Journal of Pragmatics 43 (2011) 3001–3027

Contents lists available at ScienceDirect

Journal of Pragmatics

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

Pragmatic aspects of temporal accommodation in turn-taking

Štefan Beňuš [a,b,*], Agustín Gravano [c], Julia Hirschberg [d]

[a] Constantine the Philosopher University, Štefánikova 67, 94974 Nitra, Slovakia
[b] Slovak Academy of Sciences, Institute of Informatics, Dúbravská cesta 9, 84507 Bratislava, Slovakia
[c] Departamento de Computación (FCEyN) and Laboratorio de Investigaciones Sensoriales, Universidad de Buenos Aires, Pabellón I, Ciudad Universitaria, 1428 Buenos Aires, Argentina
[d] Columbia University, 1214 Amsterdam Avenue, 450 CS Building MC 0401, New York, NY 10027, USA

ARTICLE INFO

Article history:
Received 1 April 2010
Received in revised form 20 May 2011
Accepted 21 May 2011
Available online 29 June 2011

Keywords:
Turn-taking
Accommodation
Rhythm
Grounding response
Dominance

ABSTRACT

This study investigates the relationship between the variability in the temporal alignment of turn initiations and the pragmatics of interpersonal communication. The data come from spontaneous, task-oriented dialogues in Standard American English. In addition to analyzing the temporal aspects of turn-taking behavior in general, we focus on the timing of turn-initial single word grounding responses such as *mmhm*, *okay*, or *yeah*, and conversational fillers such as *um* or *uh*. Based on qualitative and quantitative analyses of temporal and rhythmic alignment patterns, we propose that these patterns are linked to the achievement of pragmatic goals by interlocutors. More specifically, we examine the role of timing in establishing common ground, and test the hypothesis that the degree of accommodation to temporal and metrical characteristics of an interlocutor's speech is one aspect of turn-taking behavior that signals asymmetrical dominance relationships between interlocutors. Our results show that dominance relationships linked to floor-control, as well as mutual common ground, are pragmatically constructed in part through the accommodation patterns in timing of turn-initial single word utterances.

1. Introduction

The goal of this research is to improve our understanding of the relationship between prosodic form and pragmatic meaning by studying the temporal alignment of turn-initial single-word responses such as *mmhm*, *um*, *okay*, *uh*, or *yeah* in English task-oriented dialogues. We propose that this alignment is linked to the achievement of pragmatic goals by interlocutors. More specifically, we examine the role of timing in establishing common ground, and test the hypothesis that the degree to which temporal and metrical characteristics of interlocutors' speech become similar – and the directionality of this phenomenon – is one aspect of turn-taking behavior that signals asymmetrical dominance relationships between the interlocutors. Our approach combines qualitative conversational analyses of temporal/rhythmic patterns based on representative examples with corpus-based quantitative and statistical analyses testing the validity and robustness of the observed patterns.

1.1. Turn-taking and accommodation

Turn-taking is a cognitive, dynamically evolving, pragmatic system that is fundamental for human interaction (e.g. Schegloff, 2007). The turn-taking system is also fundamentally cross-modal: it is pervasive in both speech and sign

* Corresponding author at: Constantine the Philosopher University, Štefánikova 67, 94974 Nitra, Slovakia. Tel.: +421 37 6408455.
E-mail addresses: sb513@nyu.edu, sbenus@ukf.sk (Š. Beňuš), gravano@dc.uba.ar (A. Gravano), julia@cs.columbia.edu (J. Hirschberg).

0378-2166/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.pragma.2011.05.011

language and is strongly linked to paralinguistic domains such as gaze and gestures (e.g. Coates and Sutton-Spence, 2001; Goodwin, 1996). Broadly speaking, a turn-taking floor-management organization underlies the decisions of 'who speaks when' and must include at least three components: (1) ways of signaling and perceiving cues for ***transition-relevance places*** (TRPs) and turn allocation among interlocutors (e.g. Sacks et al., 1974), (2) ways of achieving suitable durations of latencies between turns, avoiding over-long overlaps or silent pauses, and (3) ways of resolving disruptions in the system (e.g. Schegloff, 2000).

Prosody plays an important role in the system of turn-taking because it participates in all three components. Extensive research has been done on prosodic (as well as syntactic and pragmatic) cues for TRPs (e.g. Ford and Thompson, 1996; Selting, 1996; Ward and Tsukahara, 2000; Gravano and Hirschberg, 2011). In terms of temporal aspects of turn-initiations, the avoidance of both extended overlaps and long silences is assumed to play an important role in turn-organization (e.g. the principle of "no gaps/no overlaps" in Sacks et al., 1974), with notions of what constitutes extended overlaps and long pauses being culturally determined (e.g. Sidnell, 2001). For example, Jefferson (1986) argued for an unmarked switch between speakers that involves "neither haste nor delay" (Jefferson, 1986:162); in her widely used transcription scheme, deviations from this unmarked case, including ***overlap***, ***perfect latch***, and extended silence, are labeled and assumed to be communicatively meaningful.[1]

In addition to the avoidance of gaps and overlaps, several studies have contended that the temporal alignment of turn initiations is meaningfully construed through their incorporation into the rhythmic patterns of the preceding turns (Couper-Kuhlen, 1993; Auer et al., 1999; Szczepek Reed, 2010). These studies argued that conversational partners perceive speech as somewhat isochronous; that is, they perceive "constancy of intervals between rhythmic events" (Auer et al., 1999:24). Interlocutors then synchronize in their turn-productions to fit into this rhythmic isochrony. Auer et al. further proposed that the fundamental unit of such isochrony in English is the 'beat', which roughly corresponds to the temporal interval between prominent syllables associated with a pitch accent. Erickson (2004) argued that the temporal alignment of prominent syllables as well as various gestures among family members at a dinner table talk converges on a perceptually salient rhythm with a high degree of isochrony, and that the convergence of rhythmical structures signals social alignment among the participants.

Such incorporation of prior rhythmic patterns assumes that the metrical and rhythmical aspects of prosody related to turn-taking are subject to accommodation among conversational partners. We understand ***accommodation*** as a dynamic process through which the behavior of one person becomes influenced by, and eventually more similar to, the behavior of his/her conversational partner, based on their shared representations at a particular level.[2] Pickering and Garrod (2004)[3] reviewed the body of evidence for accommodation at the prosodic and pragmatic levels, in addition to the lexical and syntactic levels, and proposed a mechanistic view of conversation, in which the linguistic representations at many levels become aligned among the interlocutors through a priming mechanism, which greatly facilitates the interactions. Moreover, accommodation has been found in the metrical features of utterances (e.g. Auer et al., 1999), in intensity characteristics (Ward and Litman, 2007), in phonetic and prosodic characteristics of individual words (Pardo, 2006), and in accent and other socio-phonetic variables (e.g. Gregory and Webster, 1996; Aubanel and Nguyen, 2010). Additionally, Brazil et al. (1980) used the term 'pitch concord' for turn onsets at the same pitch level as a previous speaker's turn completion; Couper-Kuhlen (1996) described participants' matching of pitch register and its interactional implications; and Szczepek Reed (2006) described prosodic accommodation between turns as 'prosodic orientation'. Finally, at the paralinguistic level, conversational partners entrain their body swaying motions (Shockley et al., 2003), and breathing (McFarland, 2001). Scott et al. (2009) proposed that neural pathways responsible for smooth turn-taking monitor the rhythm and rate of incoming speech, thus facilitating the interactional synchrony and communicative convergence in conversations.

In addition to facilitating interactions, accommodation among the interlocutors through spoken interactions is believed to be prominent in strategic negotiations of social distance among conversational partners (Giles et al., 1991). Particularly relevant to our study is the use of linguistic means, such as varying speech rate or pitch range, for creating, maintaining, or decreasing this distance.[4]

The fundamental question that we address in the present study is how variability in, and accommodation to, the timing patterns of turn initiations affects the evolution of common ground understanding and the power relationship between interlocutors during task-oriented spoken dialogues. The remainder of this section discusses the relationship between turn-taking behavior on the one side and common ground and dominance on the other, motivates the selection of grounding responses and conversational fillers as the focus for our investigation, describes our research questions in more detail, and presents our approach to answering them.

[1] An overlap occurs when a new speaker starts her turn before the current speaker finishes hers and a perfect latch occurs when a new speaker's turn is aligned precisely to the end of the current speaker's turn.

[2] This broad concept, or some particular aspects of it, is also commonly referred to as 'entrainment', 'alignment', 'convergence', 'priming', or 'adaptation'. We will use the term *accommodation* throughout the paper except when reviewing other approaches. *Alignment* in this paper will be used only to describe a temporal relationship – for example, the start of a backchannel is aligned 0.3 s after the end of the preceding utterance.

[3] This is a target article that is followed by multiple squibs in open peer commentary and finally a response to the comments by the authors.

[4] In a review of accommodation in communicative interaction, Giles et al. (1991, Tables 1 and 2, pp. 7 and 11) discussed additional features that have been showed to accommodate in conversations such as information density, self-disclosure, head nodding, and other phenomena; and several characteristics of accommodation such as its direction (upward vs. downward), modality (unimodal vs. multimodal), and symmetry (symmetrical vs. asymmetrical).

1.2. Turn-taking and common ground

Following Clark (1996), we construe ***common ground*** as mutual knowledge that is shared among interlocutors and that is known to be shared by them. Grounding is thus a basic principle of discourse organization through which information is collaboratively acknowledged as mutually shared by conversational participants. The collaborative nature of grounding suggests that this process is facilitated by the accommodation between interlocutors. Despite the wealth of research analyzing the relationship between the process of establishing common ground and intonation (e.g. Brown, 1983; Pierrehumbert and Hirschberg, 1990; Steedman, 2000; Dahan et al., 2002), its temporal aspects have not been widely studied. Mushin et al. (2003) looked at the relationship between prosody and the complexity of ***common ground units*** (CGUs). A simple CGU consists of one adjacency pair, and a complex one contains more than one pair. In addition to a systematic relationship with intonation, they found that complex CGUs were realized with more overlaps between the two turns of the CGU's first adjacency pair, and more latches between the end of the first adjacency pair and the following utterance, than simple CGUs. Hence, an initial overlap correlated with the need for further elaboration before common ground was established.

Shimojima et al. (2002) asked how prosody affects the integration rate of information in Japanese echoic responses that repeat some of the material used in the preceding turn. They found that longer delays between turns signaled significantly lower integration rates, and thus less effective establishment of common ground, than shorter delays.

Finally, Fox Tree (2002) investigated how a turn-initial silent and/or filled pause affects the perception of a speaker's second turn in question-answer adjacency pairs. She found that speakers were perceived as having more production difficulty, and being less honest and less comfortable with topics, when they started their answer with *um* or a silent pause – the effect was even more pronounced when they used both. Hence, turn-initial silent pauses and conversational fillers appear to signal less certainty toward the proposition expressed in the preceding utterance.

In sum, turn-initiations that either overlap the preceding turn or start long after the turn is finished do not seem to be conducive to smooth common ground establishment.

1.3. Turn-taking and dominance

While mutual accommodation facilitates successful grounding, asymmetries in accommodation to an interlocutor's speech may be utilized for constructing power relationship. We consider dominance as one instantiation of power that is mutually negotiated through the use of linguistic signals, and construe dominance as a communicative strategy (e.g. Poggi and D'Errico, 2010). Following dyadic power theory (Dunbar and Burgoon, 2005), we assume that the dominance of an individual depends on the submissiveness of other participants as negotiated during a conversation. One way this submissiveness may be realized is through the accommodation of a less dominant speaker to the linguistic behavior of a more dominant speaker. There are several cues to dominance that can be studied through linguistic means. For example, one might examine the loudness of voice or the slope of pitch at the end of utterances. Our focus in this paper, however, is on turn-taking behavior: we study dominance in relation to floor-control.

In this respect, dominant individuals in socially or institutionally imbalanced environments, such as work places, schools, job interviews, or court proceedings (e.g. Andersen and Bowman, 1999; Gnisci and Bakeman, 2007), have been found to speak in longer turns, hence holding the conversational floor for a relatively longer time than their interlocutors; they also tended to interrupt their interlocutors more often. Research on conversational dominance in gender studies has also identified interactional features that are indicative of the type of conversational dominance typically associated with male speech, such as more overlaps, greater frequency of interruptions, and initiations of topic changes (e.g. Itakura and Tsui, 2004). One must bear in mind, however, that short overlaps have also been shown to signal a highly collaborative structure in spoken interactions. Tannen (1998) discusses so called ***cooperative overlaps*** that function to support, affirm, or acknowledge what the other speaker is saying.

In sum, interlocutors in the position of power tend to signal their dominance by starting their turns before the preceding turn has been completed, but similarly overlapped turns signaling agreement, affirmation, or acknowledgment are considered cooperative and thus not commonly linked to dominance. Less dominant interlocutors might accommodate by simply etting themselves be interrupted or not interrupting the interlocutor, which results in longer and more frequent turns of more dominant interlocutors.

1.4. Single word grounding responses and conversational fillers

In addition to our general focus on the temporal aspects of turn-taking behavior, we are particularly interested here in the timing of turn-initial ***single word grounding responses*** (SWGRs) and ***conversational fillers*** (CFs). SWGRs include positive polarity items such as *mmhm*, *okay*, *yeah*, or *uhhuh* that participate in creating common ground, and can serve the pragmatic functions of backchannel, agreement, or acknowledgments. CFs such as *um* and *uh* are linguistic expressions connected to the cognitive load and/or planning difficulties associated with a choice (see Stewart and Corley, 2008 for a recent review). Speakers tend to use CFs to signal pragmatic, discourse, or syntactic boundaries (Swerts, 1998; Ferreira et al., 2004) and other functions (Clark and Fox Tree, 2002).

There are several reasons for our focus on SWGRs and CFs. First, they display rich ambiguity expressed through the mapping between their prosodic realization and pragmatic meaning, as is the case for many discourse markers. In other words, *how* these

lexical items are produced is closely linked to *what* they convey pragmatically. Several studies have already established the effect of intonational contours on the pragmatic meanings of cue words (e.g. Hirschberg and Litman, 1993; Hirschberg and Nakatani, 1996; Schiffrin, 1987). We aim to improve our understanding of the mapping between prosodic form and pragmatic meaning of these words by studying their timing in turn-initial position and its effect on the pragmatic structure of the discourse.

Second, they play several roles in spoken interactions that are related to grounding. SWGRs facilitate the addition of preceding information into the stack of concepts describing the common ground shared between interlocutors. Turn-initial CFs, on the other hand, may signal uncertainty (Fox Tree, 2002) and thus a need for further elaboration. These CFs also facilitate both production and perception of linguistic material because they allow speakers to plan their intended message and listeners to prepare to perceive important content. Listeners have been shown to be highly sensitive to the occurrence and timing of CFs in speech (e.g. Brennan and Williams, 1995). CFs facilitate the process of comprehension by helping listeners better predict information in upcoming speech, and by enhancing retention of words preceded by CFs in memory (Stewart and Corley, 2008). CFs are also necessary in managing spontaneous-like conversations (Bortfeld et al., 2001; Taboada, 2006), and thus, in general, should be "understood as devices with important turn-organizational uses" (Sacks et al., 1974:720). Hence, both CFs and SWGRs play a prominent role in the pragmatics of floor and information management, and thus they are crucial for better understanding of interpersonal conversations.

Third, in addition to the temporal aspects of grounding with SWGRs and CFs mentioned above, these both play a role in establishing and maintaining power relationships. Due to their positive polarity (SWGRs) and uncertainty (CFs), they are likely to be more frequent in the speech of less dominant interlocutors. Moreover, while turn-initiations produced before the preceding turn ends tend to be associated with dominance, similarly timed turn-initial SWGRs may signal submissiveness and cooperation.

Lastly, these items are by definition well delimited prosodically; they typically form a single intonational phrase and are preceded and followed by silence (from the speaker). Speakers were found to accommodate both SWGRs and CFs intonationally to the surrounding material (Heldner et al., 2010; Shriberg and Lickley, 1993). They also occupy the turn-initial position, which represents a critical location in the process of turn-construction and action formation (Schegloff, 1996). Moreover, because they are relatively frequent, focusing on them provides us with good coverage and variability, and enables statistical testing.

1.5. Hypotheses and approaches

In this paper we study how speakers initiate turns in general and focus on turns starting with SWGRs and CFs. One of the pragmatic functions of such turns is clearly linked to the establishment of common ground, since SWGRs signal a successful addition while CFs signal uncertainty, hesitation, and thus the lack of success in adding a new proposition or concept into the mutual common ground. We particularly concentrate on the **timing** of these turn-initiations and hypothesize that this timing systematically participates in common ground establishment and that it also plays a role in the development and maintenance of asymmetrical dominance relationship between interlocutors. A related hypothesis that we test is that floor-control dominance is inversely related to the degree of accommodation of a speaker to the temporal and metrical characteristics of their interlocutor's speech, and that it can emerge in dialogues in which two speakers begin with an equal power status.

