interlocutor has a representation relating to what the interlocutors are trying to achieve at that point in the dialogue' (Pickering & Garrod, 2021: 116). After finishing the planning stage, a production command will be sent to the implementing phase ('dialogue implementer') for linguistics encoding. In our view, the pragmatic unit corresponds to the communication plan generated in the game model, while the syntactic and prosodic (i.e. intonational) units are manifestations of the messages represented in the situation model. Along this line, CDUs can be considered as the outcome of the situation and game models.

3 Data and method

This section introduces the dataset and the means applied to determine the syntactic, prosodic (i.e. intonational), and pragmatic units. We then present the five types of CDUs we obtained, and display our methods of annotating turn-taking.

3.1 Dataset

This study is based on six face-to-face conversations between native speakers of American English. The conversations were conducted between two speakers, enabling us to eliminate the issue of speaker selection in a multi-party conversation. It also enabled us to ensure that the response of each conversation to the tailored initiation could be anticipated. Five of these face-to-face conversations were casual conversations between two acquainted adults of a similar age (female = 5 and male = 5; ~114 min). These conversations were obtained from the Santa Barbara Corpus of Spoken American English (SBC; Du Bois et al., 2000–2005). The criteria adopted in the current study for selecting conversations from SBC have been determined with two aims. On the one hand, the criteria were designed to ensure that the recorded conversations are as natural as possible. On the other hand, we tried to rule out the factors that might influence turn-taking time to ensure that the gap between turns can be primarily attributed to linguistic information and comprehension of utterance. First, even if naturalness is a continuum, casual talk has been acknowledged to be more natural than professional talk in terms of content and turn-design (Drew, 2013). Second, restricting the data set to conversations between two individuals rules out potential delays resulting from speaker selection (Sacks, Schegloff & Jefferson, 1974). Third, when interlocutors do not know each other or have a large age difference, more pragmatic issues may arise in language management, further influencing the turn-taking time. For instance, there may be more concerns about the preference of the content of the turn, as dispreferred turns, such as challenges and refusals, tend to receive a delayed response (Heritage, 1984). Note that there were seven conversations that fulfilled the criteria in SBC. The strong background noise (i.e. music and sounds from TV shows) made two of them non-eligible for analysis.

The sixth conversation was a task-oriented dialogue (38 min in total), in which two females, who were unacquainted at the time of data collection, performed a description-drawing task with images of 2D objects (adapted from Eijk et al., 2022; Rasenberg et al., 2022). The task-oriented dialogues were collected with the objective to explore interactive alignment at the discourse level

(Hu & Degand, 2022). The motivation to include a task-oriented dialogue in this study was to take advantage of its relatively clean sequential structure to facilitate conceptualisation and operationalisation of the notions of ‘action plan’ (Section 3.2.3.1) and ‘pragmatic unit’ (Section 3.2.3.2), which were then applied to the remaining five natural conversations. The results presented in Section 4 are based on the five SBC conversations.

Conversation (selected from SBC)	CDU	Token	Length (Token/CDU)	Turn
1	354	2,368	6.69	91
2	382	2,802	7.34	97
3	358	3,233	9.03	93
4	500	3,628	7.26	121
5	735	5,313	7.23	105
Total	2,329	17,344	7.45 (by average)	507

Table 1. Descriptive overview of the dataset.

3.2 Annotating syntactic, intonation and pragmatic units

3.2.1 Syntactic unit

The assumption held by all varieties of dependency grammar is that, within a clause, every element is embedded within binary asymmetrical relations called dependency relations, such that none remains isolated (De Marneffe & Nivre, 2019; Heringer, 1993; Tesnière, 1959). This is in line with the way people temporally build up their utterances: ‘as the speaker continues from point to point in interaction in real-time ... the units, bits and pieces that are uttered or added are not autonomous atoms but are subject to syntactic interdependencies’ (Linell, 2013: 62). Therefore, even if the notion of the syntactic unit is neither stable nor uncontroversial in current linguistics, the theoretical basis of the annotation of the syntactic unit in the current study is dependency syntax.

The starting point of analysis is verbal syntax in which a verb and its governed dependents (including core and oblique arguments, Nivre et al., 2016) are central. This analysis leads to the so-called ‘dependency clause’ that demonstrates maximal syntactic completeness. In addition to the verbal dependency clauses, we also consider nonverbal dependency clauses, in which the nonverbal head (e.g. noun, adjective) is modified (cf. Degand & Simon, 2009). In natural conversations, certain elements are not governed by the main verb but are semantically or pragmatically linked to the dependency clause. These elements, such as discourse markers (e.g. *and*, *so*, *but*, *frankly*, or *well*), primarily signal semantic or discourse relations between utterances or utterance-clusters, rather than contributing to the meaning of the proposition itself. These relations tend to occur at the peripheries of the clause (Degand & Crible, 2021) to manage the flow of thought and the structure of a conversation. While elements of dependency and embedding are considered as microgrammatical expressions, discourse particles, interjections (e.g. *oh*, *wow*), polite expressions (e.g. *please*), and vocatives (e.g. *Scott*) are considered as macrogrammatical expressions (Haselow, 2017). The linearization of these macrogrammatical expressions is not constrained by the relations of dependency, but rather by real-time cognitive activity, the requirement of text organisation, and the need to manage interlocutors’ relationship. Therefore, the current study treats these macrogrammatical expressions as independent syntactic units.

3.2.2 Intonation unit

In broad terms, an intonation unit (IU) is a stretch of speech produced under a single coherent intonation contour. Operationally, Du Bois et al.’s (1993) schema in Figure 2 indicates that an IU is generally marked by cues such as a pause along with an initial pitch reset and a lengthening of

the final syllable. The IU can represent a stretch of speech that conveys substantive content or regulates information flow (Chafe, 1993). Recently, Inbar et al (2020) have shown that the IU, as conceptualised in Du Bois et al.'s schema, is neurologically motivated. Although Du Bois et al.'s criteria to identify the boundary of IU have shortcomings (Barnwell, 2013; Barth-Weingarten, 2013) and alternative proposals on IU identification have been put forward in recent years, such as relying on the layperson's perceptive judgement, it remains a practical way to identify IUs, especially when the dataset is large. We have thus abided to this schema for the prosodic segmentation in the current study.

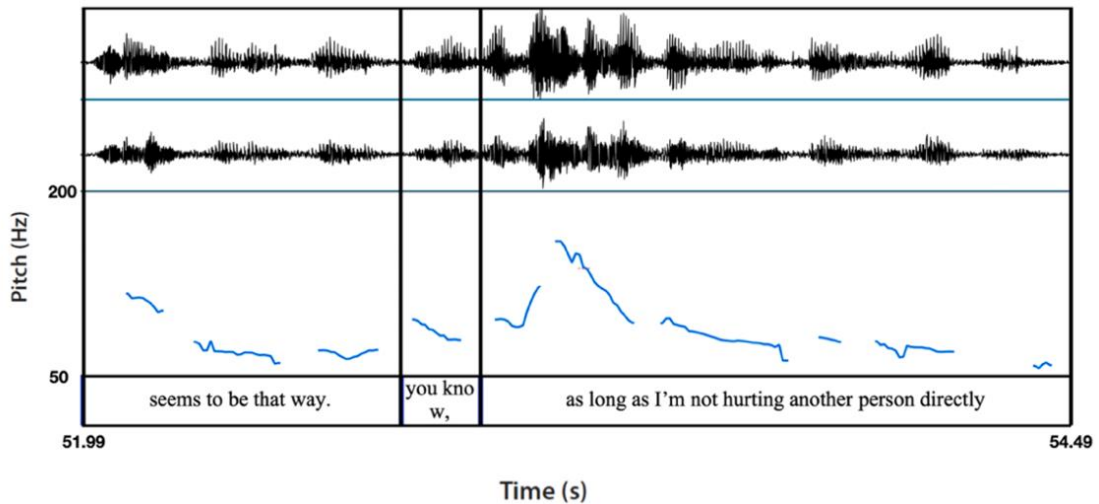


Figure 2. Illustration of intonation units.

3.2.3 Pragmatic unit

Section 2 discussed the need to include a pragmatic dimension when analysing conversational discourse and defining CDUs. This paper surmises that the speaker's communicative act originates from the primitive motive of achieving a goal, such as making a request, sharing ideas or seeking information, which are represented somewhere in the mind (Levelt, 1989; Pickering & Garrod, 2021). The utterance, then, is a verbal reflection of a conceptual and communicative goal. The listener is expected to put effort into understanding the speaker's intention (Grice, 1975, 1989) and reply to the speaker in a cooperative way. Most interlocutors have at least a (rudimentary) plan of their conversational intentions: either initiating action on the speaker's part or performing a responsive action on the listener's part. As such, we consider a pragmatic unit as a step towards implementing the speaker's action plan.

3.2.3.1 Action plan

In a nutshell, the action plan is the implementation of the speaker's interactive goal. Yet, assuming the existence of an action plan for any interaction by no means implies that interlocutors must have a full-fledged procedural implementation plan before initiating the interaction. In general, the way in which the plan is implemented emerges from the negotiation between interlocutors. As the interaction unfolds, the original motivation for initiating the conversation can be abandoned, replaced with an intention to do other things or co-exist with other motivations that arise during the conversation. The consequence is that interlocutors can suspend the conversation at any time, change the topic as the conversation unfolds or approach several topics simultaneously. Pragmatic units being the building blocks of the action plan, some pragmatic units can end up as incomplete, because the underlying action plan was aborted.

Next, it is important to consider that given the wide variety of ways in which speakers may implement their action plan, there is no exhaustive list of types of action plans (or pragmatic units). Lists of this type would be insufficient, by definition, to holistically explaining human action (Levinson, 2013), also because no specifically tailored practice exists to form a certain action (Enfield & Sidnell, 2017). It follows that the identification of the action plan and the pragmatic units should depend on the analysis of the situated context.

