8. Dialogue

Raquel Fernández, University of Amsterdam Final Draft: October 28, 2013

Abstract

Research on dialogue deals with the study of language as it is used in conversation. Dialogue is a multi-agent activity and this makes conversational language markedly different from the kind of language found in texts. This chapter introduces the main phenomena that characterise language in dialogue interaction — including disfluencies, dialogue acts, alignment, grounding, and turn taking — and discusses some of the key approaches to modelling dialogue that are fundamental in computational dialogue research.

Keywords: linguistic interaction, conversation, spoken language, dialogue acts, interaction management

8.1 Introduction

We use language effortlessly to converse with each other. Research on dialogue is concerned with formulating formal and computational models of how we do this. This is a fascinating enterprise that is also necessarily interdisciplinary, where the concerns of linguistics interface with those of other fields such as psycholinguistics, sociology, and artificial intelligence. The results of this research have an important role to play in computational linguistics and language technology as they provide the basis for the development of systems for the automatic processing of conversational data and of dialogue

systems for human-computer interaction. The present chapter concentrates on how foundational models of dialogue connect with problems in computational linguistics.

Dialogue is, by definition, a multi-agent phenomenon. The central questions in dialogue modelling are therefore concerned with how the dialogue participants interact and coordinate with each other. Conversations mainly serve to exchange information. Thus, one of the key aspects that dialogue models seek to explain is how the dialogue participants collaboratively come to share information during a conversation (contribute to their *common ground*), and how they coordinate the ongoing communicative process. Dialogue is furthermore a highly contextualised form of language use, with speech being produced spontaneously and in an online fashion. An additional challenge for models of dialogue is thus to explain how participants coordinate to take turns in speaking and how they exploit the conversational context to assign meaning to forms that are not always sentential.

We will touch upon all these issues in the following sections. The article is structured in two main parts. Section 8.2 describes the main phenomena observable in natural dialogue that make it a challenging subject for computational linguistics. Here we will introduce basic notions such as utterance, turn, dialogue act, feedback, and multi-modality. In Section 8.3, we will then present particular approaches that have been put forward to model these phenomena. Some of these approaches are experimental or theoretical in nature, but they all have formed the basis for computational exploration of different dialogue issues. The article closes in Section 8.3.8 with pointers to further reading and the main venues for dissemination in the field.

8.2 Basic Notions in Dialogue Research

The most common form of dialogue is spoken face-to-face conversation. The kind of language we use in this setting has specific features that distinguish it from written text. Some of these features have to do with the spontaneous nature of spoken language, while others are the product of the coordination processes that dialogue participants engage in during dialogue. In this section, we describe the main characteristics of language in dialogue and introduce basic notions used in dialogue research.

8.2.1 Turns, Utterances, and Dialogue Acts

What are the basic units of analysis in dialogue? Unlike written text, where language is segmented into sentences by punctuation marks, spoken language as used in natural conversation does not lend itself to be analysed in terms of canonical sentences. Transcribing unrestricted dialogue is indeed a difficult task, which involves making tricky decisions about how to carve up the flow of speech into units. Consider the following excerpt from a transcription of a telephone conversation between two participants, part of file `2028_1086_1101` of the Switchboard corpus (Godfrey et al. 1992):

(8.1) A.1: Okay, {F um. } / How has it been this week for you? /

B.2: Weather-wise, or otherwise? /

A.3: Weather-wise. /

B.4: Weather-wise. / Damp, cold, warm <laughter>. /

A.5: <laughter> {F Oh, } no, / damp. /

B.6: [We have, + we have] gone through, what might be called the four

seasons, {F uh, } in the last week. /

A.7: Uh-huh. /

B.8: We have had highs of seventy-two, lows in the twenties. /

This conversational transcript exemplifies several key aspects of language in dialogue. First and foremost, in contrast to monologue, dialogue involves several participants who take **turns** in speaking. In the example above, there are two dialogue participants, labelled A and B, who exchange eight turns, numbered from 1 to 8. It is not straightforward to define what a turn is, but informally turns may be described as stretches of speech by one speaker bounded by that speaker's silence—that is, bounded either by a pause in the dialogue or by speech by someone else. Turns are important units of dialogues. Later on, in Section 8.3.7, we will look into models that attempt to characterise how dialogue participants organise their turn taking.