Our material comes from dyadic task-oriented conversations. We ask the following questions:

- What pattern of timing of turn-initiations (if any) best characterizes the process of common ground establishment?
- Who controls the conversational floor more?
- Who accommodates their timing of turn-initiations more?
- What is the relationship between the degree of this type of accommodation and more traditional measures of dominance such as frequency of interruptions?

We approach these questions in several ways. For example, we measure raw ***latency*** (difference between current turn beginning and previous turn end), especially for turns starting with SWGRs and CFs, and look for patterns in the distribution of these latencies. We study the development of these latencies over time in particular symptomatic examples. However, describing speech patterns in terms of absolute characteristics such as latency durations in seconds may provide conceptual insights, but it also runs the danger of failing to generalize to other conversations. This is because all prosodic, and in fact all linguistically meaningful, aspects of speech are fundamentally relative. For these reasons, we complement the information from raw latencies with features describing the relationship between latency and previous speech. For example, as discussed above, several studies have argued that people perceive the speech of their conversational partners as rhythmically isochronous to some degree and are able to entrain to this perceived pattern by producing turns with rhythm similar to that of their interlocutor. One prediction of this model is that, if a speaker aligns her turn-initiation with the rhythm of the interlocutor's preceding utterance, turn-latency should positively correlate with speech rate. In other words, a speaker should start her turn sooner when the interlocutor's utterance is more rapid. In this way we can test our hypothesis that the differences between the two speakers in the degree of such accommodation are systematically related to floor-control dominance: the less dominant speaker should accommodate more to the more dominant speaker than vice versa.

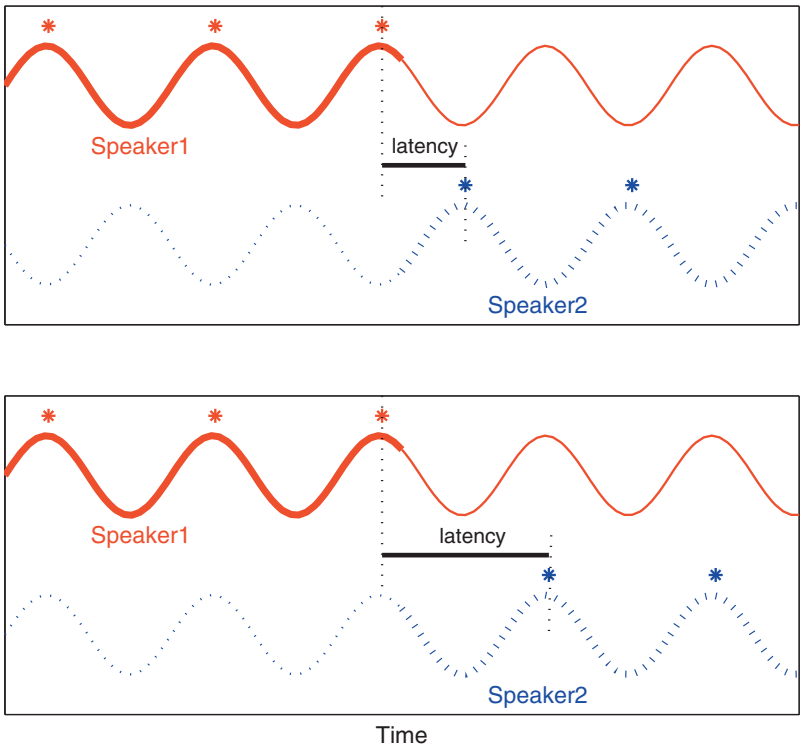

Fig. 1. Rhythmical entrainment as two oscillators defined by pitch accents (*).

Another model of rhythmical accommodation among interlocutors (Wilson and Wilson, 2005) proposes that the generation of the sequential structure of turn-taking in a dialogue can be captured by two dynamically defined oscillators, each representing the potential for initiating speech at any given moment for one interlocutor. Simplifying this notion for our purposes, we can consider that the peaks of such a function represent pitch accents; rate is the number of pitch accents in a turn divided by the turn duration; and latency is the time elapsed from the last pitch accent before the exchange and the first pitch accent after the exchange. The model is sketched in Fig. 1.[5]

The entrainment of a speaker to the rhythm of her interlocutor at a turn-exchange can thus be expressed in (1) as the ratio of the rate of pitch accents in the last utterance before the turn-exchange and the latency of the first pitch accent after the turn-exchange.

(1) Entrainment index (EI) = Latency/Rate

Wilson and Wilson's model, illustrated in the top panel of Fig. 1, assumes that the oscillators for two interlocutors are counter-phased: the peaks of one oscillator correspond to the valleys in the other. Then, perfect entrainment would correspond to EI values of 0.5, 1.5, 2.5, etc. If the interlocutors are in-phase, which is indirectly assumed in Couper-Kuhlen (1993) and Bull (1996) and illustrated in the bottom panel of Fig. 1, perfect entrainment would corresponds to EI values of 1.0, 2.0, 3.0, etc. In both cases, nevertheless, rhythmical entrainment among the interlocutors should result in a non-monotonic cluster-like distribution of EI values, with clusters separated by around 1 unit from each other. We will test the assumption that differences between speakers in their EI distributions patterns are related to their degree of accommodation, and consequently to the power relationship holding between them.

We also investigate the distribution of turn-taking types such as response elicitation, overlap, or interruption for the interlocutors and compare with observed patterns in dominance and accommodation. In sum, our approach attempts to integrate two ways of analyzing discourse. The first, typically conducted in the Conversational Analysis framework, presents insightful qualitative observations from individual transcribed examples and attempts to relate them to the general framework of sequence organization of turn-taking (e.g. Schegloff, 2000; Jefferson, 1986; Auer et al., 1999). The second approach, benefiting from advances in computational and corpus linguistics, studies the factors affecting timing patterns using large corpora of transcribed speech and (primarily) automatically extractable features from the acoustic signal, in an effort to improve the effectiveness and naturalness of interactive dialogue systems (e.g. Bull and Aylett, 1998; Yuan et al., 2007; Gravano and Hirschberg, 2009). These two approaches have to date produced somewhat contradictory results, for

[5] Wilson and Wilson's model assumes that a single oscillation cycle corresponds to the duration of one syllable rather than the interval between two pitch accents assumed in this paper. Data reported in Beňuš (2009) show a slightly better fit to the predictions of the Wilson and Wilson's model if the oscillation cycle is defined by pitch accents than by syllables.

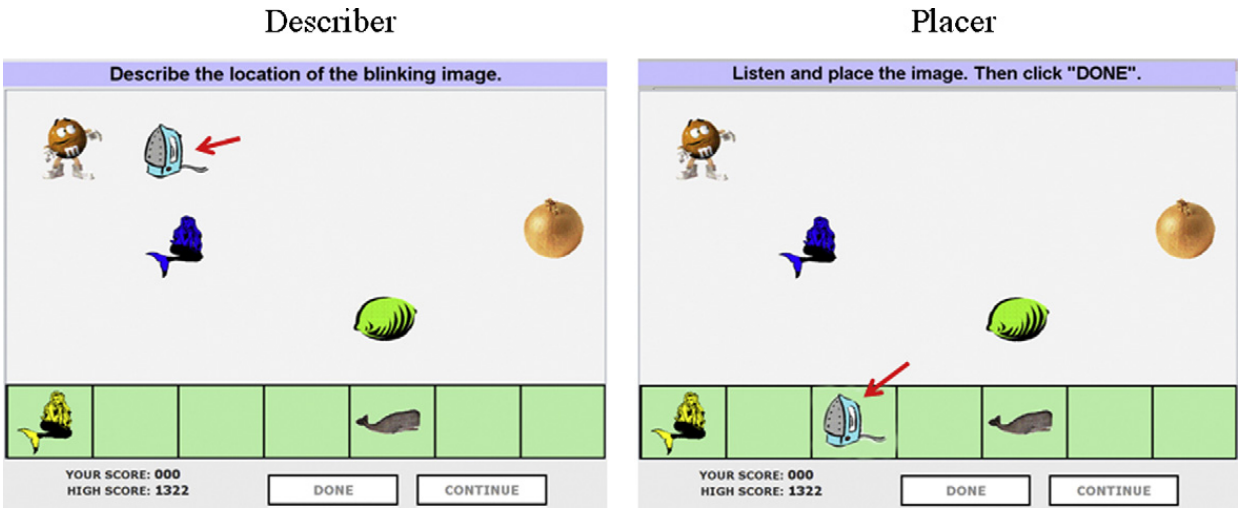

Fig. 2. Illustrations of the subjects' screens during one of the tasks of the OBJECTS game.

example with respect to rhythmical accommodation at turn boundaries. While several conversational analyses suggest that there is high-level rhythmical isochrony and wide-spread presence of accommodation among speakers at turn exchanges (e.g. Erickson, 2004; Auer et al., 1999; Szczepek Reed, 2010), larger corpus studies by and large fail to find robust quantitative support for the existence of these patterns (Bull, 1996; Beňuš, 2009). In this study, we test the possibility of combining a conversational analysis of temporal/rhythmic patterns based on representative examples and validating these observations with quantitative and statistical corpus-based analysis.

The remainder of this paper is structured as follows. Section 2 describes the methodology of data collection and annotation. Section 3 presents an in-depth qualitative analysis based on multiple examples drawn from our corpus, and section 4 follows with quantitative analyses that test the validity and robustness of the observations presented in section 4. Section 5 compares the turn-taking behavior of the two target speakers in their conversations with different interlocutors. Section 6 summarizes our findings and discusses broader implications of the study.

2. Methodology

2.1. Corpus

Our data are taken from the Columbia Games Corpus, a corpus of spontaneous task-oriented dialogues in Standard American English.[6] 13 Subjects (7 female and 6 male) played two types of collaborative games (CARDS and OBJECTS); 11 subjects played with two different partners in two different sessions, and 2 played a single session. The dialogues were recorded in a soundproof booth; subjects could not see each other due to a curtain hung between them. This scenario provided conversations in which interlocutors could not use visual cues such as facial expressions or body language and had to rely only on oral cues.

In this study we focus on the OBJECTS games. In these, one player (the Describer, whose screen is shown in the left panel of Fig. 2) describes the position of a target object with respect to other fixed objects on her screen, while the other player (the Placer, whose screen is shown in the right panel of Fig. 2) tries to move her representation of the target object to the same position on her own screen. Points are awarded based on the proximity of the target object to its correct location. The subjects switch the roles of Describer and Placer repeatedly. The recordings were orthographically transcribed, and words were aligned to the source acoustic signal by hand. On average, each OBJECTS game session took 21.5 min, totaling 4 h 29 m of dialogue for this corpus.

The target conversation discussed in this study involves two female speakers (A and B). This particular dialogue was selected for analysis because it is the longest dialogue in the corpus; this was the second session for both speakers so they were familiar with the game; and interlocutors were of the same gender. The speech analyzed in this conversation covers 35.7 min.[7] We also analyze the two (earlier) dialogues in which the target speakers (A and B) played the same games with different interlocutors. The analyzed speech in these two games is 25 and 15.2 min long, respectively.

[6] See Gravano (2009) for a detailed description of the collection and annotation of this corpus.

[7] The total length of the dialogue is 42.6 min but during one of the screens, one laptop went to sleep, so the game had to be re-started. The time and conversation during this break (5.3 min) are omitted from the analysis. Also omitted are times when speakers commented on their results, typically several seconds after each task.

Table 1
Discourse/pragmatic functions of affirmative cue words.

Agr	Agreement/acknowledgment	Chck	Check; "Is that okay?"
BC	Backchannel	PEnd	Pivot ending; Ack + CEnd
CBeg	Cue beginning discourse segment	Mod	Literal modifier
CEnd	Cue ending discourse segment	Stl	Stall/filler
BTsk	Back from a task	?	Cannot decide
PBeg	Pivot beginning; Ack + CBeg		

2.2. Data annotation

Temporal aspects of the dialogues relevant for this study include the identification of interpausal units (IPUs) and turns. We define an IPU as a maximal sequence of words from one speaker surrounded by silence longer than 50 ms. A turn is then a maximal sequence of IPUs from one speaker, such that between any two adjacent IPUs there is no speech from the interlocutor. Each IPU thus can be identified as either beginning, ending, or continuing a turn.

Prosodic and acoustic information comes from two sources. First, the corpus was intonationally transcribed using the Tone and Break Indices conventions (ToBI, Beckman et al., 2004). This labeling, among other things, provides information about which words receive pitch accents, and are thus intonationally prominent, and which are not. From each pitch-accented word we extracted the time of the acoustic energy peak as a rough estimate of the temporal point signaling prosodic prominence. The series of these temporal points was then used for the calculation of pitch-accent rate as the primary rhythmical feature in this study. The second source of prosodic information comes from continuous acoustic features for pitch and intensity that were automatically extracted from the signal using Praat software (Boersma and Weenink, 2005).

All single-word positive polarity items (termed affirmative cue words or ACWs) were identified and labeled for their discourse/pragmatic meanings (Gravano, 2009; Benus et al., 2007). Three labelers used both speech and transcripts for assigning one of 10 functions listed in Table 1. Inter-labeler reliability was measured by Fleiss' kappa (Fleiss, 1971) at 0.69.[8] In this study we use majority labels, where at least 2 labelers assigned a token to the same class. The target conversation of this study contains 531 such items, of which 480 have a majority label. SWGRs thus correspond to the Agr and BC categories.[9]

Turn-taking behavior was characterized using a slightly modified annotation scheme based on Beattie (1982), illustrated in Fig. 3 (Gravano and Hirschberg, 2009). Two labelers, who were different from the three labelers of positive polarity items, annotated each switch between the speakers in the following way. First, the presence of simultaneous speech between the speakers' turns was determined automatically. Since the back-channel annotation (BC labels) was available from the labeling described above, the decision as to whether a turn is a backchannel or not (at the root of the decision tree) was adopted without change from that annotation. Then, for all non-backchannel turns, if the exchange did not have simultaneous speech, the turns were labeled as Smooth Switches (S) if the preceding utterance was complete and as Pause Interruptions (PI) if the preceding turn was not complete. We used Beattie's informal definition of utterance completeness: "Completeness was judged intuitively, taking into account the intonation, syntax, and meaning of the utterance" (Beattie, 1982:100). If simultaneous speech was present, the turns were labeled as Butting-ins (BI) if the speaker did not succeed in grabbing the floor, and as Overlaps (O) or Interruptions (I) if the speaker did take the floor. Here, the switch was labeled Overlap if the previous speaker's utterance was complete and Interruption if it was syntactically, semantically or intonationally incomplete.

Three additional turn-related features were annotated: X1, X2, and X3. The most relevant for this study is X2, which represents a continuation of previous speech by the same speaker after a backchannel (BC_NO, or BC_O) from the other speaker.[10] The annotators labeled separately and reached Cohen's κ score (Cohen, 1960) of 0.91 corresponding to 'almost perfect' agreement. After correcting potential labeling errors, the κ score improved to 0.99 and the remaining unresolved disagreements were assigned the label "?". Finally, we also have annotations for questions and IPUs eliciting a response from the interlocutor, made jointly by two expert annotators.

3. Qualitative analysis of the target conversation

3.1. Complete task 1

The dialogue below took place when the subjects were solving the first task of the game illustrated in the two screens of Fig. 2. In the transcription below, numbers in brackets show the duration of silences within turns, bold numbers show the

[8] According to Fleiss, values between 0.6 and 0.8 correspond to substantial agreement. Some authors consider Fleiss' descriptions arbitrary, depending on the number of categories, and thus controversial; Bakeman and Gottman, for example, are inclined to regard values of k less than .7 "with some concern" (1997:66).

[9] The guidelines for labelers identified Agr category as indicating "I believe what you said", and/or "I agree with what you say", and BC as a response to another speaker's utterance that indicates only "I'm still here/I hear you and please continue". We do not analyze Pivot beginning (PBeg) and Pivot ending (PEnd) in this paper since a) there are not many of them and their inclusion would skew the distribution, and b) we found that cue-beginnings differ robustly and significantly in their prosody from Agr and BC items (Gravano et al., 2007).

[10] X1 label was used for turns that begin a new task, that is, the first turn after the change on the laptop screens. X3 marked a simultaneous start. If two turns began almost simultaneously (formally, within 210 ms of each other, see Fry, 1975) then both speakers were most probably reacting to the preceding turn. Both X1 and X3 labels are not considered in the quantitative section of this study.