3.2.3.2 Identification of the pragmatic unit (PU)

To our knowledge, Ford and Thompson (1996) is the first study that directly tries to segment speech into pragmatic units. They suggest that the PU has ‘to be interpretable as a complete conversational action within its specific sequential context’ (p. 150) with a distinct boundary after which no further content is expected to follow. Accordingly, they expect PUs to end in a final intonation contour. As for how to practically determine a complete conversational action, Ford and Thompson (1996) mainly rely on the listener’s reaction. They distinguish local pragmatic boundaries from global pragmatic boundaries. Thus, a point where recipients respond in a non-taking-turn manner, e.g. by means of the backchannel *uh-hum* to show interest and encourage more details, should be treated as a local pragmatic boundary. At a global pragmatic boundary, no additional content is expected and genuine speakership alternation takes place. This approach contrasts with that of Steen (2005) who assumes that syntactic clauses contain sufficient information to convey an action (cf. Thompson & Couper-Kuhlen, 2005). There is hence overlap between pragmatic units and syntactic units, the latter constituting the minimal unit in dialogue that possesses sufficient information (i.e. complete proposition) to make a listener act (i.e. illocution).

Our approach to determining PUs was inspired by the aforementioned studies. We agree that a complete PU should be intact in meaning-expression and action-conduction. However, our operationalisation is intrinsically different from the aforementioned criteria. The identification of the pragmatic unit does not rely on the response of the addressee in current study (cf. Ford & Thompson, 1996). And a single pragmatic unit may be interpreted as having more than one illocutionary act. Therefore, we define the pragmatic unit as one step in the implementation of the action plan of the speaker rather than as the representation of one illocutionary act (cf. Steen, 2005). For instance, when the addressee simply replies ‘okay’ to the addresser’s instruction, the ‘okay’ can be an acknowledgement of understanding while simultaneously closing the sequence. In line with the approach advocated by Degand and Simon (2009), we made segmentation independent of the syntactic and prosodic information. To do so, we relied on the situated context of the utterance to determine what action it invoked and which PU boundary was applicable.

As shown in Table 2, utterances in 1.1 to 1.4, taken from the experimental dataset, form one pragmatic unit (marked in angle brackets). They are produced to describe the cup-shape in the red rounded rectangle. In 1.5, Speaker 2 acknowledges her understanding by means of the marker ‘okay’, which forms a complete pragmatic unit. From 1.6 to 1.17, four pragmatic units can be identified. From 1.6 to 1.10, there is one pragmatic unit describing the position and shape of attachment in the red rounded rectangle. The utterances from 1.11 to 1.13 do not provide any new information. Speaker 1 basically reformulates her utterances to make the description clearer. We treat these utterances from 1.11 to 1.13 together as a single pragmatic unit. From 1.14 to 1.16, the speaker provides further detailed description of the shape of the rectangle, indicating that the long bottom side of it is at an angle, to make the previous description more specific and clearer. After the additional explanation, utterance in 1.17 is produced to reiterate that the sticking-out shape is basically a normal rectangle. Speaker 2, who needs to draw the rectangle, feels confused by the information. She says ‘wait’ (1.18) to indicate that she has problems understanding and asks that the description be suspended. Utterances from 1.19 to 1.22 are used to conduct the interrogative action as a whole. They are considered as one pragmatic unit. As shown by the illustration, pragmatic units can be expressed by either a single word (1.5 and 1.18), a clause, (1.6 to 1.11) or a clause-combination (1.19 to 1.22). They can end with either a final or non-final intonation contour.

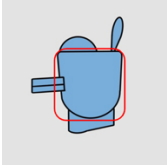
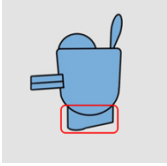
Interlocutors	Communicative action the pragmatic unit embodies	Contents of the task-oriented dialogue	Target
Speaker 1	Make a description	1. so? 2. there is a, 3. cup? 4. that is rounded at the bottom?	
Speaker 2	Acknowledgment	5. okay.	
Speaker 1	Make a description	6. and, 7. below the rounded part of the bottom of the cup, 8. there's a kind of a, 9. rectangle, 10. sticking out,	
	Clarify and summarise the previous description	11. at the bottom, 12. from the rounded part, 13. a rectangle sticking out,	
	Further explanation	14. and it's a pretty normal rectangle with the exception that- 15. the bottom side, 16. the long bottom side is at an angle.	
	Summarise	17. but other than that it's a regular rectangle.	
	Stop the description	18. wait.	
Speaker 2	Ask for a clarification	19. so, 20. it's a cup, 21. sitting on a rectangle like this? 22. or the rectangle is like that?	

Table 2. Example of the segmentation of the pragmatic units.

In addition to independence from syntactic and prosodic information, the operationalisation of the current study does not rely on recipient feedback to determine where a PU starts or ends. As noted before, the pragmatic unit corresponds to one step of an action plan. Since the plan in principle does not need to be full-fledged, people may in some cases fail to produce a complete pragmatic unit. In Extract 2, a face-to-face informal conversation, Karen encounters difficulties of verbalisation. Scott helps her finish the utterance. In that case, we treat Scott's utterance as a complete pragmatic unit. In contrast, since Karen fails to implement her action, her utterance in l.1 is identified as an incomplete pragmatic unit.

Extract 2 (Du Bois et al., 2000–2005)

- 1 → Karen: It wouldn't be as=,
 2 → Scott: It wouldn't be as aesthetically pleasing.
 3 Karen: Mhm.
 4 (TSK) <P Right P>.

An incomplete pragmatic unit may also result from one of the speakers ceasing talking when the interlocutors notice that their utterances are overlapping with each other (Sacks, Schegloff & Jefferson, 1974). In Extract 3, Karen utters 'would' (l.5) after hearing Scott's one-word response

(1.3). Evidently, however, she notes that Scott does not intend to alter the speakership. Thus, she aborts her production immediately, which results in the incomplete pragmatic unit with a truncated intonation (1.5).

Extract 3 (Du Bois et al., 2000–2005)

- | | | |
|---|----------|--|
| 1 | Karen: | I think we thought about doing that in the springtime, |
| 2 | | then I thought we'd replant them. |
| 3 | Scott: | Yeah, |
| 4 | | but [I] think there's more babies than ... we need. |
| 5 | → Karen: | [Would] -- |
| 6 | | Yeah. |
| 7 | | And we've got three new babies that I could replant, |
| 8 | | but I haven't, |

A further phenomenon that merits attention concerns instances in which the speaker either interrupts the progression of his/her own speech to fix or fine-tune issues, which is known as 'self-initiated' self-repair (Schegloff, 1977, 2013). We consider self-initiated self-repair in most of the cases as part of one pragmatic unit, because the repair is one step in implementing the action plan. Additionally, when pragmatic units are segmented, it is common to find that the speaker is quoting others' words. Regardless of the inner structure of the quotation, we consider that the quotation as a whole forms a single pragmatic unit (cf. Houtkoop & Mazeland, 1985; Jefferson, 1978).

3.2.4 Types of conversational discourse units

As discussed previously, syntactic, prosodic, and pragmatic units combine to form a conversational discourse unit. The example below illustrates the mapping procedure. In extract 4, Speaker 1 attempts to describe the cuplike shape illustrated by the rounded rectangle in red to Speaker 2, who is required to draw it.

Extract 4

- | | | |
|---|------------|---|
| 1 | Speaker 1: | The first shape it kind of looks like a cup with a bunch of stuff attached to it So like a little rounded at the bottom and you see it at a bit of an angle |
| 2 | Speaker 2: | Okay is the rounded part sticking out of the cup or is the cup |
| 3 | Speaker 1: | No it's the bottom of the cup shape is rounded |
| 4 | Speaker 2: | Is rounded at the bottom |
| 5 | Speaker 1: | Yes |

On the pragmatic level, this extract contains nine pragmatic units displayed in Table 3. The action plan of Speaker 1 is to instruct Speaker 2 to draw the cuplike shape. Meanwhile, the aim of Speaker 2 is to obtain information to draw the shape. Speaker 1 first provides a general description of the entire object (pragmatic unit 1). Then, she focuses on the cuplike shape and provides two pieces of information about it in two pragmatic units. One is that the bottom of the cup is rounded (pragmatic unit 2). The other is that the cup is placed at an angle (pragmatic unit 3). Then, Speaker 2 says, 'Okay', to convey her acceptance and understanding of the description, which forms one independent pragmatic unit. However, the description so far is insufficient to accomplish the drawing. Speaker 2 then asks for more specific information about the relation between the rounded part and the cup shape (pragmatic unit 5). Speaker 1 says, 'No', which reflects the sixth pragmatic unit in the extract to deny that the rounded part is separated from the cup. She then offers a clarification to further indicate that the bottom of the cup is rounded (pragmatic unit 7). Speaker 2 then paraphrases Speaker 1's description to confirm her understanding (pragmatic unit 8). Speaker 1 then says, 'Yeah', as a confirmative reply (pragmatic unit 9). As can be observed, each pragmatic unit is produced to implement one step of the speakers' action plans. The whole process of achieving the communicative goal is collaborative.

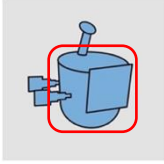
Interlocutors	Communicative action the pragmatic unit embodies	Contents of the task-oriented dialogue	Target
Speaker 1	Give a general description	1. The first shape it kind of looks like a cup with a bunch of stuff attached to it.	
	Further explanation	2. So like a little rounded at the bottom,	
	Further explanation	3. And you see it at a bit of an angle	
Speaker 2	Acknowledgment	1. Okay.	
	Request more information	2. Is the rounded part sticking out of the cup or is the cup shape itself.	
Speaker 1	Denial	1. No,	
	Clarification	2. It's the bottom of the cup shape is rounded.	
Speaker 2	Inquiry	1. Is rounded at the bottom.	
Speaker 1	Confirmation	1. Yeah.	

Table 3. Example of the segmentation of the pragmatic units.