The transcript in (8.1) also shows that spoken conversational language is often fragmented. This is due in part to the presence of speech **disfluencies**—repetitions, self-corrections, pauses and so-called *filled pauses*, such as "`um`" and "`uh`" in our example above, that interrupt the flow of speech. Such disfluencies are a hallmark of spoken language and, as the reader may have guessed, they complicate matters for parsers and natural language understanding components in general. As can be seen in (8.1), disfluencies are marked with special annotation characters in the Switchboard corpus, such as square and curly brackets. We will describe the features and the structure of disfluencies in more detail in Section 8.3.6.

Language in dialogue is fragmented in yet another sense. In conversation, unlike in written discourse, it is commonplace to use elliptical forms that lack an explicit predicate-

argument structure, such as bare noun phrases. In (8.1), turns 2 to 5 show examples of such fragments, also called **non-sentential utterances**. These fragments are considered elliptical because despite their reduced syntactic form, when uttered within the context of a dialogue they convey a full message. As we shall see in Section 8.3.5, non-sentential utterances are inherently context-dependent and resolving their meaning requires a highly structured notion of dialogue context.

Disfluencies and elliptical fragments render dialogue language substantially different from written text. Because of this, researchers working on dialogue rarely refer to *sentences* but rather to **utterances**. As with turns, to give a precise definition of utterance is not an easy task. Nevertheless, an utterance may be described as a unit of speech delimited by prosodic boundaries (such as boundary tones or pauses) that forms an *intentional unit*, that is, one which can be analysed as an action performed with the intention of achieving something.[1] In (8.1), utterances are separated by slash symbols, following the transcription conventions of the Switchboard corpus. Note that turns may contain more than one utterance. For instance, the turns in A.1 and B.4 include two utterances each. Sometimes interlocutors complete each other's utterances, as in (8.2) below.[2] In such cases, we may want to say that one single utterance is split across more than one turn.

(8.2) Dan: When the group reconvenes in two weeks=

Roger: =they're gunna issue strait jackets.

Split utterances of this sort, also called **collaborative utterances** (or *collaborative completions*), are common in natural dialogue. We will come back to them in Sec-

tion 8.3.5.

The characterisation of utterances as intentional units is in accordance with the intuition that conversations are made up of sequences of *actions*, each of them reflecting the intentions of its speaker and contributing to the ongoing joint enterprise of the dialogue. For instance, part of the dialogue in (8.1) could intuitively be described as follows: A poses a question requesting some information from B ("How has it been this week for you"); B responds with a request for clarification ("Weather-wise, or otherwise"), which A answers ("Weather-wise"); B then acknowledges that information (" Weather-wise") and goes on to reply to A's initial question ("Damp, cold, warm").

This common-sense view of dialogue as a sequence of actions is at the root of the analytic research tradition initiated by Austin's (1962) work on pragmatics and developed in Searle's (1969, 1975) speech act theory, as discussed in Chapter 7. In contemporary dialogue modelling, actions such as Question, Clarification Request, Answer, or Acknowledgement are considered examples of types of **dialogue act**—a term originally introduced by Bunt (1979) that extends the notion of *speech act* as defined by Searle (1975).[3] In contrast to speech acts, which specify the type of illocutionary force of an utterance, dialogue acts are concerned with the functions utterances play in dialogue in a broader sense. Note that utterances can play more than one function at once. For instance, an utterance such as "Bill will be there" can simultaneously function as an information act and as a promise (or a threat). This will be made more precise in Section 8.3.4 when we look into existing taxonomies of dialogue acts.