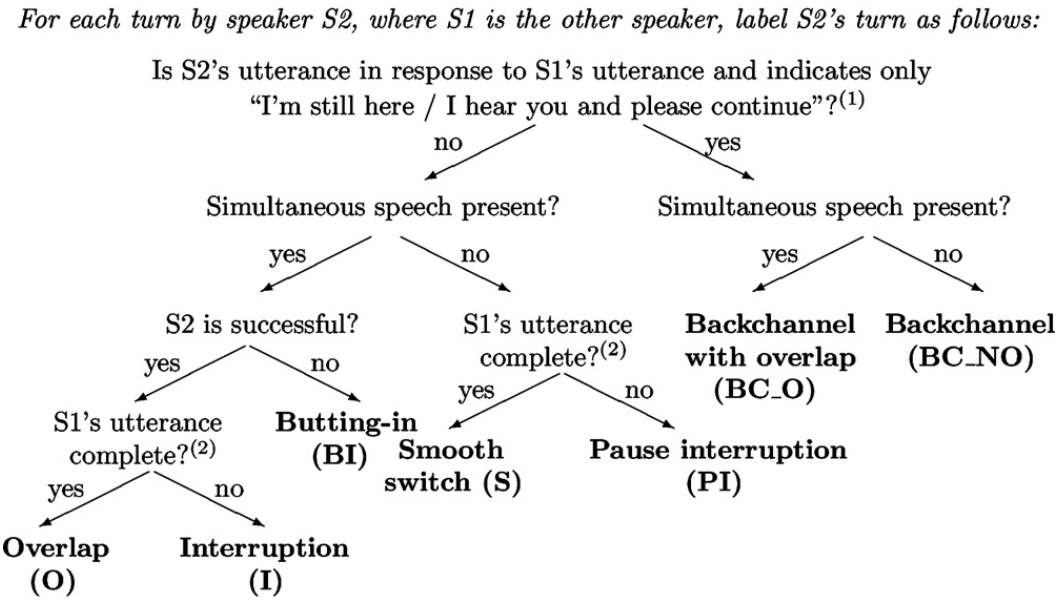

Fig. 3. Scheme for labeling turn-exchanges.
Adapted from Beattie (1982).

duration of silences across turns, and all transitions without brackets occurred without perceivable pauses. Utterance-final rising, falling, and level melody curves are shown with arrows ↑, ↓, and →, respectively; square brackets indicate overlaps; question marks in the transcription indicate undecipherable words.[11] Labels in small caps refer to DAMSL dialogue act tags (Dialogue Act Markup in Several Layers, Core and Allen, 1997). This scheme, originally motivated by Speech Act theory, defines a set of primitive communicative acts that are used primarily for manipulating common ground, and are thus suitable for a crude description of the interactional context typically present in collaborative task dialogues such as those in our corpus.

(2) Complete Task 1 of the target session (duration: 92 s)

1. A: *okay um* (0.68) *the blinking image is the iron*↑ (ASSERT, **0.18**)
2. B: *okay* → (ACKNOWLEDGE, **1.19**)
3. A: *and* (0.7) *it's ali-* (ABANDONED, 0.18) *okay* (0.13) *it's* (0.31) *above the mermaid*↑ (1.54) *and* (0.34) *vertically it's aligned to the f-* (0.25) *to the foot of the M&M guy*↑ (0.42) *like to the bottom* (0.22) *of the iron*↑ (ASSERT, **0.13**)
4. B: *okay* (ASSERT, 0.26) *lines up*↑ (INFO-REQUEST, **0.36**)
5. A: *yeah it's it's almost it's just barely* (0.27) *like over*↓ (ANSWER, **0.45**)
6. B: *o*[*kay*] (ACCEPT)
7. A: [*but*] *it's basically that same* (1.1) *line um* (ANSWER-CONT. 0.52) *so the black part at the bottom of the iron*↑ (ASSERT, **0.08)**
8. B: *mmhm*↑ (ACKNOWLEDGE, **0.13**)
9. A: *it's not necessarily like on the same line as the white foot it's just a little bit over*↓ (ANSWER-CONT. **2.75**)
10. B: *um* (0.11) *but i- it's like* (0.12) *between* (0.19) *the m-* (INFO-REQUEST)
11. A: [*okay*] (ACKNOWLEDGE)
12. B: [*M&M*] *and the mermaid right like* → (INFO-REQUEST-CONT. **0.16**)
13. A: *uh well okay* (ACCEPT, 0.24) *?-* (0.16) *the tail*↑ (ANSWER **0.17**)
14. B: *mmhm*↑ (ACKNOWLEDGE, **0.08**)
15. A: *of the iron*↑ (ANSWER-CONT., **0.28**)
16 B: *mmhm*↑ (ACKNOWLEDGE, **0.52**)
17. A: *is* (2.37) *past the* (0.23) *it's* (0.42) *a little bit past the mermaid's body*↑ (ANSWER-CONT **0.81**)
18. B: [*okay*] (ACKNOWLEDGE)

[11] This is a partial transcript; only the annotations most relevant for the discussion are shown in the interest of readability. The arrows reflect the boundary tones of the Tones and Break Indices framework for labeling prosodic events (Beckman et al., 2004) in the following way: ↓ corresponds to L–L% tone, → corresponds to H–L% (or H- if the utterance was cut off), and ↑ to all others tones.

19. A: [*like when you l*]*ook at* (ABANDONED) *okay when you look at the lower left corner of the iron*↑ (ASSERT, **0.41**)
20 B: *oka*[*y*] (ACKNOWLEDGE)
21. A: [*w*]*here the turquoise stuff is*↑ [*and*] *you know the bottom* (0.16) *point* (0.29) *out*
22. B: [*mmhm*] (ACKNOWLEDGE)
23. A: *to the farthest left for that region*↑ (ASSERT, **0.01**)
24. B: *mmhm*↑ (ACKNOWLEDGE **0.49**)
25. A: *that point is aligned to the* (ABANDONED, 0.9) *it's just about aligned to the um* (0.24) *the blue fin*↑ (ASSERT, **0.61**)
26. B: [*l- l-*] [*l*]*ike the tip of the fi*[*n*]↑(INFO-REQUEST)
27. A: [*like to the*] *lef* [*t*]↑(ASSERT-CONT.)
28. A: [*y*]*eah it's just just about but the fin is a little more left*↓ (ANSWER **0.34**)
29. B: *oka*[*y*]↓(ACCEPT)
30. A: [*y*]*eah*↓ (ACCEPT, 2.02) *so it's sort of like the same situation as* (0.14) *how the bottom black part is* (0.44) *almost aligned to the white feet of the M&M guy*↑ (ASSERT, **0.04**)
31. B: *oka*[*y*]↓(ACCEPT)
32. A: [*ye*]*ah*↓(ACCEPT)

The first two lines in this example show a prototypical example of an adjacency pair in which the first utterance provides some information (or asks a question) and the second is a SWGR. The short latency between the turns (0.18 s) signals a temporally smooth adjacency pair, in which B readily processes the new information from the preceding turn. Line 3, however, shows that this smoothness between the speakers is short lived. Speaker A provides the first part of the adjacency pair and, after producing *mermaid* with rising intonation, she seems to expect the same pattern as that established in lines 1–2. But B does not respond, and A waits for a long 1.54 s. The next chance for B to provide some feedback to A's descriptions is also in line 3 after A produces the phrase ending in *M&M guy*; this time A waits for only 0.42 s before continuing. The approximately 1.00 s difference between the first and the second silent pauses in this turn is an example of A's local adjustment in the temporal sequencing of turn productions. Finally, speaker B produces her smoothly aligned response after *iron*, and the 0.13 s pause is virtually identical to the first pause (0.18 s) between lines 1 and 2.

B's response in line 4 is followed by a question. After answering, A receives no response within the temporal window of B's two preceding responses (0.18 s and 0.13 s); thus, A provides additional information, which results in a complete overlap with B's acceptance in line 6. Perhaps realizing that her previous response in line 6 was 'too late', B avoids another overlap by aligning her backchannel in line 8 with only a 0.08 s latency. Fig. 4 gives a visual representation of this adjustment.

The difference in the timing of B's responses in lines 6 and 8 represents another temporal adjustment in turn-initiation. Based on our discussion so far, it seems that both speakers attend to a mutually constructed relationship between the turn latencies of grounding responses and their 'degree of understanding': longer latencies signal sub-optimal understanding and shorter latencies signal more optimal understanding. Moreover, within only several turns, a salient instantiation of the boundary between 'smoothly-aligned' and 'loosely-aligned' turns seems to emerge at around 0.3 s.

The rest of the excerpt provides several additional examples of this emergent relationship between temporal patterns of turn initiations and the establishment of common ground. An almost 3 s pause at the end of line 9 shows that A not only yields the floor but is unwilling to self-select this time. After perceiving significant difficulties from B signaled by this long pause and subsequent multiple disfluencies in line 10, A takes the initiative again.[12] This time, A adjusts her strategy and uses rising intonation to elicit feedback after short utterances containing single concepts (*tail*, *iron*), and B provides two smoothly aligned backchannels in lines 14 and 16. Assuming this pattern, A expects another temporarily well-aligned grounding response after *mermaid's body* in line 17. When B's response in line 18 comes very late (after 0.81 s), A infers problems with B's comprehension, and explains the position of the object in another way. B's acknowledgment in line 20 comes after 0.41 s, which is an adjustment after the previous 0.81 s, and seems to be near the emerging boundary between smooth understanding on the one hand and problematic common ground creation on the other. It is unclear if A's addition *where the turquoise stuff is* is a response to B's presumed problems, or was planned independently. Speaker B then adjusts even more, and the following two responses in lines 22 and 24 come with no perceivable pauses. The next information from A elicits a response after *blue fin* in line 25 and a familiar pattern recurs: B's response is initiated with sub-optimal timing (0.61 s), which prompts A to provide additional information. The pre-final *okay* from B in line 29 comes with the latency of 0.34 s, and, despite this tight temporal alignment, A's *yeah* comes with a slight overlap. Speaker B adjusts again and produces her final *okay* with a 0.04 s latency. However, A's final *yeah* still overlaps B's *okay*.

When we look at the distribution of DAMSL tags in this dialogue, we see that the variability of communicative and interactive actions is low. There are 9 acknowledgments, 6 assertions, 5 accepts, and 3 tokens each of abandoned utterances, information requests, and answers. All tasks in the corpus are similar in that they involve placing a target object in relation to

[12] Note the use of discourse marker *okay* for this purpose in lines 11 and 13. This usage is very different from its use as the second member of adjacency pairs; see Gravano et al. (2007) for a discussion of *okay* functions in this corpus.

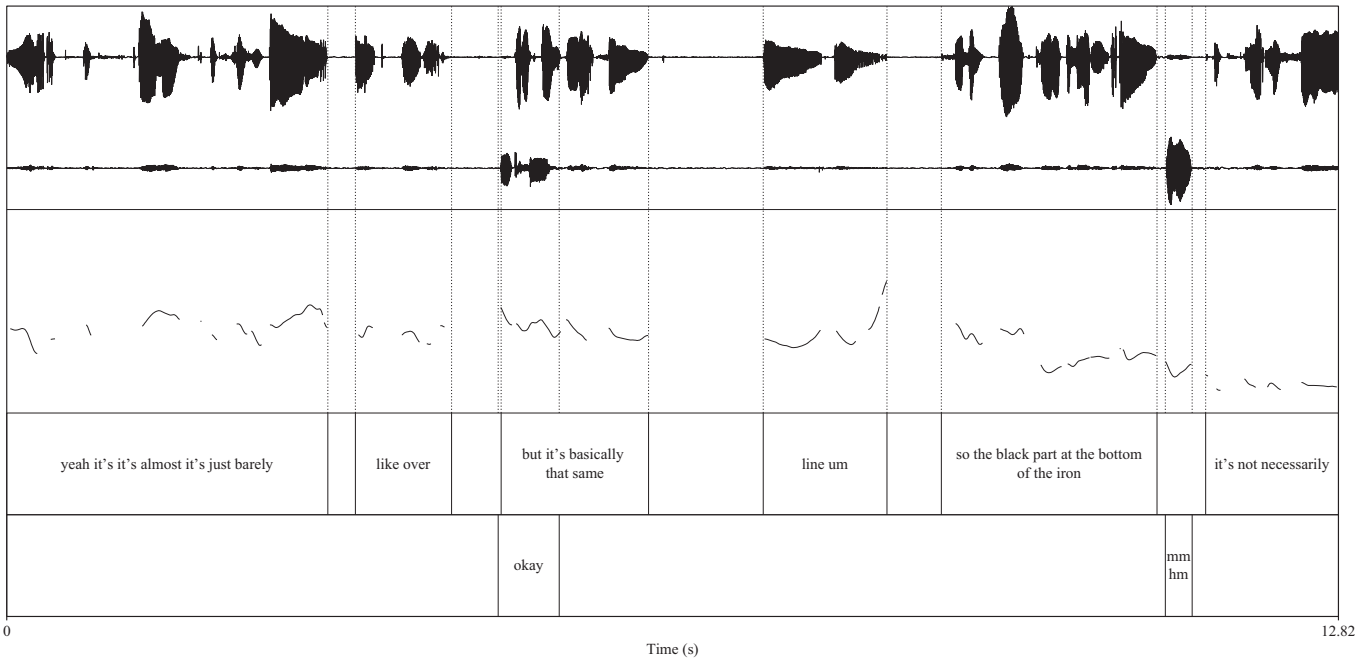

Fig. 4. Adjustment in the timing of two consecutive turn-initial grounding responses. The top panel shows the sound waves, the middle panel shows the fundamental frequency between 80 and 350 Hz, and the bottom two panels show the transcript.

other objects on the screen. Therefore, all dialogues in our corpus are similar to this one. In the DAMSL tagging scheme, the vast majority of communicative actions are Asserts, Info-requests, Accepts, Acknowledgements, and Answers. Typically, each task dialogue is naturally divided into two phases. First, the Describer produces a relatively long, monologue-like description, primarily achieved by Asserts and optionally acknowledged by the Placer. Afterwards, both speakers engage in a more interactive phase during which they refine the original description, until the Placer is satisfied. In this phase, Info-requests, Answers, and Accepts are the most common tags.[13]

To summarize the discussion of the excerpt in (2), we argued (a) that the timing of SWGRs such as *mmhm*, *okay*, or *yeah* signals pragmatic accommodation between interlocutors, (b) that this accommodation evolves dynamically, and (c) that it appears to exhibit two broad patterns: grounding turns with short latency (roughly less than 0.3 s) signal optimal understanding, and turns with latency longer than 0.5 s signal sub-optimal understanding. We also argued that speaker B is adjusting the timing of her responses more than speaker A. We conceive these temporal adjustments as one of the means by which speakers accommodate to one another. Therefore, the observed differences between the speakers in the deployment of this strategy suggest an asymmetrical power relationship. Moreover, this relationship can be observed in a discourse where two interlocutors of the same gender begin the conversation in the same power position. In sum, the analysis of Excerpt (2) suggested that the timing of turn-initiations is achieved interactionally and negotiated between interlocutors. In the following subsections we provide further support for this suggestion, focusing on the pragmatic domains of common ground and dominance.

3.2. Timing of conversational fillers and common ground

The main observation from the analysis of Excerpt (2) in section 3.1 was that the establishment of common ground between the interlocutors can be seen through the timing of turn-initial SWGRs with respect to the end of preceding turns, and that this type of accommodation evolves dynamically and seems to reach a stable critical value at around 0.3 s. In this section we expand this approach and test whether the timing of turn-initial CFs has a systematic relationship to common ground establishment that is similar to the one proposed for the grounding moves in section 3.1. Recall from section 1 that these CFs signal more production difficulty and less certainty with preceding utterances, and thus may imply problems in grounding.

Consider first the exchange in (3) below.

(3) Excerpt from task 4 of the target session (duration: 40 s)

1. B: mm let's see (0.24) so like how far up and down should I put it↓ **(0.08)**

[13] Due to the similarity of all tasks, the variability in the type and distribution of DAMSL tags across tasks is low. Excerpt (2) is thus a prototypical illustration of the types of interactional communicative moves in the entire corpus, and we will not include DAMSL tags in the discussion of subsequent examples to improve the readability of the transcripts.