Because of space limitations, we will focus on the utterances within the first pragmatic unit only to display the segmentations into intonation and syntactic units. Prosodically, two obvious pauses divide the utterances in the first pragmatic unit into three intonation units (Figure 3). On the syntactic level, there are two syntactic units (Figure 4), a nonverbal dependency clause, in which the nonverbal head (i.e. ‘shape’ in this case) is modified, followed by and a verbal dependency clause. Mapping between the three levels in this case gives rise to the pragmatics-bound CDU where one pragmatic unit corresponds to two syntactic dependency clauses and three intonation units.

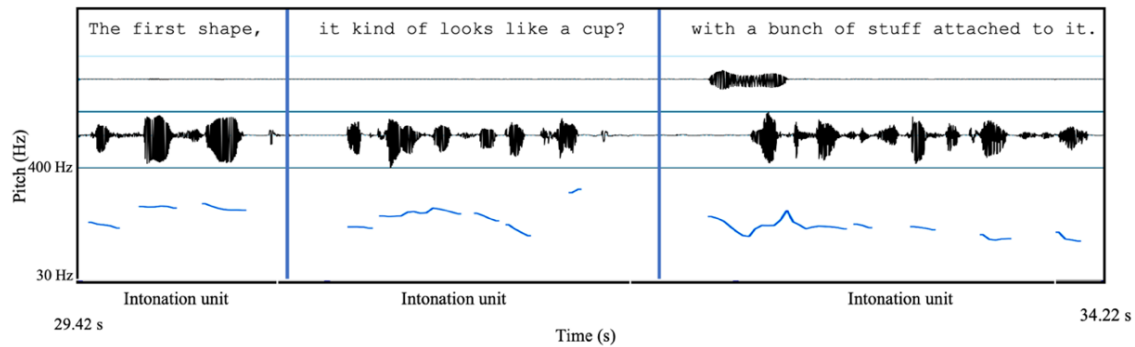


Figure 3. Example of the segmentation of the intonation units.

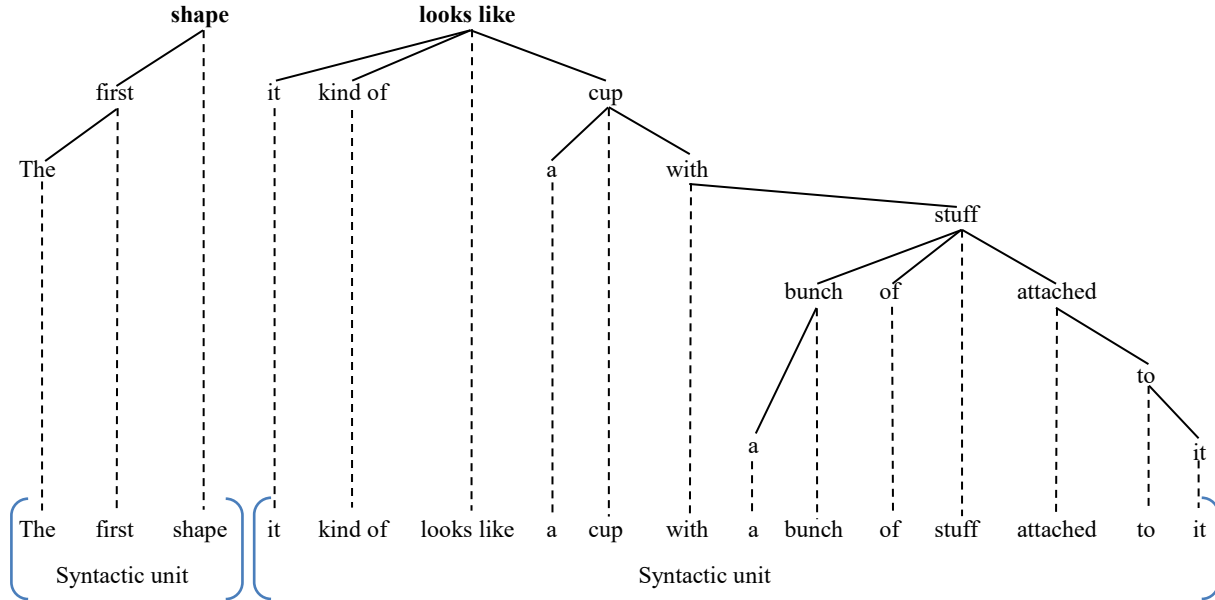


Figure 4. Example of the segmentation of the syntactic units.

Following the same criteria and procedures, Extract 4 can be segmented into a series of syntactic and intonation units. Table 4 presents the three-level segmentations of Extract 4 and the resulting type of CDU from the mapping.

Interlocutor	Pragmatic unit	Syntactic unit	Intonation unit	CDU Type
S1	Give a general description	1. the first shape	1. the first shape,	Pragmatics-bound
		2. it kind of looks like a cup with a bunch of stuff attached to it	2. it kind of looks like a cup?	
			3. with a bunch of stuff attached to it,	
S1	Further explanation	1. so	1. so like a little=	Pragmatics-bound
		2. like a little rounded at the bottom	2. rounded at the bottom,	
S1	Further explanation	1. and	1. and you see it at a bit,	Pragmatics-bound
		2. you see it at a bit of an angle	2. of angle,	
S2	Request more information	1. okay	1. okay,	Pragmatics-bound
		1. is the rounded part sticking out of the cup	1. is the rounded part,	
		2. or	2. sticking.	
		3. is the cup shape itself	3. out of the cup?	
			4. or is,	
S1	Denial	1. no	1. no,	Congruent
		1. it's the bottom of the cup chape is rounded	1. it's the bottom,	
S2	Inquiry	1. is rounded at the bottom	2. of the cup shape is rounded	Congruent
			1. is rounded at the bottom	
S1	Confirmation	1. yeah	1. yeah.	Congruent

Table 4. Three-level segmentations and CDUs of Extract 4.

Mapping between the three levels yielded five CDU types, with coinciding intonational ending, pragmatic boundary, and syntactic closure.³ The units listed below are the five types observed in the dataset we examined. Table 5 illustrates them with examples from the dataset.

- congruent type: one pragmatic unit corresponds to one syntactic dependency clause and one intonation unit;
- pragmatics-syntax-bound type: one pragmatic unit and one syntactic dependency clause correspond to two or more intonation units;
- pragmatics-prosody-bound type: one pragmatic unit and one intonation unit correspond to two or more syntactic dependency clauses;
- pragmatics-bound type: one pragmatic unit corresponds to two or more syntactic dependency clauses and intonation units;
- prosody-bound type: one intonation unit corresponds to two or more syntactic dependency clauses and pragmatic units.

There are several other approaches to defining and identifying discourse units (DU); however, each differs in some way from our approach. For instance, Prévot et al. (2015) attempted to analyse the discourse structures of spoken French and Mandarin Chinese by breaking up the discourse into utterance- or clause-like units. First, following Stede's (2012) elementary discourse unit definition, the syntactic information was used to obtain the semantic units, under the belief that the segment 'is usually a clause, but in general, ranging from minimally a (nominalisation) NP to maximally a sentence. It denotes a single event or type of events, serving as a complete, distinct unit of information that the subsequent discourse may connect to' (p. 89). Then, the segment was further refined based on the discourse and pragmatic completion considerations proposed in Ford and Thompson (1996). While the segmentation in Prévot et al. (2015) shows similarities with ours, their operationalisation of prosodic boundaries was different from ours, thus resulting in different unit structures. Further, Biber et al. (2021) and Egbert et al. (2021) both proposed a new method to identify DU based on their communicative purposes and claimed that DU had identifiable boundaries at which interlocutors shift to different communicative goals, that is, each discourse unit has one major communicative goal, for which they listed nine basic communicative goals, such as sharing feelings and evaluations, giving advice and instructions, and engaging in conflict. Even if they specifically emphasised the importance of taking the communicative action into account when segmenting DU, their "single communicative goal" concept was different from the pragmatic unit in the current study, which is defined as one step in implementing the action plan. Thus, Biber et al. (2021) and Egbert et al.'s (2021) 'single communicative goal' is similar to our 'action plan' (also see the idea of 'project' proposed by Levinson [2013]). Still, it is important to acknowledge that our 'action plan' is generated in one interlocutor's mind only, while their 'communicative goal' must be accomplished by all dialogic parties. Accordingly, they segmented the discourse into a series of 'topic-like units' (Prévot et al., 2015: 74) that included all interlocutors' utterances.

³ In theory, mapping the three dimensions should have yielded seven CDU types; however, we failed to observe the 'two-one-one' type (i.e. the 'syntax-prosody-bound' type, with two or more pragmatic units, one syntactic unit, and one prosodic unit) and the 'two-one-two' type (i.e. the 'syntax-bound' type, with one dependent clause and two or more pragmatic and intonation units). Given that one pragmatic unit corresponds to one step in the implementation of the communicative plan, and one intonation unit corresponds to one idea (Chafe, 1987, 1994), it is unsurprising that we could not identify any of the syntax-prosody-bound type, in which interlocutors attempt to stuff two ideas into one intonation unit. However, the syntax-bound type might still be observed in other datasets.

Type	Example (pragmatic unit: // //, syntactic unit: < >, intonational unit: { })			Count
Congruent (one-one-one)	1	Alice:	And with all the breaks that they've gotten,	1,100
	2		they're never gonna have hard times.	
	3 →	Mary:	//<{Hard times do train you}>//.	
	4	Alice:	Yep.	
	5	Mary:	They do.	
Pragmatics-syntax-bound (one-one-two)	1	Alice:	Yeah.	284
	2 →		//<{Ron was singlehandedly there}.	
	3 →	Mary:	{with one wrecker}>//.	
	4	Alice:	Yeah=	
Pragmatics-prosody-bound (one-two-one)	1	Michael:	We have very little control over it.	351
	2		But once we do=	
	3		We'll be able to program biology as well.	
	4 →	Jim:	//<{Well> <that's pretty frightening concept}>//.	
	5	Michael:	It is frightening but,	
	6	Jim:	<i>We can't even control our freeways</i>	
Pragmatics-bound (one-two-two)	1 →	Michael:	//<{And}>,	547
	2 →		<{It might have started a chain reaction that just blew up [the whole earth]}>//.	
	3	Jim:	[That's right].	
Prosody-bound (two-two-one)	1	Mary:	What is it.	37
	2	Alice:	Norplant?	
	3 →	Mary:	//<{Oh}>// //<really}>//.	