If we examine data from dialogue corpora, it becomes apparent that certain patterns of dialogue acts are recurrent across conversations. For instance, questions are

typically followed by answers and proposals are either accepted, rejected, or countered. Such frequently co-occurring dialogue acts (question-answer, greeting-greeting, offer-acceptance/rejection) have been called **adjacency pairs** by sociolinguists working within the framework of Conversation Analysis (Schegloff 1968, Schegloff and Sacks 1973). Adjacency pairs are pairs of dialogue act types uttered by different speakers that often occur next to each other in a particular order. As discussed by Levinson (1983), the key idea behind the concept of adjacency pair is not strict adjacency but *expectation*. Given the first part of a pair (e.g. a question), the second part (e.g. an answer) is immediately relevant and expected in such a way that if the second part does not immediately appear, then the material produced until it does is perceived as an *insertion sequence* or a *sub-dialogue*—as is the case for the question-answer sequence in B.2-A.3 in (8.1) and in the following example from (Clark 1996, p. 242):

(8.3) Waitress: What'll ya have girls?

Customer: What's the soup of the day?

Waitress: Clam chowder.

Customer: I'll have a bowl of clam chowder and a salad.

This indicates that dialogues are in some way structured. In Section 8.3.3 we will describe models that aim at characterising the dynamics and coherence of dialogue.

8.2.2 Joint Action and Coordination

We have seen that dialogues are made up of turns and utterances, and that the functions that utterances play can be analysed in terms of dialogue act types that may be

structured into adjacency pairs. A conversation, however, is not simply a sequence of individual actions (or pairs of actions) performed by independent agents, but also a form of *joint activity* that requires coordination among its participants. In fact, many of the actions performed by the dialogue participants do not directly advance the topic of the conversation but rather function as coordination devices that serve to manage the interaction itself.

One type of utterance with an *interactive communicative function* are **feedback utterances**, which are dedicated to coordinate the speakers' mutual understanding. For instance, acknowledgements such as "Uh-huh" in (8.1) A.7 above and "Yuh" and "Yes" in (8.4) from Levinson (1983) serve to give *positive feedback* regarding the understanding process.

(8.4) B.1: I ordered some paint from you uh a couple of weeks ago some vermilion

A.2: Yuh

B.3: And I wanted to order some more the name is Boyd

A.4: Yes // how many tubes would you like sir

Speakers also employ systematic linguistic means to give *negative feedback* when they encounter trouble during the communication process. This is typically done by means of clarification requests that range from conventional forms such as "Pardon?" to more contentful queries that refer back to particular aspects of previous utterances. We saw one such example in (8.1) above ("Weather-wise, or otherwise"). Turns 2, 6, and 8 in excerpt (8.5) from the British National Corpus (BNC, file KP5) (Burnard 2000) show further examples of clarification requests:

(8.5) B.1: There is not one ticket left in the entire planet! So annoying!

C.2: Where for?

B.3: Crowded House. My brother is going and he doesn't even like them.

A.4: Why doesn't he sell you his ticket?

B.5: Cos he's going with his work. And Sharon.

A.6: Oh, his girlfriend?

B.7: Yes. They are gonna come and see me next week.

A.8: Not Sharon from Essex?

B.9: No, she's Sharon from <laughing> Australia.

Feedback utterances are an example of an explicit mechanism used by the dialogue participants to keep the conversation on track and to manage the collaborative process of ensuring mutual understanding — a process that has been called **grounding** by Clark and Schaefer (1989).[4] Besides the mechanisms that are at play in explicit grounding behaviour, there is also additional evidence of coordination among agents engaged in dialogue. It has been observed that speakers and hearers tend to converge in their choice of linguistic forms at different levels of language processing — a phenomenon that has come to be known as **alignment**. For instance, speakers often adapt their pronunciation to that of their interlocutors and tend to converge on their choice of syntactic constructions and referring expressions. These adaptations take place online during a single dialogue and, according to some models, they are due to automatic psychological processes (Pickering and Garrod 2004). Regardless of its underlying causes, the presence of alignment seems to be pervasive in dialogue. We will describe models of grounding and alignment in Sections 8.3.1 and 8.3.2, respectively.