2. A: um (0.94) well from the b- (0.09) okay the bottom of the onion↑ [in re]lation to (0.66)
   B: [mmhm]
   A: the bottom of the lime I think it's probably like maybe a (1.42) like mm (0.34) a little less than a centimeter↓ **(0.63)**
3. B: ok[ay]
4. A: [yeah] (0.88) it's like basically if you look at the lime and the onion↑ **(0.2)**
5. B: mmhm↑ **(0.15)**
6. A: it would be (0.44) like centered to → **(0.93)**
7. B: the onio[n]↑
8. A: [y]eah (0.27) yeah like the lime would be centered to the o[nion]↓
9. B: [okay] (1.48) so the fifth like squiggle↑ **(0.22)**
10. A: mm[hm]
11 B: [from] the right is totally covered↓ **(0.06)**
12. A: yeah↓ **(0.05)**
13. B: ok[ay]
14. A: [totally] covered but then you'll still have that green space between the fourth line and the fifth line↓ **(0.14)**
15. B: okay

The excerpt starts with a question from speaker B who plays *Placer* in this task. The response from speaker A in line 2 starts with a turn-initial CF that aligns closely with the end of B's turn (0.08 s). The presence of *um* in this position is very common in our corpus; in fact, one third of all CFs in the OBJECTS games occur in turn-initial position, and 14% of all turns begin with a CF. These observations provide support for the pre-start function of CFs (Sacks et al., 1974), as markers of discourse and prosodic boundaries (Swerts, 1998), and as hesitation markers associated with cognitive load and the presence of choice (Stewart and Corley, 2008). What is less common, however, and is symptomatic for both speakers, is the tight temporal alignment of the CF with the end of the preceding turn. Since the CF in line 2 follows a question, there is no need for A to hurry to take the floor, since speaker B has explicitly yielded the floor and selected speaker A to continue. There is also no need in this context to employ an explicit attention getting device, another common function of turn-initial CFs. This is because speaker B is presumably fully attending to A, expecting an answer to her question. A turn-initial CF latched to the end of the preceding turn may also sometimes signal the lack of agreement with the proposition expressed in this preceding turn, but this is not the case in our example. Finally, if the CF signaled planning difficulty, it would probably be preceded by a relatively long silent pause representing cognitive processing, and not temporarily aligned almost perfectly with the end of the preceding turn. In consequence, we propose that, in addition to the above mentioned functions, CFs may participate in floor-management by signaling to the interlocutor that her utterance was understood, that no more additional information is needed, and that the speaker may need some time for planning her response. Hence, despite the hesitation nature of these CFs, their short latency signals understanding and contributes to the common ground establishment.

In this sense, some uses of turn-initial CFs have floor-management and grounding function similar to that of affirmative cue words such as *okay* or *mm-hm*, and their timing plays a role in the pragmatics of the spoken interaction. The following excerpts in (4) from Task 8 support these points and describe several uses of CFs; this time speaker A plays *Placer* and B is *Describer*.

(4) Excerpts from task 8 of the target session

1. B: okay the yellow lion is blinking↑ **(0.07)**
2. A: mmhm↑ **(0.08)**
3. B: and the yellow lion is directly on top of the owl↑ **(0.4)**
4. A: okay↑ **(1.12)**
5. B: [**um**]
6. A: [so] his feet are [or]
7. B: [the] whole thing they're like kind of like it's like sitting like right on the owl↑ [like the]
   A: [okay]
   B: owl's (0.08) hidden in my picture↑ **(0.27)**
8. A: so you can't see his face↑ **(0.28)**
9. B: **um** you can see his ey[es]↑
10. A: [o]kay

- - - - - - - - - - - - - - - - - - - -

11. B: the b- (0.5) lion's directly on top of him↑ **(0.06)**
12. A: okay↑ **(0.88)**
13. B: **um** (0.45) so it looks like the owl's kinda like peeking out from behind the lion↑ **(0.21)**
14. A: okay ↑ **(0.62)**
15. B: [**and**]
16. A: [so the bot]tom part like where like a cows udders would be is that like over (0.74) over the eyes↑ **(1.03)**
17. B: [**mm**]
18. A: [like over the] head a little bit and then just the eyes are showing↑ **(0.44)**
19. B: you mean like if the lion had udders↑

- - - - - - - - - - - - - - - - - - - -

20. B: but you can see the eyes of the owl↑ **(0.31)**
21. A: how about like the nose and stuff↓ **(0.08)**
22. B: [**um**]
23. A: [or whatever that] is → **(0.09)**
24. B: no you can't see his beak→

- - - - - - - - - - - - - - - - - - - -

25. A: okay (0.36) how about the little (0.33) black part (0.58) um (0.48) where the beak starts (0.66) do you see [that]→
26. B: [**um**] (0.17) it's like blinking in and out let me see (0.89) um yeah there's like black above the beak righ[t]↑
27. A: [o]kay [just a little bit of that]
28. B: [yeah you can see that]
29. A: okay and um (0.58) anything el[se]↑
30. B: [**um**] let me think (1.41) mm (1.04) see → (**2.00**)
31. A: is the tail sticking out from th- b- where the branch is like it's not aligned↑ **(1.18)**
32. B: [**um** yeah it's]not [aligned with the] branch
33. A: [the tail of the lion] [okay]
34. A: and either is the foot like it's ?- [sticking] out a little bit more↑ **(0.25)**
35. B: [the feet]
36. B: **um** (2.11) oh the branch on the left side↑

Ten turns begin with a CF in these four short sections of a single task. All of them come from speaker B: there are 8 *ums*, 1 *mm*, and one prolonged *and* functioning as a hesitation filler. These turn-initial CFs fall into two broad categories. Either they are preceded by a short or negative latency (i.e. overlap), or they follow a relatively long silence. We argue that the first type, exemplified in lines 9, 22, 26, 30 and 36, functions as a floor-keeping device used to signal to the interlocutor to wait. All of them follow questions from speaker A. The second use of CFs is exemplified in lines 5, 13, 15, 17 and 32 where they function as typical hesitation markers and occur after significant silences. Note how the use of CFs in their default meaning as hesitation markers results in frequent overlaps from speaker A (in lines 6, 16, 18, and 33). We suggest that the observed bi-modal distribution of latencies for turn-initial CFs represents another aspect of speaker B accommodating to the temporal pattern of turn-taking imposed by speaker A. In other words, speaker B uses CFs to acknowledge and secure the floor, and, crucially, uses their tight temporal alignment to signal this pragmatic meaning. This strategy of speaker B is employed as a local adjustment in temporal turn-sequencing, following her apparent realization that her default use of fillers as hesitation markers following significant silences results in frequent overlaps. This analysis provides additional support for our proposal that turn latencies are interactionally negotiated between the interlocutors and, in relation to one of our research questions, shows that timing of turn-initial CFs is systematic and plays an important role in the process of common ground establishment.

Consider again the latency values in these turn-initial CFs. All CFs from the first group are tightly aligned and have latencies shorter than 0.3 s (including two cases of negative latencies represented as overlaps). All of the loosely aligned CFs in the second group have latencies longer than 0.6 s. Therefore, the interlocutors display a contrast in the pragmatic meaning of turn-initial CFs. This pragmatic contrast is realized as a bi-modal distribution of prosodic temporal turn-initial latencies with a boundary differentiating the two modes somewhere between 0.3 and 0.6 s. In this sense, interactions in Excerpt (4) support the observations from the initial task, analyzed in section 3.1 above, about the dynamically evolving non-linear relationship between the pragmatic meaning and prosodic realization.

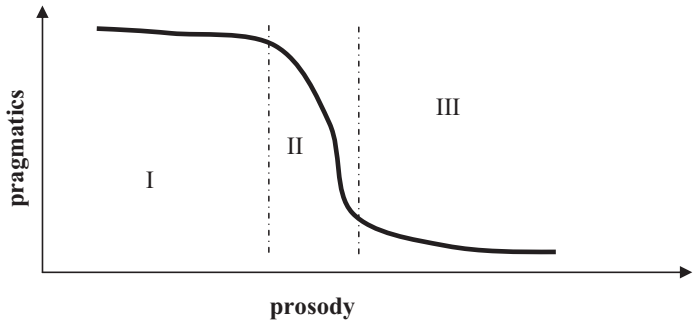

Fig. 5. Non-linearity between prosody and pragmatics.

A general form of a non-linear relationship between prosody and pragmatics is illustrated in Fig. 5. A two-dimensional prosody-pragmatics space is carved into three regions. Regions I and III represent stable regions where changes in the prosodic parameter have little effect on the pragmatic parameter. In the unstable region II, however, even small changes in prosody have a great effect on pragmatics. Similar non-linear relationships are abundant in speech (e.g. categorical perception, Liberman et al., 1957, or the alignment of F0 peaks with respect to stressed syllables for cuing contrasts between narrow and broad focus, between questions and statements, or between different dialects, Smiljanic and Hualde, 2000; D'Imperio and House, 1997; Atterer and Ladd, 2004). In our case, the prosodic parameter may represent turn latency and the pragmatic one common ground. Regions I and III thus represent two stable degrees of common ground understanding, with shorter latencies signaling more understanding and longer latencies signaling less understanding. The prosodic interval in region II represents a critical boundary for these two macroscopically stable pragmatic meanings.

We conclude that the pragmatic accommodation between the interlocutors is enabled partly through the dynamically evolved non-linear relationship between the temporal alignment of both SWGUs and CFs and the establishment of mutually shared knowledge.

3.3. Timing of SWGR/CFs and dominance

In addition to common ground establishment, the second major hypothesis presented in section 1.5 was that the timing of SWGRs participates in constructing an asymmetrical power relationship between the interlocutors. In this section, we examine this hypothesis qualitatively.

We noticed that when speaker B plays the objectively less dominant Placer role, she typically assumes the passive role of simply acknowledging the information coming from her interlocutor. Speaker A in this position, however, frequently asks questions or uses final rising intonation to produce response-eliciting statements from speaker B. Several such examples can be observed in the excerpts with CFs in (4) above. In these four excerpts, at least six turns from speaker A effectively establish her control of the floor (lines 6, 8, 16, 21, 25, 29). Assuming that, if a speaker produces a question or a response-eliciting utterance, this speaker is holding the floor and controlling the flow of the conversation, speaker A thus seems to be a more dominant speaker.[14] We test the validity of this analysis for the entire conversation quantitatively in section 4.

We next return to the qualitative analysis of the temporal patterns of affirmative cue words in transcript (3). The difference between B's latencies in lines 3 and 5 provides another example of the pattern described in section 3.1. A slightly delayed response from B in line 3 (0.63 s), results in an overlap from speaker A, followed by a more tightly aligned backchannel in line 5 (0.2 s). This time, however, the overlap in lines 3 and 4 does not seem to be connected to the relationship between the timing of turn-initiation and common ground, as we have argued so far. Speaker A's turn in line 2 is produced in significantly slower tempo, two silent pauses, an overlong CF *mm* (1.38 s), and a slowed production of *centimeter*. Speaker B accommodates by producing her backchannel with a relatively longer latency. However, the following turn-initial *yeah* from speaker A seems to reset the tight temporal alignment pattern, to which speaker B again accommodates with her backchannel in line 5. This example thus supports the utilization of the timing in SWGR/CFs for the negotiation of the power relationship, and more specifically, the dominance of speaker A and the accommodation of speaker B.

Another pattern of temporal alignment in SWGRs discussed in section 1 which we assume to be linked to the power relationship is the degree of incorporation of the current utterance into the metrical patterns of the interlocutor's utterance. We illustrate this incorporation below with examples from adjacency triplets. In this unit of interaction, the first speaker provides some information or poses a question; the second speaker acknowledges, backchannels, or provides a short answer; and then the first speaker acknowledges this response. Lines 2–4 of Excerpt (3) exemplify this

[14] This approach to dominance corresponds to the notion of sequential dominance (Itakura and Tsui, 2004).

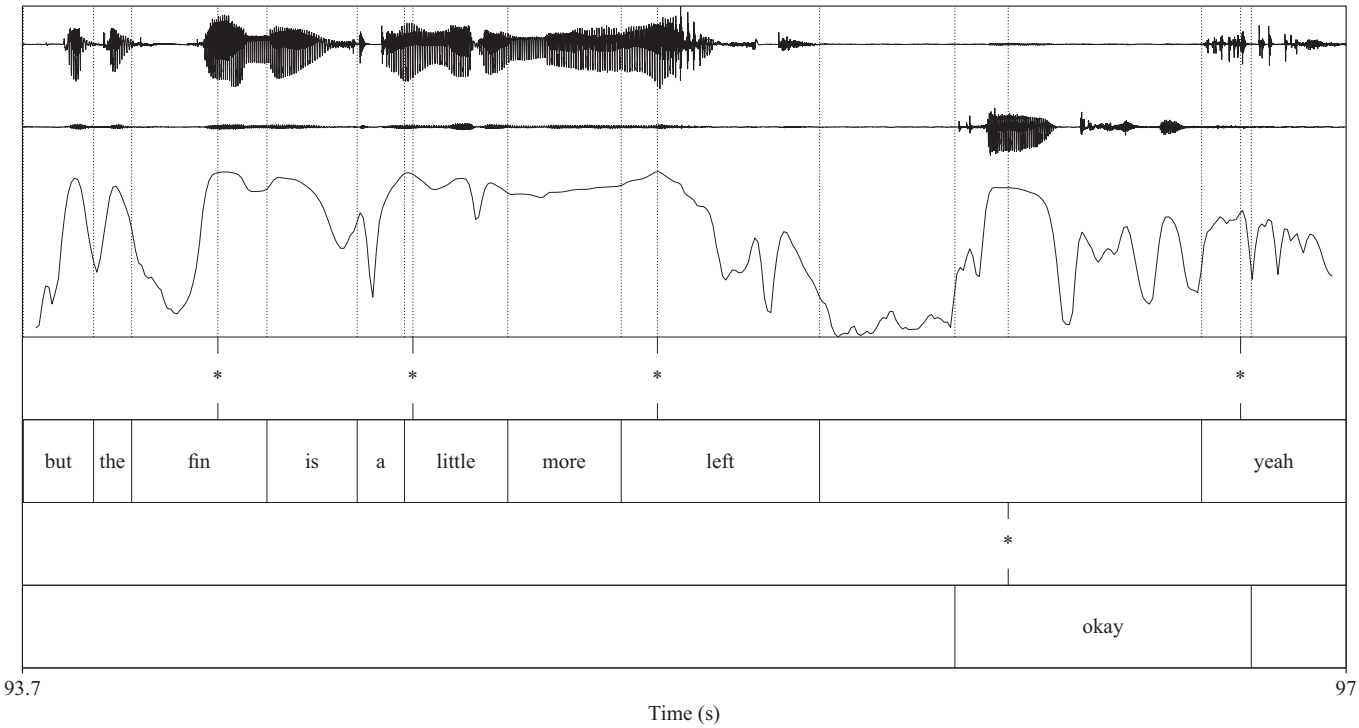

Fig. 6. Adjustment in the rhythmical alignment in consecutive adjacency triplets. The top panel shows the sound waves, the middle panel intensity in dB, and the bottom four panels the transcript with pitch accents labeled as "*".

situation when an adjacency triplet is initiated by speaker A, and lines 11–13 display another adjacency triplet, this time initiated by speaker B.

Adjacency triplets also present suitable material for comparing relative and absolute time in describing the temporal alignment of SWGRs. Following the discussion in sections 1 and 2, we relativize time by studying the timing of peaks of prosodic prominence in relation to such peaks in the preceding utterance. In other words, we establish a 'beat' of a turn-final utterance and examine how the first peak of prosodic prominence after the turn exchange fits into this beat. Beat roughly corresponds to the temporal interval between prominent syllables associated with a pitch accent. These P-centers (Couper-Kuhlen, 1993) are assumed to be linked most tightly to the amplitude (loudness) of the syllables (Cummins and Port, 1998). In our corpus, we define P-centers as the amplitude peaks of the stressed syllables in all words that receive a pitch accent mark in the labeling of the prosodic structure using the ToBI scheme (Beckman et al., 2004). Two adjacency triplets from the closing section of excerpt (2) are repeated here for convenience in Excerpt (5). Figs. 6 and 7 show the relevant portions of the exchange in (5). Fig. 6 illustrates the triplet in lines 1–3, and Fig. 7 in lines 3–5.

(5) Concluding section of Task 1

1. A: yeah it's just just about but the fin is a little more left↓ **(0.34)**
2. B: oka[y]↓
3. A: [y]eah↓ (2.02) so it's sort of like the same situation as (0.14) how the bottom black part is (0.44) almost aligned to the white feet of the M&M guy↑ **(0.04)**
4. B: oka[y]↓
5. A: [ye]ah↓

In the first triplet, the semi-isochronous pattern established by the metrical distribution of the pitch accents on A's *fin*, *little*, and *left* is followed by a slightly delayed peak in B's *okay*. This *okay* should have fallen in between A's peaks on *left* and *yeah* in order to be perfectly rhythmically aligned. Looking at the second triplet we see that speaker B makes adjustments and that her second *okay* is perfectly rhythmically incorporated into the pattern initiated by A's pitch accents. Hence, despite the fact that both triplets end in overlap by *yeah* from speaker A, the first *okay* from B is rhythmically slightly mis-aligned and B tries to correct for this with her second *okay*.