Table 5. Types and distribution of conversational discourse units.

3.3 Turn-taking annotation: floor-taking-turn vs. non-floor-taking-turn

Speakership alternation in conversation can be roughly divided into two types. The first one is the so-called *floor-taking turn*, which this study is restricted to. It is fairly easy to operationalise as a turn that is initiated by another speaker. The second one is the *non-floor-taking-turn* (Ford & Thompson, 1996), manifested mainly in the form of backchannels (Schegloff, 1982), either with one-word particles such as *yeah*, *wow*, *uh-hum* or idiom-like usage such as *I see*, *that's right* and *for sure*. They are produced to either demonstrate the listener's acknowledgement or to signal that the listener is paying attention to the speaker, but does not wish to take the floor. We acknowledge that backchannels play a critical role in managing the conversation (Jefferson, 1978). However, since interlocutors in those cases do not seem to genuinely take a turn (cf. Ford & Thompson, 1996), it is assumed that the mechanism of producing them is different from the production of the floor-taking turn (Levinson & Torreira, 2015), and might even go through a radically different mechanism in language production (Macwhinney, 2008; Snider & Arron, 2012). That is why we excluded them from the present study. Next to backchannels, we also do not consider joint-turn constructions as a floor-taking-turn. This decision is based on the observation that all 10 cases of joint-turn construction in the dataset resulted from the listener's successful prediction of what the speaker is going to say. In such cases, where the listener joins the speaker in uttering the upcoming information, we do not consider them to purposefully take the turn, rather the joint-turn construction is seen as an expression of acknowledgment and understanding. One such case appears below in Extract 6.

Extract 6 (Du Bois et al., 2000–2005)

1 Richard: I don't know if the parents a=re awa=re,
 2 that we did,
 3 you know,
 4 Fred: [Break up]?
 5 Richard: [separate].

3.4 Reliability check

The three-dimensional segmentation was manual, following the criteria presented in Section 3.2. In order to ensure the soundness of this manual process, several reliability checks were performed. First, in line with Spooren and Degand's (2010) recommendation that annotation and coding requires training, the segmentation process was conducted twice on part of the dataset, namely the entire task-oriented dialogue, with a six-week interval between the annotations. Intra-coder reliability was performed by the first author. It was calculated on two episodes randomly selected from the dialogue at the interval from the sixth to the twelfth minute and from the twenty-fourth to the thirtieth minute. This examination yielded Cohen's kappa reliability values of 0.91 for the syntactic segmentation, 0.70 for the prosodic segmentation and 0.75 for the pragmatic units, indicating high reliability of our operationalisation.

We then turned to the segmentation of the five natural SBC conversations following the same operationalisation. Here too, we operated a reliability check, this time between two coders (inter-coder reliability rather than intra-coder reliability). One of the coders was the first author, who coded the full dataset. The other coder, a student in linguistics who was blinded to the purpose of this study, was trained to identify the syntactic units, pragmatic units, and turn-taking. Following the training, an inter-coder reliability check was performed for 25% (28 minutes) of the conversations, yielding Cohen kappa's reliability scores of 0.84 for turn-taking identification, 0.91 for the identification of syntactic units and 0.94 for pragmatic units, indicating almost perfect agreement for turn-taking annotation and the identification of syntactic and pragmatic units (Landis & Koch, 1977). Note that the intonation units were already present in the original data set and were not reannotated (see Section 3.1). All cases of disagreement regarding the syntactic and pragmatic segmentation and the identification of turn-taking were resolved individually through discussion. These concerned 52 out of 1344 syntactic segmentation cases, where a case refers to one boundary, 42 pragmatic cases out of 1125 boundaries, and 22 turn-takings out of 274 cases, where a case stands for the place where the "other speaker" produces verbal content.

4 Results

This section first shows the association between turn-taking and the three types of linguistic boundaries. It then presents the influence that different types of CDU have on the transition speed. Note that the analyses have been conducted based on the five SBC conversations.

4.1 Syntactic, prosodic and pragmatic boundaries in turn-taking

A total number of 5006 boundaries and 507 turn-taking cases were identified in the dataset. Still, 21 turn-taking cases were excluded from further analysis either because unclear speech made identifying the boundary in each dimension too challenging, or because of difficulties in identifying the utterance to which the speaker was replying. Among the 486 remaining valid turn-taking cases, we observed five that occurred when none of the three boundaries were present. We kept them in the analyses.

Due to issues with multicollinearity and a low turn-taking incidence rate, we could not perform logistic regression. Three separate chi-square tests of independence were thus performed to examine the relation between the status (i.e. present vs. absent) of the three types of linguistic boundary and the occurrence of turn-taking (i.e. occurred vs. did not occur). There was a statistically significant relationship between the status of the syntactic boundary and occurrence of

turn-taking, $\chi^2(1, N = 5011) = 93.79, p < .001$. That is, turn-taking was more likely to be observed immediately after a syntactic boundary compared to locations where a syntactic boundary is absent. Identical results were observed for the prosodic boundary, $\chi^2(1, N = 5011) = 131.30, p < .001$ and pragmatic boundary, $\chi^2(1, N = 5011) = 423.51, p < .001$. In addition, as Table 6 shows, the conditional proportions of turn-taking are 11.6%, 12.6% and 18.8% when syntactic, prosodic and pragmatic boundaries are present, respectively. Each of these is significantly higher than the corresponding conditional proportions (i.e. syntactic: 1.2%, prosodic: 1.7% and pragmatic: 1.6%) for absence of the boundary. According to the tests and description of the proportions, we found that the pragmatic boundary is most likely among the three types to be associated with turn-taking. The syntactic boundary is least likely to be associated with turn-taking, and the prosodic boundary is in between.⁴

			Turn-taking (occurred)	χ^2	df	<i>p</i>
Syntactic boundary	Absent	Count	11	93.79	1	<.001
		% within syntactic boundary	1.2%			
	Present	Count	475			
		% within syntactic boundary	11.6%			
Prosodic boundary	Absent	Count	22	131.30	1	<.001
		% within prosodic boundary	1.7%			
	Present	Count	464			
		% within prosodic boundary	12.6%			
Pragmatic boundary	Absent	Count	41	423.51	1	<.001
		% within pragmatic boundary	1.6%			
	Present	Count	445			
		% within pragmatic boundary	18.8%			

Table 6. Chi-square tests: Association between linguistic boundaries and turn-taking.

4.2 Transition speed after each type of conversational discourse unit

Transition speed is defined as the temporal interval between the offset of the speaker's current turn and the onset of the listener's response (see 'Offset 2' as defined by Kendrick & Torreira, 2015: 266). Analyses included the CDUs after which turn-taking took place, yet a number of tokens were removed from the analyses for different reasons. Among the 482 valid turn-taking cases, 24 cases of turn-taking that did not take place at the boundary of a CDU (with three coinciding boundaries) were ruled out. Another 49 cases were excluded, because they were followed by a pause longer than 1,500 ms (cf. Kendrick, Holler & Levinson, 2023; Templeton et al., 2023). The reason is that such longer transition time probably involved other factors than mere prediction, such as reaction behaviour (Duncan, 1972) and/or preference issues (Kendrick & Torreira, 2015). Moreover, a further 46 cases, were excluded from analysis because the turns' exact boundaries were difficult to identify due to low recording quality and chaotic overlap. Finally, we also excluded the 8 cases of the prosody-bound type, because they were too rare. We thus ended up with 355 valid turn-taking cases involving the remaining four CDU types.

To assess the impact of the type of conversational discourse units on transition speed following the boundary of each CDU type, we conducted a repeated-measures analysis of variance

⁴ It should be acknowledged that the number of conversations is small in the current analysis. It may spark concern on the violation of the assumption of independence. Therefore, we abstained from making an overly assertive and indisputable conclusion. Further examinations with larger datasets are warranted.

(ANOVA). Table 7 presents the means and standard deviations for the gap after each type of CDU vis-à-vis each participant (Speaker).

Mauchly's test confirmed that the assumption of sphericity had been met ($X^2(5) = 8.101, p = .155$). The effect of the CDU type was statistically significant at the 5% level of significance, $F(3, 24) = 4.33, p = .014$, with a partial η^2 of .351. However, Speaker 4 was excluded from our analysis because the transition speeds after the pragmatics-syntax-bound CDU and the pragmatics-bound CDU were more than two times the standard deviation (2 SD) of each corresponding mean.

Post-hoc pairwise comparisons adjusted using the Bonferroni method revealed no significant difference between the mean transition speeds after the boundary of the pragmatics-syntax-bound CDU and those after the boundary of the pragmatics-bound CDU ($p = .991, 95\% \text{ C.I.} = [-358.03, 139.56]$). The average transition speed following the boundary of the pragmatics-syntax-bound CDU and pragmatics-bound CDU indicated a marginally significant difference ($p = .052, 95\% \text{ C.I.} = [-392.36, 1.53]$). Notably, the mean transition speed after the boundary of the pragmatics-syntax-bound CDU was significantly higher than that after the boundary of the pragmatics-prosody-bound CDU ($p = .018, 95\% \text{ C.I.} = [-401.51, -38.40]$). Finally, there were no statistically significant differences among the congruent, pragmatics-prosody-bound CDU ($p = 1.00, 95\% \text{ C.I.} = [-279.24, 230.16]$) and pragmatics-bound CDU ($p = .52, 95\% \text{ C.I.} = [-67.31, 239.67]$), as illustrated in Table 8.