8.2.3 Multimodality and Communication Medium

So far, we have been concerned with linguistic phenomena that play critical roles in dialogue. We should note, however, that dialogue is a situated activity and as such it is directly affected by the context in which it takes place. As we have mentioned, the most common setting for conversation is face-to-face spoken dialogue. In such a setting, gestures and gaze play an important role. For instance, positive feedback may be given in the form of a head nod; gaze may help to signal a turn switch; and a pointing gesture can act as an answer to a question. Thus, models of face-to-face dialogue ultimately need to be **multimodal**, and a good deal of research in Computational Linguistics nowadays looks at the integration of language with other modalities in both understanding and generation, as discussed in detail in Chapter 42. Wahlster (2006) offers a good overview of the challenges involved in developing multimodal dialogue systems.

Not all dialogue, however, takes place face to face. Other forms of communication such as telephone conversations, text chat or video conferences, also allow dialogue albeit with restrictions. The constraints imposed by each of these modes of **mediated communication** (regarding, e.g., the availability of visual contact or the affordance of simultaneous communication) have an impact on the interaction management mechanisms used by the dialogue participants and hence influence the flow and the shape of the dialogue (Whittaker 2003, Brennan and Lockridge 2006).

8.3 Models of Dialogue Phenomena

In this section we analyse in more detail the fundamental phenomena we have introduced in Section 8.2. We review some of the main approaches to modelling these phenomena and look to recent work in computational linguistics that builds on these models.

8.3.1 Models of Grounding

Conversation can by and large be described as a process whereby speakers make their knowledge and beliefs common, a process whereby they add to and modify their *common ground* (Stalnaker 1978). As we pointed out earlier, this collaborative process is known as **grounding** after Clark and Schaefer (1989). Conversation is also a multi-agent process between two or more individuals who are not omniscient and therefore mutual understanding — and hence successful grounding — is not guaranteed. Models of grounding thus need to explain not only how participants achieve shared understanding and contribute to their common ground, but also how partial understanding or misunderstanding may arise, and how interlocutors may recover from such communication problems. Allwood (1995) and Clark (1996) independently put forward similar theories of communication that take these issues into account. They propose that utterances in dialogue result in a hierarchy of actions that take place at different levels of communication and that, crucially, are performed by both the speaker of an utterance and its recipient. Figure 8.1 shows the four levels of communication proposed, using a synthesis of the terminology employed by these two authors.

This ladder of communicative functions that utterances in dialogue perform is reminiscent of Austin's (1962) classic distinction between an utterance's *locutionary*, *illocu-*

Level	Actions
1 contact:	A and B pay attention to each other
2 perception:	B perceives the signal produced by A
3 understanding:	B understands what A intends to convey
4 uptake:	B accepts / reacts to A's proposal

Figure 8.1: Levels of communication and actions at each level by speaker (A) and addressee (B)

tionary, and *perlocutionary* acts. However, while the work of Austin (and later Searle) focuses on the actions performed by the speaker, models of grounding highlight the fact that conversation requires actions of both speakers *and* addressees. Given the actions of the speaker, the addressee is expected to comply — has an "obligation of responsiveness" in Allwood's words. For instance, consider the utterance "How has it been this week for you?" from our earlier example (8.1). At level 1, speaker and hearer establish contact and mutual attention. With her utterance, the speaker is also presenting a signal for the addressee to perceive (level 2). At level 3, the speaker has the intention to convey a particular meaning and the hearer must recognise her intention for communication to succeed (in this case, the speaker is *asking* a question about a particular issue). Finally, at level 4, the speaker intends to elicit a reaction from the addressee and is hence proposing a *joint project* that the addressee can take up.

Lack of understanding or miscommunication may occur at any of these levels of action: we may not hear our interlocutor properly, we may not know the meaning of a word she uses, or we may hear her and understand the language used in her utterance

but fail to recognise its relevance. To achieve grounding, dialogue participants thus must understand each other at all levels of communication. The degree of mutual understanding they need to achieve, however, may vary with the purpose of the conversation. For instance, in a commercial transaction on the phone, understanding each digit of a credit card number is critical, while understanding every word in a closing utterance such as "Thank you very much and have a nice weekend" is not. Participants must provide evidence that they understand each other up to what Clark (1996) calls the **grounding criterion**, i.e. the appropriate degree of understanding given the communicative situation at hand. According to Clark, the different levels of action are connected by the so-called *principle of downward evidence*, according to which positive evidence of understanding at a particular level can be taken as evidence that the grounding criterion has been reached at lower levels as well. Thus, by replying with "Goodbye" to a partially perceived contribution such as the closing utterance mentioned above, a participant would give evidence of understanding at level 4, and by the principle of downward evidence she would implicitly indicate that the grounding criterion has also been fulfilled at lower levels.