The adjacency triplets show how the timing of SWGRs functions in negotiating the floor-control dominance through the patterns of accommodation of speaker B and its lack for speaker A. Despite the effort from B to accommodate her SWGR *okays* to A's metrical pattern, speaker A still produces her SWGR *yeahs* with an overlap. Fig. 8 shows another example of this pattern, this time without overlaps. The figure shows two consecutive triplets from Task 7 in which A is the Placer and B the Describer. We see that the spacing of the pitch accents in the first question is greater than in the second question, to which B

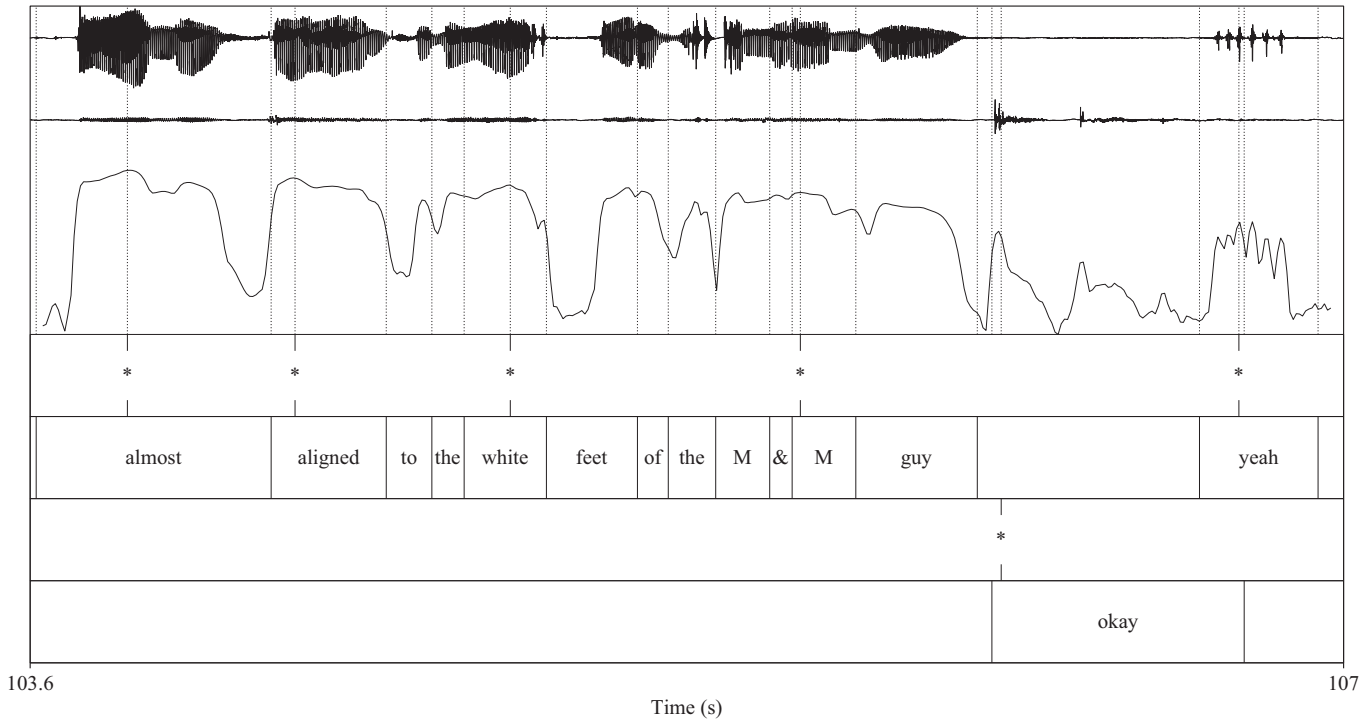

Fig. 7. Adjustment in the rhythmical alignment in consecutive adjacency triplets. The top panel shows the sound waves, the middle panel intensity in dB, and the bottom four panels the transcript with pitch accents labeled as "*".

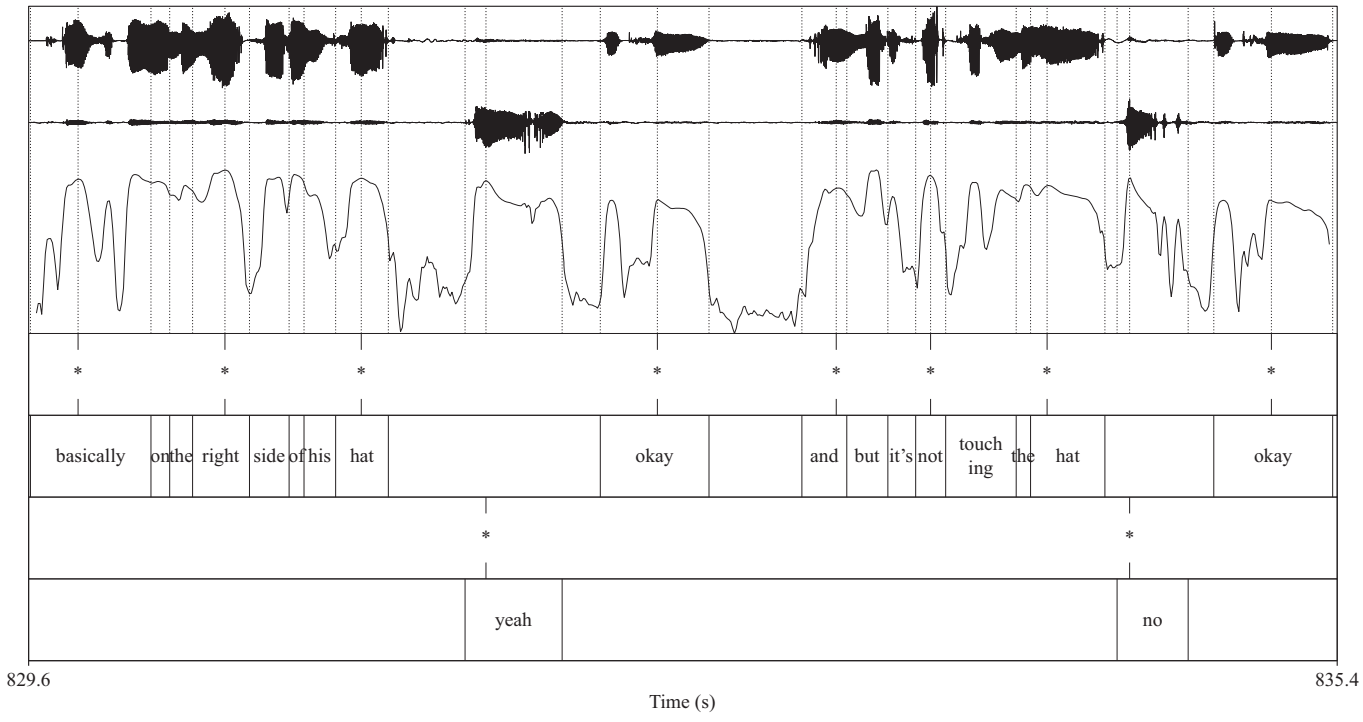

Fig. 8. Adjustment in the rhythmical alignment of two consecutive adjacency triplets. The top panel shows the sound waves, the middle panel intensity in dB, and the bottom four panels the transcript with pitch accents labeled as "*".

adjusts by aligning the P-center of the second response more tightly than the first response. The qualitative observations based on the visual inspection of these cases will be tested quantitatively in the next section.

In this section we have argued that speaker A is a more dominant speaker, and that the variation in the temporal initiations of SWGR/CFs is one of the means for establishing this dominance. This proposal makes the prediction that speaker B adjusts her turn-taking behavior more than speaker A. In the next section, we test the validity of this prediction quantitatively in several ways.

Table 2
Distribution of major turn types in the target conversation; for column labels see section 2.2.

			Turn label						Total
			S	O	BC	X2	PI	I	
Speaker	A	Count	**121**	**69**	**47**	**52**	**11**	**12**	**312**
		%	38.8%	22.1%	15.1%	16.7%	3.5%	3.8%	100.0%
		Adj. Res.	−1.6	2.1	−1.1	1.7	−1.3	.3	
	B	Count	**144**	**50**	**58**	**38**	**18**	**11**	**319**
		%	45.1%	15.7%	18.2%	11.9%	5.6%	3.4%	100.0%
		Adj. Res.	1.6	−2.1	1.1	−1.7	1.3	−.3	
Total		Count	**265**	**119**	**105**	**90**	**29**	**23**	**631**
		%	42.0%	18.9%	16.6%	14.3%	4.6%	3.6%	100.0%

4. Quantitative analysis of the target conversation

The target conversation contains 633 IPUs initiating an exchange of turns that are almost equally distributed between the two speakers (313 for A and 320 for B). We begin with quantitative testing of the first observation presented in section 3.3 that speaker A controls the floor more than speaker B by eliciting a response more frequently. Our corpus contains annotations for questions and utterances eliciting a response from the interlocutor. A Pearson chi-square test shows that speaker A elicits response significantly more often than speaker B; $\chi^2(1, 633) = 4.58$, $p = 0.032$. Hence, this result supports the analysis that speaker A controls the floor more than speaker B.[15]

We continue with an analysis of turn type distributions to assess overlaps and interruptions as possible indicators of dominance. In terms of general turn-taking behavior, Table 2 shows the frequencies of major turn types; in this table backchannels with and without overlap (BC_O and BC_NO, respectively), as well as the continuation after a backchannel with and without overlap (X2_O and X2_NO respectively), were merged into BC and X2 respectively, two instances of a turn labeled as "?" from each speaker were omitted, and no butting-ins (BI) occurred.

A Pearson chi-square test shows that the difference between the speakers in their turn-taking behavior, represented by the distribution of the major turn types, is significant at a 90% confidence level; $\chi^2(5, N = 631) = 10.02$, $p = 0.075$. The analysis of the percentages and adjusted residuals shown in Table 2 reveals that the speakers differ significantly only in the propensity for overlaps: speaker A overlaps more often than speaker B, $p < 0.05$. A significant difference between the speakers at a less conservative level of $p < 0.1$ is observed for smooth switches (S) and continuations after a backchannel (X2).[16] Additionally, we can construe our labeling scheme as identifying three main turn-taking behaviors: fluent floor changes (S, O), disfluent floor changes (I, PI), and turns without intentional floor change (BC, X2). In fluent floor changes, speaker A used more overlaps than smooth switches when compared to speaker B, and this difference is significant also in a separate 2-by-2 chi-square analysis; $\chi^2(1, N = 384) = 4.51$, $p = 0.034$.

Looking at the distribution of other turn types, speakers do not differ significantly. Their propensity for standard interruption (I) is virtually identical.[17] Given the qualitative observation of A hurrying B's SWGRs, Table 3 shows the distribution of overlaps in backchanneling. Speaker A overlaps preceding backchannels significantly more often than speaker B (X2_O).[18] This result thus supports the qualitative analysis of A as hurrying B's grounding responses presented in section 3.

To summarize the quantitative analysis of turn-type distributions, speaker A overlaps more often than speaker B, and, within the backchannel turns, A overlaps preceding backchannels more often than B (X2_O). These quantitative results are in line with our qualitative analysis of speaker A as using overlapped turn-initiations for hurrying speaker B. This analysis by itself is not conclusive since overlapping is not necessarily an index of dominance or floor control, as noted in section 1.3. Moreover, the two speakers show similar behavior in more traditional measures of dominance in turn-taking such as the frequency of standard interruptions or the total number of turns. However, taken together with the qualitative analyses of

[15] Despite the equal distribution of the game roles (Describer/Placer) in the tasks (7 tasks with speaker A as Describer and 7 with her as a Placer), the distribution of turns in these roles was slightly skewed. In the tests run separately for the game roles, speaker A elicited responses more often than speaker B when A was in the Placer role; $\chi^2(1, 338) = 6.1$, $p = 0.01$, and no difference between the speakers was observed when A was in the Describer role; $\chi^2(1, 295) = 0.227$, $p = 0.63$. A comparison among differences of proportions (Blalock, 1979) in the distributions for the two roles showed a significant difference; $z = 2.02$, $p < 0.05$. Hence, speaker A controlled the floor more than speaker B when speaker A played the Placer.

[16] Adjusted residuals are assumed to have a normal distribution with mean of zero and standard deviation of one. Thus, adjusted residuals with an absolute value greater than 2.0 allow us to conclude that the observed frequency count for that cell is significantly different from the expected value, had there been no association between the two variables in question at $p < 0.05$. Residual values greater than 1.6 allow similar conclusion at $p < 0.1$.

[17] Speaker B produces slightly more pause interruptions (5.6% vs. 3.5%) and backchannels (18.2% vs. 15.1%) than speaker A. The qualitative direction in these two non-significant patterns is consistent with greater empathy of speaker B and her accommodation to speaker A's speech. This is because PIs are commonly used to finish interlocutor's sentences or provide help when the interlocutor seems to be in trouble, and can thus be considered as collaborative rather than competitive. Determining if a pause interruption is intended as collaborative or disruptive is a complex and subjective task. Nevertheless, we checked all 29 instances of PI in this conversation and attempted to classify them informally into these two categories. 12 out of 18 PIs from speaker B were perceived as collaborative while only 2 out of 11 PIs from speaker A were so perceived. The observed difference between the speakers in the use of pause interruption turns is thus in line with the hypothesized more dominant behavior of speaker A.

[18] Note, however, that the low count in X2_O cell for speaker B means that the reported significance in the adjusted residual should be taken with caution.

Table 3
Distribution of overlaps in backchanneling.

	Backchannel label						Total
			BC_NO	BC_O	X2_NO	X2_O	
Speaker	A	Count	**36**	**11**	**45**	**7**	**99**
		%	36.4%	11.1%	45.5%	7.1%	100.0%
		Adj. Res.	−.9	−1.3	1.0	2.1	
	B	Count	**41**	**17**	**37**	**1**	**96**
		%	42.7%	17.7%	38.5%	1.0%	100.0%
		Adj. Res.	.9	1.3	−1.0	−2.1	
Total		Count	**77**	**28**	**82**	**8**	**195**
		%	39.5%	14.4%	42.1%	4.1%	100.0%

the types of overlaps presented in previous sections, the data in Tables 2 and 3 complement the picture of speaker A as being more dominant than speaker B.

So far, we have quantitatively analyzed the patterns of temporal alignment using discrete labels that characterize turn-taking behavior. We now test the potential of continuous features related to turn latency for better understanding of the relationship between the temporal alignment and pragmatic accommodation. Recall that latency, measured in seconds, was the main means of describing temporal alignment in section 3. In the target session, speaker A produces her turns with significantly shorter latencies than speaker B; $F(1, 631) = 4.3$, $p = 0.039$ with mean latencies of 0.35 and 0.23 s, respectively.

Raw latencies describing absolute timing of turn initiations provide only partial understanding of the patterns in the temporal alignment of SWGR/CFs. To complement this understanding, we investigate the variability of latencies within a well-defined temporal window in an effort to test whether speaker A hurries her interlocutor more than vice versa. Consider the pattern of turn exchanges between two speakers (Spkr1 and Spkr2) illustrated in Fig. 9, in which turn1 does not overlap turn0, turn1 and turn3 begin with a SWGR/CF, and turn2 overlaps turn1. According to our hypothesis, the overlap between turn1 and turn2 may have an effect on the duration of latency3, depending on the power relationship between the speakers. If Spkr1 is dominant, then Spkr2 will typically rush her next answer, effectively making latency3 smaller than latency1. If Spkr2 is dominant, then we should observe no effect on latency3.

The pattern of SWGR/CF timing defined in this way is summarized in Table 4. In the case when Spkr1 represents Speaker A of our target conversation, Fisher's Exact Test shows that the distribution departs significantly from random; $p = 0.02$. In the second case, when Spkr1 corresponds to Speaker B, the observed frequencies do not depart significantly from random; $p = 0.295$. In other words, when speaker A's turn2 overlaps B's turn1, this overlap has a significant effect on the timing of B's turn3; however, this effect does not seem to occur when the roles are reversed. These results, quantitatively testing the local adjustments in the timing of SWGR/CFs over the entire conversation, add support to our analysis concerning the effect of A's dominance over B on the timing of conversational exchanges.

Another timing measure we hypothesize to be linked to dominance is the accommodation of turn initiations to the rhythm of the preceding utterance: greater accommodation positively correlates with lower dominance. We start by testing the correlation between the rate of pitch accents in the last IPU before the turn-exchange (a rough measure of speech rhythm) and the latency between the last pitch accent in this IPU and the first pitch accent in the IPU immediately following a turn-exchange. In all available data from the target session, this correlation is significant and positive but very moderate; $r(588) = 0.14$, $p = 0.001$. Testing for the differences between speakers, we observe that the overall effect is due only to speaker B. This is because the correlation for speaker B improves compared to the value from the pooled data; $r(293) = 0.22$, $p < 0.001$, while the significance disappears for speaker A; $r(295) = 0.02$, $p = 0.71$. Additional improvement for speaker B is obtained when testing only for the SWGRs (labels Agr and BC, respectively in Table 1 in section 2.2). Here, the correlation for speaker B remains significant and approaches moderate values, $r(127) = 0.3$, $p = 0.001$, while speaker A's correlation remains not-significant; $r(130) = 0.01$, $p = 0.9$. It is possible that this effect is an artifact of the roles the speakers played. However, the same correlation tests run separately for turns initiated by the speaker in the Placer and Describer roles produce similar results as those reported above. Hence, the quantitative difference between the speakers in the correlation between the time of turn-initiation and the rate of pitch accents in the preceding utterance shows that speaker B accommodates more than speaker A, which supports the qualitative observation about the asymmetry between the speakers in floor-control dominance.