Speaker	Congruent		Pragmatics-syntax bound		Pragmatics-prosody bound		Pragmatics-bound	
	M	SD	M	SD	M	SD	M	SD
1	399.64	293.03	10.22	428.70	221.67	300.13	219.50	410.89
2	341.75	553.44	476.00	470.88	764.83	485.31	228.08	418.17
3	662.05	424.94	405.50	54.48	689.75	407.11	438.67	260.71
4	531.68	316.53	923.00	.	645.80	507.53	1343.50	70.00
5	623.04	374.17	238.00	.	779.20	456.28	492.00	183.20
6	713.41	409.66	377.75	303.88	404.37	372.53	722.00	63.64
7	323.80	335.89	210.00	168.77	376.75	160.57	301.21	189.51
8	358.94	429.18	287.88	254.11	457.25	464.50	77.00	350.96
9	202.29	245.17	61.00	16.97	329.25	81.03	274.75	552.50
10	374.63	416.60	174.46	227.07	197.33	111.42	470.71	330.54
Total	471.93	406.05	233.04	324.55	495.57	402.76	470.71	330.54
2 SD of the M	[-340.17, 1284.03]		[-419.06, 882.14]		[-309.95, 1301.09]		[-453.26, 1146.38]	

Table 7. Speakers' transition speeds following each type of CDU.

Dependent variable	F(3, 24)	p	η^2	Congruent		Pragmatics-syntax bound		Pragmatics-prosody bound		Pragmatics-bound	
				M	SD	M	SD	M	SD	M	SD
Gap	4.33	.014	.351	444.39 _a	176.61	248.98 _a	155.39	468.93 _b	223.55	358.21 _{a,b}	191.98

Table 8. One-way repeated-measures ANOVA: The effect of the type of CDU on the transition speed after the boundary of each CDU type. Note. Means not sharing subscripts are significantly different from one another.

5 General discussion

A close examination of the role that CDUs play in the turn-taking operation found a significant association between the presence of three linguistic boundaries and occurrence of turn-taking. Additionally, the analysis based on the current dataset suggested the duration of the transition gap between turns was influenced by the CDU type. In this section, we discuss the possible reasons underpinning these observations.

5.1 The importance of meaning-connection in managing turn-taking

The comparison of the conditional proportions of turn-taking in presence or absence of syntactic, prosodic, and pragmatic boundaries led to the conclusion that the pragmatic boundary is the best predictor for turn-taking, whereas the syntactic boundary was the worst. The prosodic boundary was in between. This observation aligns with Schegloff (2007): turn-at-talk is organised not just by grammar and phonetic properties, but also by social action. Conversation emerges from action formation and ascription (Levinson, 2013). This perspective has received support from several studies in cognitive neuroscience and psycholinguistics, which have indicated that interlocutors allocate resources to process the actions of their conversational partners. For instance, neuroscientific studies have indicated that addressees attempt to predictively identify a speech act before even hearing it (Gisladottir, Chwilla & Levinson, 2015; Gisladottir, Bögels & Levinson, 2018). In addition to predicting speech acts, listeners show differences in how direct and indirect speech acts are processed (Clark & Lucy, 1975; Coulson & Lovett, 2010).

Considering this together with the pragmatic unit, defined as one step of implementing an action plan and identified by analysing the meaning the speaker is attempting to convey within the situated context, there is a straightforward connection to the ‘meaning’ the speaker is attempting to convey. In other words, rather than following syntactic and prosodic structures *per se*, interlocutors predominantly follow the meaning their dialogic partner conveys. This idea is consistent with recent studies on joint-turn construction, which is traditionally defined as one participant completing a syntactic unit initiated by another participant (Lerner & Takagi, 1999; Suzuki & Usami, 2005). Laury and Ono (2021) found that if the two parts of the joint-turn construction are combined, they often do not form a syntactically acceptable construction. Rather, the two parts are semantically conjoined—the second speaker produces an utterance based on their understanding of prior utterances. Therefore, interlocutors may not jointly construct syntactic units but rather connect parts of joint-turn construction via meaning.

When discussing meaning connection, it is important to consider views on the connection between intonation and meaning expression. Chafe (1987, 1994) proposed that the intonation unit corresponds to a new complete idea. Following this view, the association between prosodic boundaries and turn-taking should have been similar to that between the pragmatic boundary and turn-taking. Yet, Chafe’s proposal is based on an ideal situation where the pause between two intonation units denotes the change in information activation in the speaker’s mind. In other words, activated information becomes deactivated during the pause, whereas inactive information becomes activated (Chafe, 1987; 1994). However, according to our observation, a considerable number of pauses resulted from difficulties in verbalisation as well as breath and laughter. That is, the segment before the pause/boundary failed to convey a complete idea, and such pauses might thus fail to lead to a turn-transition. While prosodic boundaries may be the most frequent type of boundary, this overall frequency decreases its degree of specific association with turn transitions.

Nevertheless, compared to the syntactic boundary, prosodic boundaries showed slightly stronger associations with turn-transition occurrence. This aligns with studies wherein prosody is considered critical to the timing of turn-taking (Bögels & Torreira, 2015, 2021) but contradicts others wherein syntax is argued to play a dominant role in turn-taking (De Ruiter, Mitterer & Enfield, 2006; Sacks, Schegloff & Jefferson, 1974). Importantly, the difference between these two factors was not substantial; thus, it is worth reiterating that these two factors may produce only

trivial independent effects (Auer, 1996; Selting, 2000). This largely resonates with Levinson and Torreira's (2015) proposal for a psycholinguistic model of turn-taking, where they argued that:

morphosyntax may provide most of the early clues to the overall structural envelope . . . , so offering some long distance projection. Within the last half second or so, the actual words will often be predicted, and, within that same late time frame, cues to imminent turn closure, usually prosodic and phonetic, are likely to appear, indicating a likely turn end. (p. 13)

It should also be noted that the weight of syntax and prosody in turn-taking may differ across languages as speakers of different languages may rely to different degrees on syntax and prosody in language processing. Evidence comes from unrelated but noteworthy studies on syllable effects (Mehler et al., 1981), where a target syllable is detected faster when it precisely matches the first syllable of the carrier (e.g. *ba* in *ba.lance*) as compared to when it does not (e.g. *ba* in *balcon*). Studies have found that this effect largely depends on whether the particular language is a stress language (Cutler, 1997).

5.2 Relation between type of CDU and transition speed: earlier emergence of the gist

We should acknowledge that turn-taking is not simply about taking the turn at the proper place. The average transition speeds after the four types of CDUs in our dataset are all shorter than 500 ms in average, faster than the time required to produce a single word (Indefrey & Levelt, 2004). In other words, interlocutors, in general, prepare their response before the arrival of the turn-end. Meanwhile, in our analysis, even if all turn-taking occurred at the place where syntactic, prosodic and pragmatic boundaries converge, the transition speeds to certain extent differ across the CDU types. It is particularly noteworthy that transition speed after the pragmatics-syntax-bound CDU, i.e. [1 pragm./1 synt./2 inton.], is the fastest (on average), which differs from that of the pragmatics-prosody-bound [1 pragm./2 synt./ 1. inton.] CDU. Uncovering the precise reasons for these differences is beyond the scope of this study. However, with the CDU segmentation in hand we would like to suggest a number of possible directions for future research.

Despite the discrepancy of when people begin to prepare their utterance (Bögels, Magyari & Levinson, 2015; Boiteau et al., 2014; Piai et al., 2015; Sjerps & Meyer, 2015), mounting evidence has indicated that ideally the semantic direction of an utterance is clarified as early as possible (Corps et al., 2018) so that speakers can anticipate the upcoming content, thus facilitating turn-taking. In other words, the earlier the emergence of the gist⁵ of an utterance, the earlier the listener starts to prepare a response and the faster the transition speed. In this regard, it is reasonable to believe that the pragmatics-syntax-bound type possesses specific linguistic features that other types do not have to make its gist become clear earlier than others.

It is notable that English fronts some key (lexico)syntactic characteristics (Levinson, 2013) that allow speakers to reveal their gist earlier than those languages users whose language provides the key (lexico)syntactic characteristics relatively late. Levinson (2013) highlighted that English tends to shift its question words (e.g. *wh*-words) right up to the front of an utterance, so that listeners can recognise the information-seeking actions earlier (Levinson, 2013). For instance, when trying to find a key, an English speaker may say 'Where is my key?'; meanwhile, the same meaning should be formulated as 'My key where' in Mandarin Chinese. Therefore, the English listener may recognise the speaker's information-seeking actions earlier than the Chinese listener under the same situation.⁶ Similar observations were made by Ono and Thompson (2017) regarding the earlier occurrence of the negative morpheme in natural English conversation than in Japanese.

⁵ We use 'gist' to refer to either the communicative action, such as interrogative or statement, the speaker is going to conduct or the concrete content their utterance is designed to convey.

⁶ Different languages could have different ways of front-loading the same gist (e.g. information-seeking). It is possible for Chinese users to have recognised the information-seeking action as early as English users did by drawing on other linguistic resources in addition to syntax, such as prosody (cf. Shen, 1990).

Consequently, they concluded that English speakers do not need to wait until the end of an utterance to recognise the negative semantic implication of the entire utterance.

Inspired by these studies, it is reasonable to speculate that the end point of the main syntactic part of the pragmatic-syntax-bound CDU occurs earlier than that in the other three types of CDUs. To this end, we considered the duration between the end of the head of the CDU (usually the predicate) and the end of the CDU.⁷ Figure 5 displays that the post-predicate duration of the pragmatic-syntax-bound CDU is longer than that of the other three types. The raw descriptive data refrain us from making any cogent conclusion, but to a certain extent it indicates that the syntactic structure of the pragmatic-syntax-bound CDU may help the listener to grasp its gist earlier than in the other types. Being aware that our measurement is oversimplified, further scrutiny is needed regarding the influence of the length (Magyari & de Ruyter, 2012), prosody (Garrod & Pickering, 2015) and transitivity (Tao & Hu, 2019; Thompson & Hopper, 2001) of the CDU.