The second measure that describes the accommodation of turn initiations to the metrical patterns of the interlocutor's preceding utterance, as discussed in section 1.5, is the entrainment index (EI). The histograms in Fig. 10 show the distribution of EI values for SWGRs divided by speaker and role.

Although small bin counts prevent conclusive quantitative analysis of these distributions, the plots do show clear differences between the two speakers. The values for speaker B (2 right panels) tend to cluster around 1.3 in both roles, and there is an additional peak/discontinuity around 2.3, that is roughly 1 unit away from the first cluster, when speaker B plays the Placer role (top right panel). Values around or below zero, effectively corresponding to significant overlaps, are extremely rare for this speaker. Hence, the histograms show some degree of rhythmical entrainment in turn-initial SWGRs for speaker B due to clustering of EI values.

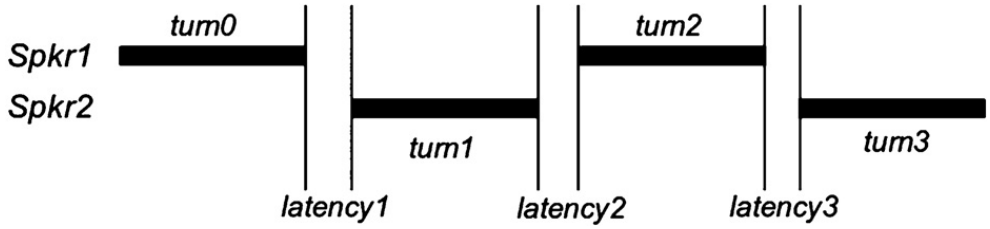

Fig. 9. Schematic representation of turn-exchanges.

Table 4
Number of cases when a speaker rushes her SWGR/CF as a function of having her SWGR/CF rushed in the preceding turn (see Fig. 9).

Spkr1 = speaker A, Spkr2 = speaker B		Did Sp B rush her turn3 (latency3 < latency1)?	
		Yes	No
Did A in turn2 overlap B in turn1 (latency2 < 0)?	Yes	17	5
	No	16	21
Spkr1 = speaker B, Spkr2 = speaker A		Did Sp A rush her turn3 (latency3 < latency1)?	
		Yes	No
Did B in turn2 overlap A in turn1 (latency2 < 0)?	Yes	14	3
	No	17	11

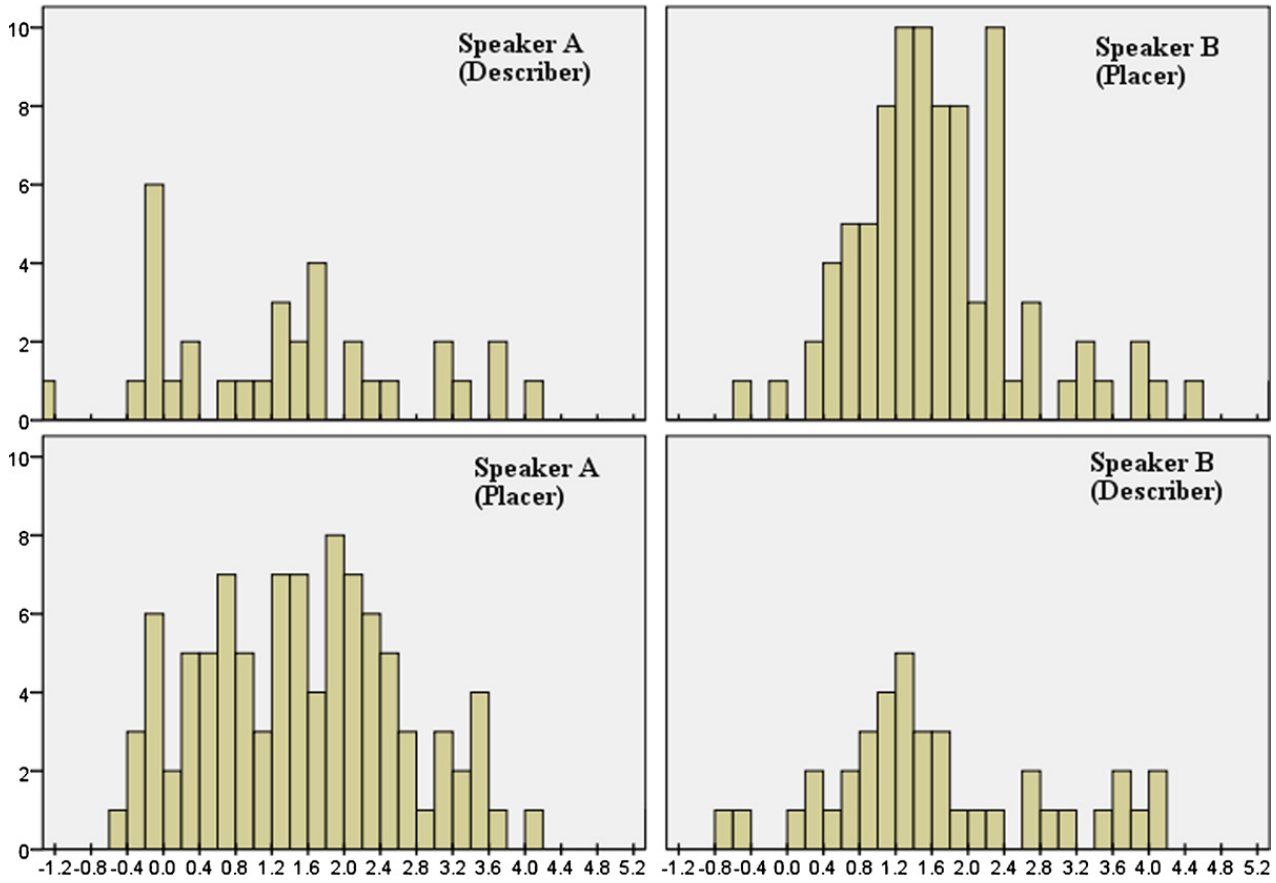

Fig. 10. Distributions of entrainment index (latency/rate) of single-word grounding responses separately for two speakers (columns) and game roles (rows).

Speaker A, on the other hand, does not show this pattern. Her values in both roles are less clustered and much more evenly spread than speaker B's values, and the difference between adjacent discontinuities seems to be around 0.5. Additionally, A's histograms from both roles show a clear peak around or below zero, corresponding to frequent overlaps. We conclude that the relationship between the rate of pitch accents before the turn-exchange and latency after the the exchange support the view that the turn-initial SWGRs from speaker A are less entrained to the rhythm of the preceding utterances than those from speaker B.

Table 5
Distribution of turn-initial IPUs in the target conversation depending on whether they start with a conversational filler or not.

			Turn starts with a CF?		Total
			No	Yes	
Speaker	A	Count	**286**	**27**	**313**
		%	91.4%	8.6%	100.0%
		Adj. Res.	2.4	−2.4	
	B	Count	**273**	**47**	**320**
		%	85.3%	14.7%	100.0%
		Adj. Res.	−2.4	2.4	
Total		Count	**559**	**74**	**633**
		%	88.3%	11.7%	100.0%

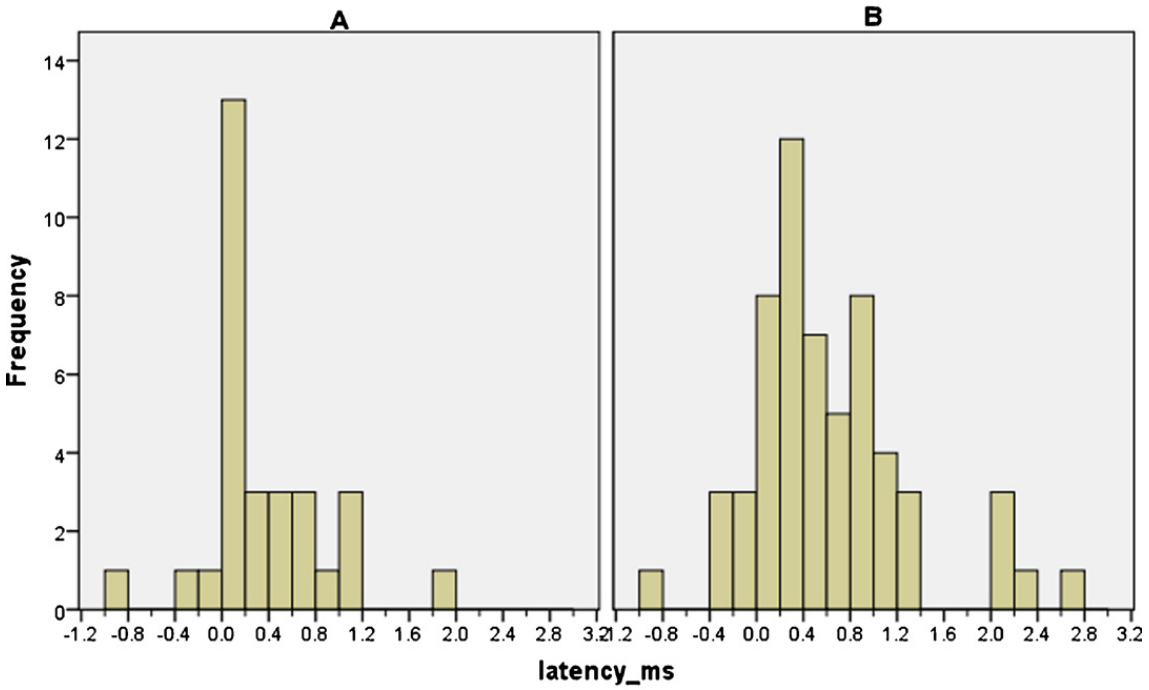

Fig. 11. Distributions of raw latencies for turn-initial conversational fillers by speaker.

After examining SWGRs, we now turn to CFs. In the qualitative observations we identified two prominent functions of turn-initial CFs related to rhythmical entrainment, and speaker B as more inclined to use them for floor-management purposes. Here we test these qualitative observations with a quantitative analysis based on the data from the entire session. First, Table 5 shows the distribution of IPUs immediately following a turn-exchange for the two speakers. We see that 12% of these IPUs begin with a CF and that two thirds of these are produced by speaker B. This difference between the speakers is significant; $\chi^2(1, 633) = 4.6$, $p = 0.032$. Additionally, speaker B produces these turn-initial CFs with greater mean pitch and tends to produce them also with greater mean intensity than speaker A; $F(1, 66) = 4.59$, $p = 0.036$ and $F(1, 66) = 3.14$, $p = 0.081$, respectively.[19] This difference signals the greater pragmatic importance of turn-initial CFs for speaker B than for speaker A.

Finally, Fig. 11 shows the distribution of raw latencies for the turn-initial CFs for the two speakers. We look at the latencies between the end of turn-final IPUs and the start of the turn-initial CF, and not the latencies between the last pitch accents and the pitch accent of the CF. We take this approach because the plateau intonation pattern, with which CFs tend to be produced, does not yield reliable information about the presence, and more importantly, about the temporal alignment, of pitch accents. Again, we see that the two speakers differ in their timing of turn-initial CFs. Speaker A has a very clear peak around 0.1, which are turn initiations with almost perfect latches. Recall that the pattern of aligning turn-initial CFs with a short overlap, or shortly after the preceding turn ends, was identified in sections 3.2 and 3.3 as a local temporal adjustment strategy of speaker B. We argued that B uses this strategy in an effort to secure the floor for herself as a reaction to speaker A's

[19] Both mean pitch and mean intensity were *z*-score normalized by speaker, which allows for straightforward comparison between the speakers despite potential physiological differences among the speakers or differences in experimental conditions.

tactics to take the floor when B does not start her turn after a short latency. Here we see that this pattern of close temporal alignment for turn-initial CFs is preferred for speaker A as well, and might serve as an additional source of accommodation for speaker B. The histogram for speaker B has a wider distribution, and thus, greater variation, with the local adjustments analyzed qualitatively above as a potential source of this variation. Moreover, in addition to the peak around 0.3 s, the histogram for speaker B displays additional minor discontinuity around 1 s. Such discontinuity, although statistically only suggestive, is in line with our qualitative analysis of speaker B's turn-initial CFs in section 3.2. There we argued for two pragmatically meaningful degrees of common ground understanding that are signaled by two non-overlapping intervals of latency values. The histograms, together with the qualitative observations, thus show that the temporal alignment of CFs is yet another area where speaker B accommodates more to speaker A than vice versa.

To summarize the quantitative observations from the target session, the analysis of the turn-type distribution provides support for the qualitative characterization of speaker A as less accommodating, rhythmically hurrying her interlocutor, and more dominant, while speaker B is more accommodating. Further support for this characterization comes from investigating the relationship between the rhythm of the utterance preceding a turn-exchange and the initiation of the following turn with special focus on affirmative cue words and CFs. Here we see significant differences between the speakers in the rhythmical alignment of turn initiations as well as in the accommodation of this behavior to the interlocutor's patterns. Note that the conclusion of A's dominance is not supported if we look only at more traditional measures of asymmetrical power relationships, such as the number of turns and frequency of interruptions, which are similar for the two speakers, or the amount of speaking time, which is greater for speaker B than speaker A.

5. Games with different interlocutors

We conclude our analysis with several observations from the dialogues in which the two speakers of the target conversation (A and B) play identical games but with different conversational partners. Our aim is to find out if, and how, the patterns of turn-taking behavior, observed in the target session depend on the conversational partner. We will argue that the temporal patterns of turn-initiations of the more dominant speaker (A) are minimally affected by her interlocutor, while the turn-taking behavior of the less dominant speaker (B) varies to some extent as a function of her interlocutor's turn-taking behavior. Speaker A played her first game with another female speaker D, and speaker B played her other game with a male speaker C.

Consider first speaker C's turn-taking behavior, illustrated in the transcript of the penultimate task of the session shown in (6) below.

(6) Excerpt from task 13 of session 2, B plays the Describer (duration: 124 s)

1. B: The M&M↑ (0.32) is directly on top of the onion↑ (6.57) um (3.84) the (2.65) right hand of the M&M↑ (1.36) is kind of (0.73) a- crosses over to the onion↑ so (0.92) and the left hand is completely on top of the onio[n]↑
2. C: [mm] hm↑ **(0.91)**
3. B: um → **(1.76)**
4. C: What about the eyebrow↓ **(0.79)**
5. B: the eyebrows (0.54) um (0.32) ther- the head of the (0.24) M&M actually (0.45) covers (0.28) the o- like its u- b- (0.66) goes past the boundary of the onion↑ **(0.03)**
6. C: mmhm → **(0.48)**
7. B: um (0.21) it looks like → **(1.33)**
8. C: How much f- uh-↓ **(0.77)**
9. B: Oh y- um (0.23) see (0.47) do you see like the left shoe on the M&M↑ **(0.05)**
10. C: mm [hm]↑
11. B: [how it] looks like there s almost little shadow↑ **(0.05)**
12. C: mmhm → **(0.19)**
13. B: like line that shadow up with the (0.86) curve of the onion↓ (4.39) and (1.59) an- and there s a shadow on the right f- sh[oe]↑
14. C: [mm]hm→ **(0.32)**
15. B: that doesn't line up exactly it s a little bit to the left (0.4) like overlapping↓ (4.29) but the top↑ (0.42) the left shoe where the shadow starts↑ (0.67) like is exactly on top of onion↓ (0.92) like the the curvature of the onion lines up↓ (10.84) and it looks like the (1.25) the thumb I guess → (0.31) of the (0.32) um right hand (0.2) well (0.17) the M&M s left hand but to the right↑ (1.25) um (2.69) it kinda looks like the shadow of that lines up with (0.5) the curvature → (0.91) of the onion up there (0.32) but the thumb does → (1.16) the hand juts out over the onion↓ **(1.48)**
16. C: mm↓

The first identification of the target object to be placed by speaker B, *M&M*, ends in rising intonation and a short pause, which provides speaker C with a TRP suitable for a grounding response. After C chooses not to respond, the next IPU from speaker B ends with a clear backchannel-eliciting pattern on *onion*. The silent pause of over 6 s shows speaker B's expectation of receiving feedback. Another failure of backchannel elicitation appears after *right hand of the M&M*, still in the same turn.

In addition to failures to provide grounding responses, another discrepancy between expected and received turn initiation occurs after speaker B appears to yield the floor to speaker C. For example, speaker B yields the floor with falling intonation on *onion* and subsequent long silent pause in line 13, to which speaker C fails to respond. Two other cases occur in line 15 after *overlapping* and *lines up*. The former results in a silent pause of more than 4 s and the latter in an extreme 10 s silent pause. Hence, frequent overlong silent pauses, attributable to speaker C's failure to initiate a turn-exchange, show comparatively worse temporal alignment of turn-taking than in the target session. In contrast to the target session, in which speaker A frequently 'rushes' speaker B to align her responses in close temporal relation to the ends of A's turns, speaker C of this session does not seem to apply such pressure.

Looking at turn-internal silent pauses, a quantitative measure well suited for testing silences resulting from failed turn initiations, we find that speaker B produces significantly longer pauses than speaker C; $F(1, 387) = 11.37$, $p = 0.001$. Speaker B also produces turn-internal silent pauses significantly longer in this session than in the target session; $F(1, 432) = 10.9$, $p = 0.001$. Moreover, average turn latencies are also significantly longer in this session than in the target session; $F(1, 985) = 19.16$, $p < 0.001$, 0.54 s vs. 0.29 s, respectively. Finally, the correlation between the rate of pitch accents before the turn-exchange and the latency between the last pitch accent before the turn and the first one after the turn does not reach significance for speaker B in this session, despite the pattern of rhythmic alignment she exhibits in the target session. This might be due to the inconsistent behavior of speaker C. Hence, we conclude that the temporal aspects of speaker B's turn-taking behavior are affected by her interlocutor's pattern of turn initiations.

The use of turn-initial CFs from speaker B in this session corresponds mostly to the default hesitation pattern identified in the target session. For example, two turn-initial CFs in the preceding excerpt (in lines 3 and 7) are produced with latencies of 0.5 s or more and indicate difficulty in planning the next utterance. This analysis is supported by the response from speaker C, who in both cases perceives these as problems, attempts to take the floor, and responds with questions facilitating progress towards the task completion. However, we also observe the pattern of close alignment of turn-initial CFs with the end of preceding turn analyzed in the preceding section as a floor-securing device. Two short excerpts exemplify both alignment patterns: the default loose alignment signaling planning problems in line 4 and the closely aligned floor securing CF in lines 2, 6, and 9.

(7) Excerpts from tasks 3 and 4 of session 2, B is the Describer

1. C: how far is the ear↑**(0.38)**
2. B: [um]↑
3. C: [from the lion]↑**(1.28)**
4. B: um → (0.43) the → (0.34) uh → (0.78) uh see the black line on the top that kinda curves arou[nd]→
5. C: [y]e[s]↓
6. B: [u]m → (0.5) that is looks like there maybe two p- (0.25) or three pixels between the lion and that↑
   -
7. B: okay↑(0.66) so the nail↑(0.35) is gonna go directly on top of the lawnmower↑(1.91) and → (1.19) if you look at the nail there's↑(0.25) a line↑(0.15) n-kind of see the nail's pointed and then there's a line than kinda cuts across → **(0.52)**
8. C: mmh[m mmhm]
9. B: [um]↑(0.29) that looks like it gets lined up with the front edge of the lawnmower↑

Finally, Table 6 compares the distribution of turn types in the two OBJECT games played by speaker B. A Pearson chi-square test shows that the difference is significant ($\chi^2(5, 495) = 11{,}81$, $p = 0.038$) and is caused mainly by a greater frequency of backchannels and a lower frequency of smooth switches in the target session compared with the other session. This is also supported by looking at the frequencies of turn-initial grounding responses for speaker B: there are significantly more backchannels and fewer agreements/acknowledgements in the target session than in the other session; $\chi^2(1, 206) = 7.27$, $p = 0.007$.

Summing up the turn-taking behavior in the two sessions in which speaker B participated, we see that: (1) The temporal alignment strategies analyzed in section 3.3 as signaling the dominance relationship between the speakers seem to reflect the turn-taking behavior of the interlocutor. This is because speaker B uses them frequently both with SWGRs and CFs in the target conversation but only occasionally in her other conversation. (2) Conversely, the strategies relating to the establishment of common ground such as B's default use of turn-initial CFs are present irrespective of the conversational partner.

Table 6
Distribution of turn types in the OBJECT games by speaker B in her first session (B&C) and in the target session (B&A); BC_O is collapsed with BC_NO and X2_O with X2_NO, 3 cases of "?" omitted.

Session		Turn-type						Total
		BC	I	O	PI	S	X2	
B&C	Count	**16**	**6**	**31**	**7**	**101**	**15**	**176**
	%	9.1%	3.4%	17.6%	4.0%	57.4%	8.5%	100.0%
	Adj. Res.	−2.7	.0	.6	−.8	2.6	−1.2	
B&A	Count	**58**	**11**	**50**	**18**	**144**	**38**	**319**
	%	18.2%	3.4%	15.7%	5.6%	45.1%	11.9%	100.0%
	Adj. Res.	2.7	.0	−.6	.8	−2.6	1.2	
Total	Count	**74**	**17**	**81**	**25**	**245**	**53**	**495**
	%	14.9%	3.4%	16.4%	5.1%	49.5%	10.7%	100.0%

Speaker A played her other game with another female, speaker D. The dynamics of this conversation is again quite different from the target session. However, while speaker C is less cooperative, passive, and often not rhythmically aligned with speaker B, speaker D is very active and tries to align her turn-initiations very tightly to speaker A's turn production. An Anova test shows that average turn latencies are significantly shorter in this session than in the target session; $F(1, 915) = 8.79$, $p = 0.003$; 0.15 s vs. 0.29 s, respectively. Consider the SWGRs in the first half of the excerpt in (8).

(8) Excerpt from task 5, session 7 (duration: 91 s)

1. A: ok um (0.42) it is situated between the mermaid↑ **(0.17)**
2. D: [uhhuh]
3. A: [the yel]low mermaid↑ [and the] whale↑ **(0.06)**
4. D: [uhhuh] alright↑ **(0.4)**
5. A: and→ (0.74) it's closer to the fin of the mermaid than it is to the whale↓ **(0.17)**
6. D: alrig[ht]
7. A: [in] terms of distance vertically↑ **(0.09)**
10. D: yeah **(0.2)**
11. A: um and → (1.0) between the yello- uh betwe- excuse me between the first and the second square it (0.39) sort of (0.28) it s aligned to the fin o- of (0.71) the (0.27) [whale]↑
12. D: [mermai-]uh the wha[le]↑
13. A: [li]ke it's her if you're like cutting off the margin on the left↑ **(0.49)**
14. D: [uhhuh]
15. A: [the tip] of the lemon↑ **(0.19)**
16. D: yeah↑ **(0.22)**
17. A: is↓ (0.5) if you put a ruler against all the fins↑ **(0.37)**
18. D: [yeah]
19. A: [like it] would be (0.4) like right there that's where it would be↓ **(0.04)**
20. D: that little point [that little nubbin of the (0.33) would be]
21. A: [that little knob yeah yeah]
22. D: on the same level as the fins of the whale↓ **(0.21)**
23. A: yeah if you were looking (0.07) to the left margin↑ **(0.06)**
24. D: and the tip an- and the (0.74) ok[ay]↑
25. A: [ye]ah and it's closer to the mermaid [there's]
26. D: [<whisper> closer to] the mermaid↓ **(0.06)**
27. A: but um↑ **(0.35)**
28. D: okay (0.25) so like the uh nubbi[n of]

29. A: [but it's not] touching th- l- mer- there's about like maybe a centimeter between the fin [and the] lemon↓ **(0.2)**
30. D: [oh]↓ oh so it s pretty clo[se]→
A: [yeah] (0.26) but it's [not that close]↓
D: [oh a centimeter] (0.23) so like [the nubbin] of the lemon
31. A: [a centimeter]
32. D: would it be sort of aligned with the → **(0.09)**
33. A: it's aligned to the [very] tip of the whale you know what I [mean like the fins] like the
34. D: [line]↑ [yeah the o-(ne)]
35. A: ends → **(0.08)**
36. D: yea[h]↓
37. A: [of] it yeah↓ **(0.06)**
38. D: and and vertically↑ (0.37)
it would it [be a]ligned to the line between the 2nd and the 3rd square↑ **(0.54)**
39. A: [hm]
it s not al- the 2nd and 3rd square↑ **(0.1)**
40. D: at th- of th[e]→
41. A: [th]e 1st and the [2nd squ]are you me[an]↑
42. D: [in be-] [the] 2nd and the 3rd↑ **(0.32)**
43. A: the 2nd and the 3rd no↑ **(2.41)**
44. D: w- I thi[nk I-]
45. A: [where] is your stuff located↓

Speaker D produces 8 SWGRs between lines 1 and 17. The first five SWGRs in lines 1–8 all occur with a latency under 0.2 s after speaker A elicits them; yet four of them result in overlap from the subsequent A's turn. Especially the overlap in line 7 is unexpected, since A starts a new discourse sub-segment but fails to signal this rhythmically. The only non-overlapped SWGRs in the first part of the excerpt are those where A seems to hesitate or has not fully planned her utterances. This happens in lines 5 (*and* + long pause), 9 (*um and* + pause), or 15 (prolonged *is* + pause). Although this excerpt is taken from the first task in which speaker D is the Placer, we observe an already tight temporal alignment of her SWGRs, and consequently, little evidence for temporal adjustments described for speaker B in the target session. Nevertheless, despite the fact that A elicits feedback prosodically, and despite the extremely tight temporal alignment with which speaker D produces her SWGRs, subsequent overlaps caused by speaker A are frequent. We analyze this pattern as the failure of speaker A to temporarily adjust her turn initiations, and thus the failure to accommodate to speaker D's production of SWGRs. Additionally, speaker A frequently interrupts her interlocutor and holds the floor despite the efforts from speaker D to claim it. This occurs in the second half of the excerpt (e.g. lines 27, 33, and 41). Hence, the turn-taking behavior of speaker A in her two sessions seems to consistently display the absence of accommodation to her interlocutor.

Table 7 shows the comparison of the OBJECT games in the two sessions played by speaker A based on the turn-type distribution. Despite the greater frequency of backchannels in the target session, the difference between A's turn-taking behavior in these two sessions is not significant overall; $\chi^2(5, 447) = 6{,}65$, $p = 0.248$.[20] This is also supported by looking at the frequencies of turn-initial grounding responses for speaker A: there is no significant difference in their distributions in the two sessions played by speaker A; $\chi^2(1, 175) = 1.\ 263$, $p = 0.2$.

Hence, the comparison of turn-type distribution for the speakers in the target session with their turn-taking behavior in their other session shows that speaker A accommodates minimally to her interlocutors and behaves rather consistently in her turn-taking strategies, while speaker B tends to adjust her turn-taking behavior when interacting with a different interlocutor.[21] This finding is in line with our analysis of speaker A as less accommodating in her turn-taking behavior than speaker B. But this time the support comes from a more coarse-grained global level of observation when we move from analyzing the behavior within a single session to comparing the speakers' behaviors across two separate sessions.

[20] In this chi-square test, contrary to the similar test in section 4, we include both BC and X2 categories because we compare the distribution of a single speaker in separate sessions and thus BC and X2 are not related. It is also interesting that in the CARD games, which have significantly less data and are communicatively less engaging, speaker A shows significant difference between her two sessions in the distribution of these variants. However, the comparison of the objects and cards games is beyond the scope of this paper.

[21] It is important to keep in mind, however, that there are other differences between the two sessions that we did not control for that could influence the result, such as familiarity with the game.

Table 7
Distribution of turn types by speaker A in her first session (A&D) and the target session (A&B) in the OBJECT; BC_O is collapsed with BC_NO and X2_O with X2_NO, 3 cases of "?" label were omitted.

Session		Turn-type						Total
		BC	I	O	PI	S	X2	
A&D	Count	10	8	34	7	56	20	135
	%	7.4%	5.9%	25.2%	5.2%	41.5%	14.8%	100.0%
	Adj. Res.	−2.2	1.0	.7	.8	.5	−.5	
A&B	Count	47	12	69	11	121	52	312
	%	15.1%	3.8%	22.1%	3.5%	38.8%	16.7%	100.0%
	Adj. Res.	2.2	−1.0	−.7	−.8	−.5	.5	
Total	Count	57	20	103	18	177	72	447
	%	12.8%	4.5%	23.0%	4.0%	39.6%	16.1%	100.0%

To summarize the investigations of the two non-target sessions, we have argued that speaker B is a more accommodating speaker than A, as seen by speaker B's adjustment of turn-type distribution in her two sessions, while no significant differences are observed for speaker A. We also saw significant differences in mean turn latencies and other general measures of temporal alignment in the vicinity of turn exchanges in the three sessions in the directions expected from the qualitative observations. The findings in this section thus support the analyses in sections 3 and 4 that were concerned with analyzing the behavior of speakers within a single session, by comparing the speakers' behaviors across two separate sessions.

6. Discussion and conclusions

There is a rich and fruitful tradition of research into the systematicities of the relationship between prosody and the semantic/pragmatic/discourse meanings of utterances; see for example Hirschberg (2003) for a review. Prosody in this line of research is mostly understood as distributions and types of intonational prominences together with phrasing of utterances into hierarchically organized prosodic units. In the last decades there have also been significant advances in our understanding of the turn-taking mechanism and especially of its prosodic correlates within the framework of Conversational Analysis. In this paper we proposed that fine aspects of the temporal and rhythmical organization of turn-initiations are also linked to the pragmatic aspects such as the power relationship, common ground understanding, and accommodation to interlocutor's speech.

Based on the analysis of task-oriented dialogues we argued that the timing of turn-initial SWGRs and CFs is a prosodic marker of an asymmetrical floor-control dominance relationship that emerges even in an originally balanced discourse. More specifically, the dominance of one speaker was linked to the way she often timed a turn initiation so that her interlocutor had insufficient time to begin or finish her turn. In contrast to this speaker, the other speaker's accommodation was linked to multiple strategies for adjusting the temporal alignment of her SWGR/CFs – as a reaction to this pressure from the first speaker. The prosody-pragmatics alignment was also demonstrated in the relationship between the time of turn initiation and mutual common ground understanding: shorter latencies signal a greater degree of common ground understanding. Finally, we tested which of the patterns are stable and robust across speakers, and which may be prone to entrainment based on conversational partner. We observed that speaker A's dominance-producing close alignment of her turn initiations was used irrespective of her interlocutor. Speaker B's use of closely aligned turn-initial CFs, that were analyzed as a floor-securing device reacting to the pressure from speaker A, were also present in both interactions of this speaker. All the conversations also displayed the correlation between short turn-latencies and mutual common ground understanding. Local adjustments in the timing of turn-initial SWGRs for speaker B were dependent on the interlocutor.

We supported these analyses by qualitative examination of multiple excerpts from our corpus. We identified core patterns by a description of symptomatic examples couched in the framework of Conversational Analysis. We validated these by the quantitative analysis of turn-taking behavior over the corpus as a whole. In this way, our study provides a comprehensive picture of the prosody-pragmatics alignment that is based on the convergence of both qualitative and several quantitative measures of relative and absolute timing of turn-initiations.

There is one apparent paradox in our results. While more traditional measures of floor control such as frequency of interruptions and distribution of turns between the speakers did not show a significant difference between the speakers, our temporal and metrical measures did. However, previous studies used these traditional measures for the analysis of floor-control dominance in conversations in which speakers clearly establish (or clearly assume) asymmetrical dominance relationships, such as those related to gender or work place. It could be expected that the type of dominance in our target conversation of persons with equal status and gender, would not be very overtly and robustly displayed in these traditional measures of floor-control dominance, which we found to be the case. However, one of the novel points of this paper is that a more subtle form of asymmetry in floor-control dominance can be observed in the conversation of equal-status interlocutors when we look at temporal and metrical patterns in turn initiations. Moreover, this type of floor-control dominance evolves

internally and dynamically within a conversation, assumes no external factors such as status or gender, and can be meaningfully approached with the methods suggested in this paper.

Our study exhibits a number of limitations. In our corpus, we do not have access to measures of dominance independent and external to the spoken interactions, such as personality tests, or different power status. Hence, our claims apply to the systematic relationship between prosody and pragmatics as they can be observed by studying spoken data. Additionally, our qualitative analyses simplify over many subtle and interesting facets of multi-layered negotiating processes of conversation in general and of turn latencies in particular.[22] Moreover, the study is limited to three conversations of two target speakers. For these reasons, the current paper should be considered as providing testable hypotheses for further large-scale research rather than as representative analysis of prosody–pragmatics relationship.