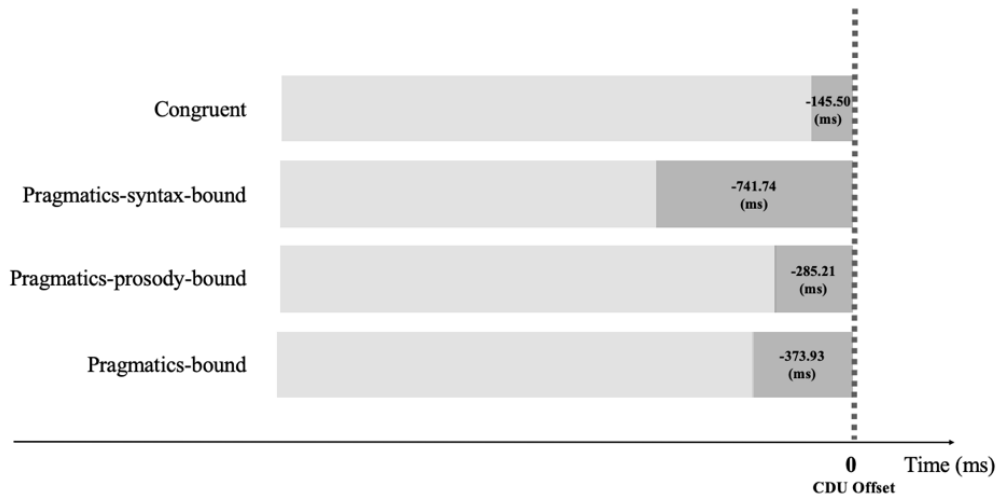


Figure 5. Post-predicate duration of CDU (Dark Gray Part).

6 Conclusion

The aim of this study was to extend our understanding of the building blocks of conversational discourse and the role they play in turn transition. We propose that the discourse unit should be an amalgam of three types of units: syntactic (dependency) clauses, intonation units and pragmatic actions. The mapping of these three types of segments form Conversational Discourse Units that are in line with human communication in which interlocutors exploit linguistic information to convey and interpret one another’s intentions. Such mapping provides a linguistic perspective for defining the analytical conversational unit, which is considered basic for organising conversational turn-taking. Based on the investigation of natural conversation, this study demonstrated that compared to prosodic and syntactic boundaries, interlocutors are subject to take turns immediately after a full pragmatic unit. This result indicates that meaning-connection and action ascription underpin the management of turn-taking. Meanwhile, transition speed after the pragmatic-syntax-bound CDU denotes a statistical difference in one of the three remaining types of CDU. This finding may be attributed to the linguistic structure of the pragmatic-syntax-bound CDU, which enables interlocutors to understand its gist earlier than they do for other types of CDU.

The investigation of the management of turn-taking began with identifying CDUs. The question may arise why we would need a new concept, when TCU-based analysis is a well-

⁷ The post-predicate duration of the congruent CDU was calculated based on 80 cases randomly selected from the 130 congruent CDU cases.

established method for investigating turn-taking. Clearly, this study does not intend to disregard or dismiss the TCU framework. Rather, it seeks to provide an additional perspective to understand turn-taking. More importantly, as a linguistic unit, CDU enriches our toolkit not only for exploring the turn-taking operation but also for extending our understanding of language production and comprehension in several other ways, such as exploring the possibility of speaker-hearer alignment at the discourse level, uncovering cross-linguistic differences in conversational structure or evaluating the impact of the communicative context on the flow of discourse.

Our segmentation extends the work of Degand and Simon (2009), whose study was restricted to the interplay between the prosodic and syntactic dimensions in defining units of discourse segmentation. To the best of our knowledge, this study is the first to incorporate the pragmatic dimension into discourse segmentation in a (at least fairly) objective way. It overcomes the long-time criticism on the subjectivity and ambiguity of the definition of pragmatic completeness (Skantze, 2021). Yet, the analysis of the situated context underlying the identification of the pragmatic units being based mainly on semantic understanding of the utterance, further work is needed on identifying linguistic (surface) rules related to pragmatic boundaries, so that our CDU proposal could be applied to a larger data analysis, thus gaining generalisation.

Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N°859588. We are grateful for detailed and constructive comments from three anonymous reviewers. All remaining errors are ours.

References

- Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cécile Fabre, Mai Ho-dac, Anne Le Draoulec, Philippe Muller, Marie-Paule Péry-Woodley, Laurent Prévot, Josette Rebeyrolles, Ludovic Tanguy, Marianne Vergez-Couret, Laure Vieu (2012). An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus. In *The Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’ 12)*, Istanbul, Turkey. http://www.lrec-conf.org/proceedings/lrec2012/pdf/836_Paper.pdf
- Nicholas Asher and Alex Lascarides (2003). *Logics of Conversation*. Cambridge University Press, Cambridge, East of England.
- Gerry T. M. Altmann and Jelena Mirković (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33:583–609. <https://doi.org/10.1111%2Fj.1551-6709.2009.01022.x>
- Peter Auer (1996). On the prosody and syntax of turn-continuations. In Elizabeth Couper-Kuhlen and Margret Selting, editors, *Prosody in Conversation: Interactional Studies*, pages 57–100. Cambridge University Press, New York City, New York.
- Peter Auer (2005). Projection in interaction and projection in Grammar. *Text*, 25(1):7–36. <https://doi.org/10.1515/text.2005.25.1.7>
- Brendan Barnwell (2013). Perception of prosodic boundaries by untrained listeners. In Beatrice Szczepek Reed and Geoffrey Raymond, editors, *Units of Talk – Units of Action*, pages 125–165. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania.
- Dagmar Barth-Weingarten (2013). From “intonation units” to censoring – an alternative approach to the prosodic-phonetic structuring of talk-in-interaction. In Beatrice Szczepek Reed and Geoffrey Raymond, editors, *Units of Talk – Units of Action*, pages 91–124. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania.

- Douglas Biber, Jesse Egbert, Daniel Keller, Stacey Wizner (2021). Towards a taxonomy of conversational discourse types: An empirical corpus-based analysis. *Journal of Pragmatics*, 171:20–35. <https://doi.org/10.1016/j.pragma.2020.09.018>
- Sara Bögels (2020). Neural correlates of turn-taking in the wild: Response planning starts early in the free interviews. *Cognition*, 203:104347. <https://doi.org/10.1016/j.cognition.2020.104347>
- Sara Bögels, Lilla Magyari and Stephen C. Levinson (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, 5:12881. <https://doi.org/10.1038/srep12881>
- Sara Bögels and Francisco Torreira (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, 52:46–57. <https://doi.org/10.1016/j.wocn.2015.04.004>
- Sara Bögels and Francisco Torreira (2021). Turn-end estimation in conversational turn-taking: The roles of context and prosody. *Discourse Processes*, 58(10):903–924. <https://doi.org/10.1080/0163853X.2021.1986664>
- Timothy W. Boiteau, Patrick S. Malone, Sara A. Peters and Amit Almor (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General*, 143(1):295–311. <https://doi.org/10.1037/a0031858>
- Galina B. Bolden (2009). Implementing incipient actions: The discourse marker so in English conversation. *Journal of Pragmatics*, 41(5):974–998. <https://doi.org/10.1016/j.pragma.2008.10.004>
- Brentari Diane, Joshua Falk, Anastasia Giannakidou, Annika herrmann, Elisabeth Volk and Markus Steinbach (2018). Production and comprehension of prosodic markers in Sign Language imperatives. *Frontiers in Psychology*, 9:770. <https://doi.org/10.3389/fpsyg.2018.00770>
- Andrew Carnie (2021). *Syntax: A Generative Introduction (Fourth Edition)*. Wiley-Blackwell, Chichester, West Sussex.
- Wallace Chafe (1987). Cognitive constraints on information flow. In Russell S. Tomlin, editor, *Coherence and Grounding in Discourse*, pages, 21–51. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania.
- Wallace Chafe (1993). Prosodic and functional units of language. In Jane A. Edwards and Martin D. Lampert, editors, *Talking Data: Transcription and Coding in Discourse Research*, pages 33–43. Lawrence Erlbaum, Hillsdale, New Jersey.
- Wallace Chafe (1994). *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press, Chicago, Illinois.
- Herbert H. Clark (1996). *Using Language*. Cambridge University Press, Cambridge, East of England.
- Herbert H. Clark and Peter Lucy (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, 14:56–72. [https://doi.org/10.1016/S0022-5371\(75\)80006-5](https://doi.org/10.1016/S0022-5371(75)80006-5)
- Ruth E. Corps, Abigail Crossley, Chiara Gambi and Martin J. Pickering (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. <https://doi.org/10.1016/j.cognition.2018.01.015>
- Seana Coulson and Christopher Lovett (2010). Comprehension of non-conventional indirect requests: An event-related brain potential study. *Italian Journal of Linguistics*, 22(1):107–124. https://www.italian-journal-linguistics.com/app/uploads/2021/05/6_coulsonlovett.pdf
- Elizabeth Couper-Kuhlen (2006). Prosodic cues of discourse units. In Keith Brown, editor, *Encyclopedia of Language & Linguistics (Second Edition)*, pages 178–182. Elsevier, Amsterdam, Noord-Holland. <https://doi.org/10.1016/B0-08-044854-2/00588-5>.
- Anne Cutler (1997). The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes*, 12:839–845. <https://doi.org/10.1080/016909697386718>