In conclusion, this study adds to the growing body of evidence about the role of lower-level embodied perception-production mechanisms in co-creating higher-level social interactions (e.g. Knoblich and Sebanz, 2008) and approaches that analyze the rhythm of speech as an affordance for entrainment among speakers (Cummins, 2009). Affordance was originally developed by Gibson (1979) and later extended by Norman (1990) to include a 'perceived possibility of action'.[23] For our purposes, affordance relates the environment – the speech of the interlocutor(s) – to the abilities of the speaker to produce speech as movement in time. The organization of turn-initiations is thus conceived not as the property of the acoustic signal, nor as a linguistic structure residing in the speaker or hearer separately. It is rather conceived as a perception-production link that has a potential to facilitate interpersonal accommodation (in terms of convergence or divergence). Hence, we propose that the rhythmical aspects of the interlocutor's speech are one of the affordances to be perceived and to be acted upon by temporally aligning turn initiations, especially in the case of turn-initial single word utterances. When temporal aspects of turn-taking are seen as affordances for accommodation, we can study not only discourse features such as 'overlap', or 'long pause', but also more fine-grained metrical features and how they relate to pragmatic aspects of spoken interpersonal interactions.

Acknowledgments

This work was supported in part by the project of European structural funds (ITMS 26240220060), Alexander von Humboldt Foundation, CONICET, PICT-2009-0026 and UBACYT 20020090300087. We are grateful to the audiences at talks at the University of Bielefeld and COST 2102 meeting in Dresden and two anonymous reviewers. All shortcomings are due only to the authors.

References

Andersen, Peter, Bowman, Lou, 1999. Positions of power: nonverbal influence in organizational communication. In: Guerrero, L.K., DeVito, J.A., Hecht, M.L. (Eds.), The Nonverbal Communication Reader, Waveland Press, Long Grove, IL, pp. 317–376.

Atterer, Michaela, Ladd, Robert D., 2004. On the phonetics and phonology of segmental anchoring of F0: evidence from German. Journal of Phonetics 32, 177–197.

Aubanel, Vincent, Nguyen, Noel, 2010. Automatic recognition of regional phonological variation in conversational interaction. Speech Communication 52 (6), 577–586.

Auer, Peter, Couper-Kuhlen, Elizabeth, Müller, Frank, 1999. Language in Time. Oxford University Press, Oxford.

Bakeman, Roger, Gottman, John M., 1997. Observing Interaction: An Introduction to Sequential Analysis, 2nd ed. Cambridge University Press, Cambridge, UK.

Beattie, Geoffrey, 1982. Turn-taking and interruption in political interviews: Margaret Thatcher and Jim Callaghan compared and contrasted. Semiotica 39 (1/2), 93–114.

Beckman, Mary E., Hirschberg, Julia, Shattuck-Hufnagel, Stefanie, 2004. The original ToBI system and the evolution of the ToBI framework. In: Jun, S.-A. (Ed.), Prosodic Typology: The Phonology of Intonation and Phrasing, Oxford University Press, Oxford, pp. 9–54.

Beňuš, Štefan, 2009. Are we 'in sync': turn-taking in collaborative dialogues. In: Proceedings of 10th Interspeech, ISCA, Bonn, pp. 2167–2170.

Benus, Stefan, Gravano, Agustín, Hirschberg, Julia, 2007. The prosody of backchannels in American English. In: Proceedings of the 16th International Conference of Phonetic Sciences, pp. 1065–1068.

Blalock, Hubert M., 1979. Social Statistics. McGraw-Hill, New York.

Brazil, David, Coulthard, Malcolm, Johns, Catherine, 1980. Discourse Intonation and Language Teaching. Longman, London.

Brennan, Susan E., Williams, Maurice, 1995. The feeling of another's knowing: prosody and conversational fillers as cues to listeners about the metacognitive states of speakers. Journal of Memory and Language 34, 383–398.

Brown, Gillian, 1983. Prosodic structure and the given/new distinction. In: Cutler, A., Ladd, D.R., Brown, G. (Eds.), Prosody: Models and Measurements, Springer-Verlag, Berlin, pp. 67–77.

Bortfeld, Heather, Leon, Silvia, Bloom, Jonathan, Schrober, Michael, Brennan, Susan, 2001. Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. Language and Speech 44 (2), 123–147.

Boersma, Paul, Weenink, David, 2005. Praat: doing phonetics by computer, http://www.praat.org.

Bull, Mathew, 1996. An analysis of between-speaker intervals. In: Cleary, J., Moll, aacute-Aliod, D. (Eds.), Proceedings of the Edinburgh Linguistic Conference, pp. 18–27.

Bull, Mathew, Aylett, Mathew, 1998. An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. In: Mannell, R.H., Robert-Ribes, J. (Eds.), Proceedings of ICSLP-98, vol. 4, Australia, Australian Speech Science and Technology Association (ASSTA), Sydney, pp. 1175–1178.

Clark, Herbert H., 1996. Using Language. Cambridge University Press, Cambridge.

[22] An anonymous reviewer correctly observes that our analyses may seem somewhat 'mechanistic'. For example, pauses at turn-exchanges might not necessarily be linked to grounding, but in some adjacency pairs, first pair parts may make subsequent pauses conditionally relevant, and an immediate onset may be noticeably immediate. However, given the mentioned little variability of conversational move types in our corpus, our detailed analysis of the target conversation shows that vast majority of pauses at turn-exchanges can be primarily analyzed as relating to grounding, and signaling alignment with the interlocutor. This, however, does not mean that these pauses do not have other functions.

[23] The notion of affordance constitutes one of the building blocks of ecological psychology, branches of cognitive science, and other disciplines.

Clark, Herbert H., Fox Tree, Jean E., 2002. Using uh and um in spontaneous speaking. Cognition 84, 73–111.
Coates, Jennifer, Sutton-Spence, Rachel, 2001. Turn-taking patterns in deaf conversation. Journal of Sociolinguistics 5, 507–529.
Cohen, Jacob, 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20 (1), 37–46.
Core, Mark G., Allen, James F., 1997. Coding dialogs with the DAMSL annotation scheme. In: In Traum, David (Ed.), Working Notes: AAAI Fall Symposium on Communicative Action in Humans and Machines, pp. 28–35.
Couper-Kuhlen, Elizabeth, 1996. The prosody of repetition: on quoting and mimicry. In: Couper-Kuhlen, E., Selting, E.M. (Eds.), Prosody in Conversation, Cambridge University Press, Cambridge, pp. 366–405.
Couper-Kuhlen, Elizabeth, 1993. English Speech Rhythm. John Benjamins, Amsterdam.
Cummins, Fred, 2009. Rhythm as an affordance for the entrainment of movement. Phonetica 66 (1–2), 15–28.
Cummins, Fred, Port, Robert, 1998. Rhythmic constraints on stress-timing in English. Journal of Phonetics 26 (2), 145–171.
Dahan, Delphine, Tannenhouse, Michael K., Chambers, Craig C., 2002. Accent and reference resolution in spoken-language comprehension. Journal of Memory and Language 47, 292–314.
D'Imperio, Mariapaola, House, David, 1997. Perception of questions and statements in Neapolitan Italian. In: Proc. EUROSPEECH, vol. 1. pp. 251–254.
Dunbar, Norah E., Burgoon, Judee K., 2005. Perceptions of power and interactional dominance in interpersonal relationships. Journal of Social and Personal Relationships 22, 231–257.
Erickson, Frederick, 2004. Talk and Social Theory: Ecologies of Speaking and Listening in Everyday Life. Polity Press, Cambridge.
Ferreira, Fernanda, Lau, Ellen F., Bailey, Karl G.D., 2004. Disfluencies, language comprehension, and Tree Adjoining Grammars. Cognitive Science 28 (5), 721–749.
Fleiss, Joseph L., 1971. Measuring nominal scale agreement among many raters. Psychological Bulletin 76 (5), 378–382.
Ford, Cecilia E., Thompson, Sandra A., 1996. Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In: Ochs, E., Schegloff, E., Thompson, S.A. (Eds.), Interaction and Grammar, Cambridge University Press, Cambridge, pp. 134–184.
Fox Tree, Jean, 2002. Interpreting pauses and ums at turn exchanges. Discourse Processes 34 (1), 37–55.
Fry, D.B., 1975. Simple reaction-times to speech and non-speech stimuli. Cortex 11 (4), 355–360.
Gibson, James J., 1979. The Ecological Approach to Visual Perception. Houghton Mifflin, Boston.
Giles, Howard, Coupland, Nikolas, Coupland, Justine, 1991. Accommodation theory: communication, context and consequence. In: Giles, H., Coupland, J., Coupland, N. (Eds.), Contexts of Accommodation: Developments in Applied Sociolinguistics, Cambridge University Press, Cambridge, pp. 1–68.
Gnisci, Augusto, Bakeman, Roger, 2007. Sequential accommodation of turn taking and turn length: a study of courtroom interaction. Journal of Language and Social Psychology 26 (3), 234–259.
Goodwin, Charles, 1996. Transparent vision. In: Ochs, E., Schegloff, E., Thompson, S.A. (Eds.), Interaction and Grammar, Cambridge University Press, Cambridge, pp. 370–404.
Gravano, Agustín, 2009. Turn-taking and affirmative cue words in task-oriented dialogue. Unpublished Ph.D. thesis. Columbia University, NY.
Gravano, Agustín, Benus, Stefan, Chávez, Héctor, Hirschberg, Julia, Wilcox, Lauren, 2007. On the role of context and prosody in the interpretation of 'okay'. In: Proceedings of 45th Conference of Association of Computer Linguistics, pp. 800–807.
Gravano, Agustín, Hirschberg, Julia, 2011. Turn-taking cues in task-oriented dialogue. Computer Speech and Language 25 (3), 601–634.
Gravano, Agustín, Hirschberg, Julia, 2009. Turn-yielding cues in task-oriented dialogue. In: Proceedings of SIGDIAL, Association for Computational Linguistics, pp. 253–261.
Gregory, Stanford W., Webster, Stephen, 1996. A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status. Journal of Personality and Social Psychology 70, 1231–1240.
Jefferson, Gail, 1986. Notes on latency in overlap onset. Human Studies 9, 153–183.
Heldner, Mattias, Edlund, Jens, Hirschberg, Julia, 2010. Pitch similarity in the vicinity of backchannels. In: Proceedings of 11th Interspeech Conference, pp. 3054–3057.
Hirschberg, Julia, 2003. Pragmatics and intonation. In: Horn, L., Ward, G. (Eds.), Handbook of Pragmatics, Blackwell, New York, pp. 515–537.
Hirschberg, Julia, Litman, Diane, 1993. Empirical studies on the disambiguation of cue phrases. Computational Linguistics 19 (3), 501–530.
Hirschberg, Julia, Nakatani, Christine, 1996. A prosodic analysis of discourse segments in direction-giving monologues. In: Proceedings of 34th Conference of Association of Computer Linguistics, pp. 286–293.
Itakura, Hiroko, Tsui, Amy, 2004. Gender and conversational dominance in Japanese conversation. Language in Society 33, 223–248.
Knoblich, Günther, Sebanz, Natalie, 2008. Evolving intentions for social interaction: from entrainment to joint action. Philosophical Transactions of the Royal Society B 363, 2021–2031.
Liberman, Alvin M., Harris, Katherine S., Hoffman, Howard S., Griffith, Belver C., 1957. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54 (5), 358–368.
McFarland, David H., 2001. Respiratory markers of conversational interaction. Journal of Speech, Language, & Hearing Research 44, 128–143.
Mushin, Ilana, Stirling, Lesley, Fletcher, Janet, Wales, Roger, 2003. Discourse structure, grounding, and prosody in task-oriented dialogue. Discourse Processes 31 (1), 1–31.
Norman, Donald, 1990. The Design of Everyday Things. Doubleday, New York.
Pardo, Jennifer, 2006. On phonetic convergence during conversational interaction. Journal of Acoustical Society of America 119 (4), 2382–2393.
Pickering, Martin, Garrod, Simon, 2004. Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences 27, 169–226.
Pierrehumbert, Janet, Hirschberg, Julia, 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P., Morgan, J., Pollack, M. (Eds.), Intentions and Plans in Communication and Discourse, MIT Press, Cambridge, MA, pp. 271–311.
Poggi, Isabella, D'Errico, Francesca, 2010. Dominance signals in debates. In: Salah, A.A., et al. (Eds.), HBU 2010, LNCS 6219, Springer-Verlag, Berlin, Heidelberg, pp. 163–174.
Sacks, Harvey, Schegloff, Emanuel, Jefferson, Gail, 1974. A simplest systematic for the organization of turn-taking for conversation. Language 50, 696–735.
Schegloff, Emanuel, 1996. Turn organization: one intersection of grammar and interaction. In: Elinor, Ochs, Sandra, Thompson, Emanuel, Schegloff (Eds.), Interaction and Grammar, Cambridge University Press, Cambridge, pp. 52–133.
Schegloff, Emanuel, 2000. Overlapping talk and the organization of turn-taking for conversation. Language and Society 19, 1–63.
Schegloff, Emanuel, 2007. Sequence Organization in Interaction. Cambridge University Press, Cambridge.
Schiffrin, Deborah, 1987. Discourse Markers. Cambridge University Press, Cambridge.
Scott, Sophie K., Gettigan, Carolyn Mc, Eisner, Frank, 2009. A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. Nature Reviews. Neuroscience 10 (4), 295–302.
Selting, Margret, 1996. On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. Pragmatics 6, 357–388.
Shimojima, Atsushi, Katagiri, Yasuhiro, Koiso, Hanae, Swerts, Marc, 2002. Informational and dialogue-coordinating functions of prosodic features of Japanese echoic responses. Speech Communication 36 (1–2), 113–132.
Shockley, Kevin, Santana, Marie-Vee, Fowler, Carol A., 2003. Mutual interpersonal postural constraints are involved in cooperative conversation. Journal of Experimental Psychology: Human Perception & Performance 29, 326–332.
Shriberg, Elizabeth, Lickley, Robin, 1993. Intonation of clause-internal filled pauses''. Phonetica 50, 172–179.
Sidnell, Jack, 2001. Conversational turn-taking in a Caribbean English Creole. Journal of Pragmatics 33, 1263–1290.
Smiljanic, Rajka, Hualde, Jose I., 2000. Lexical and pragmatic functions of tonal alignment in two Serbo-Croatian dialects. In: Okrent, A., Boyle, J. (Eds.), Chicago Linguistic Society, vol. 36(1). Chicago Linguistic Society, Chicago, pp. 469–482.
Steedman, Mark, 2000. Information structure and the syntax-phonology interface. Linguistic Inquiry 31 (4), 649–689.
Stewart, Oliver W., Corley, Martin, 2008. Hesitation disfluencies in spontaneous speech: the meaning of um. Language and Linguistics Compass 4, 589–602.

Swerts, Marc, 1998. Conversational fillers as markers of discourse structure. Journal of Pragmatics 30, 485–496.
Szczepek Reed, Beatrice, 2010. Speech rhythm across turn transitions in cross-cultural talk-in-interaction. Journal of Pragmatics 42 (4), 1037–1059.
Szczepek Reed, Beatrice, 2006. Prosodic Orientation in English Conversation. Palgrave, Basingstoke.
Taboada, Maite, 2006. Spontaneous and non-spontaneous turn-taking. Journal of Pragmatics 16 (2–3), 329–360.
Tannen, Deborah, 1998. You Just Don't Understand: Women and Men in Conversation. Ballantine, New York.
Ward, Arthur, Litman, Diane, 2007. Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In: Proceedings of SLaTE Workshop on Speech and Language Technology in Education, Farmington, PA.
Ward, Nigel, Tsukahara, Wataru, 2000. Prosodic features which cue back-channel responses in English and Japanese. Journal of Pragmatics 23, 1177–1207.
Wilson, Margaret, Wilson, Thomas P., 2005. An oscillator model of the timing of turn-taking. Psychonomic Bulletin and Review 12 (6), 957–968.
Yuan, Jiahong, Liberman, Mark, Cieri, Christopher, 2007. Towards an integrated understanding of speech overlaps in conversation. In: Proceedings of ICPhS XVI, pp. 1337–1340.

Štefan Beňuš received his PhD in Linguistics from New York University in 2005. His research centers around two areas: (1) the relationship between speech prosody and pragmatic/discourse aspect of the message as well as emotional state of the speaker delivering the message and (2) the relationship between phonetics and phonology with the emphasis on articulatory characteristics of speech.

Agustín Gravano received his PhD in Computer Science from Columbia University in 2009. His main research topic is the representation of prosodic variation in spoken dialogue, aimed at improving the models used in spoken dialogue systems, both for understanding the user's input and for generating natural responses.

Julia Hirschberg received her PhD in Computer Science from the University of Pennsylvania, after previously doing a PhD in sixteenth-century Mexican social history at the University of Michigan. Her main area of research is computational linguistics, specifically the relationship between intonation and discourse. Her current interests include emotional speech (including deceptive and charismatic speech); intonation variation in spoken dialogue systems; speech synthesis; speech search and summarization over large corpora of broadcast news and voicemail; and interfaces to speech corpora.