- Liesbeth Degand (2019). Review of the book *Spontaneous spoken English: An integrated approach to the emergent grammar of speech* by A. Haselow. *Language*, 95(1):185–187. <https://doi.org/10.1353/lan.2019.0018>
- Liesbeth Degand and Ludivine Crible (2021). Discourse markers at the peripheries of syntax, intonation and turns. Towards a cognitive-functional unit of segmentation. In Daniël Van Olmen and Jolanta Šinkūnienė, editors, *Pragmatic Markers and Peripheries*, pages 19–48. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania. <https://doi.org/10.1075/pbns.325.01deg>
- Liesbeth Degand and Anne-Catherine Simon (2009). On identifying basic discourse units in speech: theoretical and empirical issues. *Discours: Revue de Linguistique, Psycholinguistique et Informatique*. <https://doi.org/10.4000/discours.5852>
- Liesbeth Degand, Anne-Catherine Simon, Noalig Tanguy and Thomas van Damme (2014). Initiating a discourse unit in spoken French: Prosodic and syntactic features of the left periphery. In Salvador Pons Bordería, editor, *Discourse Segmentation in Romance Languages*, pages 243–273. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania. <https://doi.org/10.1075/pbns.250.09deg>
- Marie-Catherine de Marneffe and Joakim Nivre (2019). Dependency grammar. *Annual Review of Linguistics*, 5:197–218. <https://doi.org/10.1146/annurev-linguistics-011718-011842>
- Jan Peter de Ruiter, Holger Mitterer and Nick J. Enfield (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82:515–535. <https://doi.org/10.1353/lan.2006.0130>
- Paul Drew (2013). Turn design. In Tanya Stivers and Jack Sidnell, editors, *The Handbook of Conversation Analysis*, pages 131–149. Wiley-Blackwell, Malden, Massachusetts.
- John W. Du Bois, Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson and Nii Martey (2000–2005). *Santa Barbara Corpus of Spoken American English, Parts 1–4*. Linguistic Data Consortium. <https://www.linguistics.ucsb.edu/research/santa-barbara-corpus#access>
- John W. Du Bois, Stephan Schuetze-Coburn, Susanna Cumming and Danae Paolino (1993). Outline of discourse transcription. In Jane A. Edwards and Martin D. Lampert, editors, *Talking Data: Transcription and Coding Methods for Language Research*, pages 45–89. Lawrence Erlbaum, Mahwah, New Jersey.
- Starkey Duncan (1972). Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology*, 23(2):283–292. <https://psycnet.apa.org/doi/10.1037/h0033031>
- Lotte Eijk, Marlou Rasenberg, Flavia Arnese, Mark Blokpoel, Mark Dingemanse, Christian F. Doeller, Mirjam Ernestus, Judith Holler, Branka Milivojevic, Asli Özyürek, Wim Pouw, Iris van Rooij, Herbert Schriefers, Ivan Toni, James Trujillo and Sara Bögers (2022). The CABB dataset: A multimodal corpus of communicative interactions for behavioural and neural analyses. *Neuroimage*, 264:119734. <https://doi.org/10.1016/j.neuroimage.2022.119734>
- Jesse Egbert, Stacey Wizner, Daniel Keller, Douglas Biber, Tony McEnery and Paul Baker (2021). Identifying and describing functional discourse units in the BNC Spoken 2014. *Text & Talk*, 41(5-6):715–737. <https://doi.org/10.1515/text-2020-0053>
- Nick J. Enfield and Jack Sidnell (2017). On the concept of action in the study of interaction. *Discourse Studies*, 19(5):515–535. <https://doi.org/10.1177/1461445617730235>
- Michael C. Ewing (2019). The predicate as a locus of grammar and interaction in colloquial Indonesian. *Studies in Language*, 43(2):402–443. <https://doi.org.ezproxy.is.ed.ac.uk/10.1075/sl.17062.ewi>
- Laurent Filliettaz and Eddy Roulet (2002). The Geneva Model of discourse analysis: An interactionist and modular approach to discourse organization. *Discourse Studies*, 4(3), 369–393. <https://doi.org/10.1177/14614456020040030601>

- Janet Dean Fodor (2002). Prosodic disambiguation in silent reading. *North East Linguistics Society*, 32, 113–132. https://scholarworks.umass.edu/nels/vol32/iss1/8?utm_source=scholarworks.umass.edu%2Fnels%2Fvol32%2Fiss1%2F8&utm_medium=PDF&utm_campaign=PDFCoverPages
- Cecilia E. Ford (2004). Contingency and units in interaction. *Discourse Studies*, 6(1):27–52. <https://journals.sagepub.com/doi/pdf/10.1177/1461445604039438>
- Cecilia E. Ford, Barbara A. Fox and Sandra A. Thompson (1996). Practices in the construction of turns: the TCU revisited. *Pragmatics*, 6(3):427–454. <https://doi.org/10.1075/prag.6.3.07for>
- Cecilia E. Ford and Sandra A. Thompson (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the projection of turn completion. In: Elinor Ochs, Emanuel A. Schegloff and Sandra A. Thompson, editors, *Interaction and Grammar*, pages 135–184. Cambridge University Press, Cambridge, East of England.
- Bruce Fraser (2009). An account of discourse markers. *International Review of pragmatics*, 1:293–320.
- Simon Garrod and Martin J. Pickering (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, 6(751):1–12. <https://doi.org/10.3389%2Ffpsyg.2015.00751>
- Rosa S. Gisladdottir, Sara Bögels and Stephen C. Levinson (2018). Oscillatory brain responses reflect anticipation during comprehension of speech acts in spoken dialogue. *Frontiers in Human Neuroscience*, 12:34. <https://doi.org/10.3389/fnhum.2018.00034>
- Rosa S. Gisladdottir, Dorothee J. Chwilla and Stephen C. Levinson (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS One*, 10(3):e0120068. <https://doi.org/10.1371/journal.pone.0120068>
- Paul Grice (1975). Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Studies in Syntax and Semantics III: Speech Acts*, pages 183–198. Academic Press, Cambridge, Massachusetts.
- Paul Grice (1989). *Studies in the Ways of Words*. Harvard University Press, Cambridge, Massachusetts.
- Alexander Haselow (2017). *Spontaneous Spoken English: An Integrated Approach to the Emergent Grammar of Speech*. Cambridge University Press, Cambridge, East of England.
- Barbara J. Grosz and Candace L. Sidner (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3): 175–204. <https://dash.harvard.edu/handle/1/2579648>
- Mattias Heldner, Jens Edlund, Anna Hjalmarsson and Kornel Laskowski (2011). Very short utterances and timing in turn-taking. In *Interspeech 2011*, pages 2837–2840, Florence, Tuscany.
- Hans Jürgen Heringer (1993). Dependency syntax - basic ideas and the classical model. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann de Gruyter, editors, *Syntax – An International Handbook of Contemporary Research (Vol. 1)*, pages 298–316. Walter de Gruyter, Berlin.
- John Heritage (1984). *Garfinkel and Ethnomethodology*. Polity, Cambridge.
- John Heritage (2013). Turn-initial position and some of its occupants. *Journal of Pragmatics*, 57:331–337. <https://doi.org/10.1016/j.pragma.2013.08.025>
- John Heritage and Marja-Leena Sorjonen (1994). Constituting and maintaining activities across sequences: And-prefacing as a feature of question design. *Language in Society*, 23:1–29. <https://doi.org/10.1017/S0047404500017656>
- John Heritage and Marja-Leena Sorjonen (2018). *Between Turns and Sequence: Turn-initial Particles Across Languages*. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania.

- Judith Holler, Kobin H. Kendrick and Stephen C. Levinson (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, 25(5):1900–1908. <https://doi.org/10.3758/s13423-017-1363-z>
- Judith Holler and Stephen C. Levinson (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, 23(8):639–652. <https://doi.org/10.1016/j.tics.2019.05.006>
- Paul J. Hopper and Sandra A. Thompson (2008). Projectability and clause combining in interaction. In Ritva Laury, editors, *Crosslinguistic Studies of Clause Combining: The Multifunctionality of Conjunctions*, pages 99–124. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania.
- Hanneke Houtkoop and Harrie Mazeland (1985). Turns and discourse units in everyday conversation. *Journal of Pragmatics*, 9:595–619. [https://doi.org/10.1016/0378-2166\(85\)90055-4](https://doi.org/10.1016/0378-2166(85)90055-4)
- Junfei Hu and Liesbeth Degand (2022). Do speakers align at the discourse level? The role of discourse segments in task-oriented dialogue. In *The CogLing Day 2022*, Tilburg, North Brabant.
- Maya Inbar, Eitan Grossman and Ayelet N. Landau (2020). Sequences of intonation units from ~ 1 Hz rhythm. *Nature*, 10(1):1–9. <https://doi.org/10.1038/s41598-020-72739-4>
- Peter Indefrey and Willem J. M. Levelt (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2):101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Shlomo Izre’el (2020). The basic unit of spoken language the interfaces between prosody, discourse and syntax: A view from spontaneous spoken Hebrew. In Shlomo Izre’el, Heliana Mello, Alessandro Painunzi and Tommaso Raso, editors, *In Speech of Basic Units of Spoken Language: A Corpus-driven Approach*, pages 77–106. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania.
- Shlomo Izre’el, Heliana Mello, Alessandro Panunzi and Tommaso Raso (2020). *In Search of Basic Units of Spoken Language: A Corpus-Driven Approach*. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania.
- Gail Jefferson (1978). Sequential aspects of storytelling in conversation. In Jim Schenkein, editor, *Studies in the Organization of Conversational Interaction*, pages 219–248. Elsevier Inc, Amsterdam, Noord-Holland. <https://doi.org/10.1016/B978-0-12-623550-0.50016-1>
- Gail Jefferson (1983). On a failed hypothesis: “conjunctionals” as overlap-vulnerable. *Tilburg Papers in Language and Literature*, 28:1–33.
- Kobin H. Kendrick, Judith Holler and Stephen C. Levinson (2023). Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society B*, 378(1875):20210473. <https://doi.org/10.1098/rstb.2021.0473>
- Kobin H. Kendrick and Francisco Torreira (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, 52(4):255–289. <https://doi.org/10.1080/0163853X.2014.955997>
- Jieun Kiaer (2014). *Pragmatic syntax*. Bloomsbury, London, England.
- Hye Ri Stephanie Kim and Satomi Kuroshima (2013). Turn beginnings in interaction: An introduction. *Journal of Pragmatics*, 57:267–273. <https://doi.org/10.1016/j.pragma.2013.08.026>
- Sotaro Kita and Asli Özyürek (2003). What does crosslinguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48: 16–32. [https://doi.org/10.1016/S0749-596X\(02\)00505-3](https://doi.org/10.1016/S0749-596X(02)00505-3)
- J. Richard Landis and Gary G. Koch (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174. <https://doi.org/10.2307/2529310>

- Ritva Laury and Tsuyoshi Ono (2021). Co-construction reconsidered: Over-syntacticization in interactional linguistics. In *The Fourth Interactional Conference on Interactional Linguistics and Chinese Language Studies*, Beijing.
- Gene H. Lerner and Tomoyo Takagi (1999). On the place of linguistic resources in the organization of talk-in-interaction: A co-investigation of English and Japanese. *Journal of Pragmatics*, 31(1):49–75. [https://doi.org/10.1016/S0378-2166\(98\)00051-4](https://doi.org/10.1016/S0378-2166(98)00051-4)
- Willem J. M. Levelt (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, Massachusetts.
- Stephen C. Levinson (2013). Action formation and ascription. In Tanya Stivers and Jack Sidnell, editors, *The Handbook of Conversation Analysis*, pages 103–130. Wiley-Blackwell, Malden, Massachusetts.
- Stephen C. Levinson and Francisco Torreira (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6: 731. <https://doi.org/10.3389/fpsyg.2015.00731>
- Xiaoting Li (2016). Some interactional uses of syntactically incomplete turns in Mandarin conversation. *Chinese Language and Discourse*, 7(2), 237–271. <https://doi.org/10.1075/cld.7.2.03li>
- Per Linell (2013). The dynamics of incrementation in utterance-building: Processes and resources. In Beatrice Szczepek Reed and Geoffery Raymond, editors, *Units of Talk – Units of Action*, pages 57–89. John Benjamins Publishing Company, Amsterdam, Noord-Holland.
- Lilla Magyari and Jan Peter de Ruiter (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, 3:367. <https://doi.org/10.3389/fpsyg.2012.00376>
- William C. Mann and Sandra A. Thompson (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281. <https://doi.org/10.1515/text.1.1988.8.3.243>
- Yael Maschler and Deborah Schiffrin (2015). Discourse markers, language, meaning, and context. In Deborah Tannen, Heidi E. Hamilton and Deborah Schiffrin, editors, *The Handbook of Discourse Analysis (Second Edition)*, pages 189–221. John Wiley & Sons, Chichester, West Sussex.
- Jacques Mehler, Jean Yves Dommergues, Uli Frauenfelder and Juan Segui (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20:298–305. [https://doi.org/10.1016/S0022-5371\(81\)90450-3](https://doi.org/10.1016/S0022-5371(81)90450-3)
- Marianne Mithun (2020). Prosody and the organization of information in Central Pomo, a California indigenous language. In Shlomo Izre'el, Heliana Mello, Alessandro Painunzi and Tommaso Raso, editors, *In Speech of Basic Units of Spoken Language: A Corpus-driven Approach*, pages 107–126. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania.
- Jacqueline Monschau, Rolf Kreyer and Joybrato Mukherjee (2003). Syntax and semantics at tone unit boundary. *Anglia*, 121:581–609. <https://doi.org/10.1515/ANGL.2003.581>
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty and Daniel Zeman (2016). Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666, Portorož, Piran. <https://aclanthology.org/L16-1262>
- Tsuyoshi Ono and Sandra A. Thompson (2017). Negative scope, temporality, fixedness, and right- and left-branching: Implications for typology and cognitive processing. *Studies in Language*, 41(3):543–576. <https://doi.org/10.1075/sl.41.3.01ono>

- Vitória Piai, Ardi Roelofs, Joost Rommers, Kristoffer Dahlsätt and Eric Maris (2015). Withholding planned speech is reflected in synchronized beta-band oscillations. *Frontiers in Human Neuroscience*, 9:549. <https://doi.org/10.3389/fnhum.2015.00549>
- Martin J. Pickering and Chiara Gamb (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144: 1002–1044. <https://doi.org/10.1037/bul0000158>
- Martin J. Pickering and Simon Garrod (2021). *Understanding Dialogue Language Use and Social Interaction*. Cambridge University Press, Cambridge, East of England.
- Paola Pietrandrea, Sylvain Kahane, Anne Lacheret and Frédéric Sabio (2014). The notion of sentence and other discourse units in corpus annotation. In Tommaso Raso and Heliana Mello, editors, *Spoken Corpora and Linguistic Studies*, pages 331–364. John Benjamins Publishing Company, Amsterdam/Philadelphia, Noord-Holland/Pennsylvania. <https://benjamins.com/#catalog/books/sc1.61.12pie/details>.
- Laurent Prévot, Shu-Chuan Tseng, Klim Peshkov and Alvin Cheng-Hsien Chen (2015). Processing units in conversation: A comparative study of French and Mandarin Data. *Language and Linguistics*, 16(1):69–92. <https://doi.org/10.1177/1606822X14556605>
- Marlou Rasenberg, Asli Özyürek, Sara Bögels and Mark Dingemanse (2022). The primacy of multimodal alignment in converging on shared symbols for novel referents. *Discourse Processes*, 59(3):209–236. <https://doi.org/10.1080/0163853X.2021.1992235>
- Geoffrey Raymond (2003). Grammar and social organization: Yes/No interrogatives and the structure of responding. *American Sociological Review*, 68(6):939–967. <https://psycnet.apa.org/doi/10.2307/1519752>
- Geoffrey Raymond (2004). Prompting action: The stand-alone “so” in ordinary conversation. *Research on Language and Social Interaction*, 37:185–218. https://doi.org/10.1207/s15327973rlsi3702_4
- Carina Riest, Annett B. Jorschick and Jan Peter de Ruiter (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology*, 6:89. <https://doi.org/10.3389/fpsyg.2015.00089>
- Harvey Sacks, Emanuel A. Schegloff and Gail Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735. <https://doi.org/10.2307/412243>
- Emanuel A. Schegloff (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In Deborah Tannen, editor, *Analyzing Discourse: Text and Talk*, pages 71–93. Georgetown University Press, Georgetown, Washington DC.
- Emanuel A. Schegloff (1996a). Turn-organization: One intersection of grammar and interaction. In Elinor Ochs, Emanuel A. Schegloff and Sandra A. Thompson, editors, *Interaction and Grammar*, pages 52–133. Cambridge University Press, Cambridge, East of England.
- Emanuel A. Schegloff (1996b). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, 102(1):161–216. <https://doi.org/10.1086/230911>
- Emanuel A. Schegloff (2007). *Sequence Organization in Interaction*. Cambridge University Press, Cambridge, East of England.
- Emanuel A. Schegloff (2013). Ten operations in self-initiated, same turn repair. In Makoto Hayashi, Geoffrey Raymond and Jack Sindell, editors, *Conversational Repair and Human Understanding*, pages 41–70. Cambridge University Press, Cambridge, East of England.
- Emanuel A. Schegloff, Gail Jefferson and Harvey Sacks (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382. <https://doi.org/10.2307/413107>
- Margret Selting (2000). The construction of units in conversational talk. *Language in Society*, 29(4):477–517. <https://www.jstor.org/stable/4169050>
- Xiao-nan Susan Shen (1990). *The Prosody of Mandarin Chinese*. University of California Press, Berkeley, California.

- Matthias Sjerps and Antje S. Meyer (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, 136:304–324. <https://doi.org/10.1016/j.cognition.2014.10.008>
- Gabriel Skantze (2021). Turn-taking in conversational system and human-robot interaction: A review. *Computer Speech & Language*, 67:101178. <https://doi.org/10.1016/j.csl.2020.101178>
- Neal Snider and Inbal Arnon (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In Stefan Th. Gries and Dagmar Divjak, editors, *Frequency Effects in Language*, pages 127–164. Mouton de Gruyter, Berlin.
- Manfred Stede (2012). *Discourse Processing*. Morgan & Claypool Publishers, San Rafael, California.
- Gerard Steen (2005). Basic discourse acts: Towards a psychological theory of discourse segmentation. In Francisco J. Ruiz de Mendoza and M. Samdra Peña, editors, *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*, pages 283–312. Mouton de Gruyter, Berlin.
- Tanya Stivers, Nick J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Geritje Hoyman, Federico Rossano, Jan Peter de Ruiter, Kyung-Eun Yoon and Stephen C. Levinson (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26):10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Takashi Suzuki and Mayumi Usami (2005). Co-constructions in English and Japanese revisited: A quantitative approach to cross-linguistic comparison. In *The Ninth International Pragmatics*, pages 263–276, Riva del Garda, Trento.
- Beatrice Szczepek Reed (2010). Prosody and alignment: a sequential perspective. *Culture Studies of Science Education*, 5:859–869. <https://doi.org/10.1007/s11422-010-9289-z>
- Beatrice Szczepek Reed (2011). Beyond the particular: Prosody and the coordination of actions. *Language and Speech*, 55(1):13–34. <https://doi.org/10.1177%2F0023830911428871>
- Hongyin Tao and Junfei Hu (2019). Structural, semantic, and pragmatic properties of *nong* (弄) constructions in Mandarin discourse: Evidence from corpora. *International Journal of Chinese Linguistics*, 6(1):162–176. <https://doi.org/10.1075/ijchl.18003.tao>
- Emma M. Templeton, Luke J. Chang, Elizabeth A. Reynolds, Marie D. Cone LeBeaumont and Thalia Wheatley (2023). Long gaps between turns are awkward for strangers but not for friends. *Philosophical Transactions of the Royal Society B*, 378(1875):20210471. <https://doi.org/10.1098/rstb.2021.0471>
- Lucien Tesnière (1959). *Éléments de Syntaxe Structurale*. Klincksieck, Paris, Ile-de-France.
- Sandra A. Thompson and Elizabeth Couper-Kuhlen (2005). The clause as a locus of grammar and interaction. *Discourse Studies*, 7(4/5):481–505. <https://doi.org/10.1177%2F1461445605054403>
- Sandra A. Thompson and Paul J. Hopper (2001). Transitivity, clause structure, and argument structure: evidence from conversation. In Joan Bybee and Paul Hopper, editors, *Frequency and the Emergence of Linguistic Structure*, pages 27–60. John Benjamins Publishing Company, Amsterdam, Noord-Holland/Pennsylvania.