Addressees employ a variety of mechanisms to give evidence that they have understood the speaker up to the grounding criterion. Our "Goodbye" example would be an instance of implicit evidence given by means of a *relevant next contribution*. But other more explicit mechanisms such as feedback utterances are very common as well. Recipients may issue an acknowledgement (such as a nod or a backchannel like *uh huh*) or they may repeat or paraphrase part of the speaker's utterance. Feedback mechanisms of this sort can be classified according to the level of communication at which the evi-

dence of understanding is given. For instance, a repetition indicates that the repeated material has been correctly perceived (level 2) while a paraphrase may give evidence that the speaker's utterance has been not only perceived but also understood (level 3). It is important to note, however, that there is not a one-to-one correspondence between the form of feedback utterances and their function. Acknowledgements such as *yeah*, for instance, may be ambiguous between signals of attention/understanding and signals of acceptance.

Similar kinds of ambiguity apply to the forms of clarification requests. As (8.6) shows, an utterance can give rise to a range of requests for clarification that can be classified according to the communication level at which they signal a problem or an uncertainty (Schlangen (2004) provides a classification along these lines). But several corpus studies (Purver 2004, Rodríguez and Schlangen 2004) have shown that the same types of surface forms can have different functions, especially when the clarification has an elliptical form. For instance, *Goldoni street?* in (8.6) can be taken as requesting confirmation of the words used in the target utterance, as an indication that Goldoni street is unknown to B, or as a signal that B considers Goldoni street an inappropriate choice. Note furthermore that one single utterance can give positive and negative feedback simultaneously. In (8.7), B's clarification request repeats part of A's utterance – this pinpoints the source of the understanding problem while giving positive evidence that the repeated material has been grounded.

(8.6) A: I know a great tapas restaurant in Goldoni street.

B: Pardon? / A great what? / Goldoni street? / Should I consider this an invitation?

(8.7) B: A tapas restaurant where?

Which feedback mechanism is appropriate in a given situation depends on several factors, such as the degree of uncertainty regarding a possible misunderstanding and the desire to be brief and efficient. The so-called *principle of least collaborative effort* states that dialogue participants will try to invest the minimum amount of effort that allows them to reach the grounding criterion.

Giving feedback about the status of the grounding process can be considered *collateral* to the main subject matter of the conversation. Allwood (1995) and Clark (1996) explain the special status of feedback by distinguishing between two layers within the communicative process: a layer corresponding to the communication itself, containing the communicative acts that deal with the subject matter or "official business" of the conversation; and a parallel layer of interaction management or meta-communication that deals with managing the grounding process, as well as other interaction mechanisms such as turn taking. Figure 8.2 shows an extract from an earlier example where acts are classified into these two conversational layers. Unlike other acts that may have implicit consequences at layer 2, the primary function of feedback acts such as acknowledgements and clarification questions is to manage the grounding process. Thus, these feedback acts have the property of being *meta-communicative*: while other types of acts deal with the topic of the conversation, the subject matter of feedback utterances are the basic acts of communication.

The theories we have discussed regarding grounding in human-human dialogue have had an important impact in computational research on dialogue systems and conversational agents. Due to their limited abilities, dialogue systems are prone to misunder-

	Layer 1: basic communicative acts	Layer 2: meta-communicative acts
B:	There is not one ticket left in the entire planet! So annoying!	
C:		Where for?
B:		Crowded House.
B:	My brother is going and he doesn't even like them.	
A:	Why doesn't he sell you his ticket?	*implicit positive evidence*
B:	Cos he's going with his work. And Sharon.	*implicit positive evidence*
A:		Oh, his girlfriend?
B:		Yes.
B:	They are gonna come and see me next week.	

Figure 8.2: Layers of communication

standing. There is thus great need for employing grounding strategies that help to reduce the system's uncertainty regarding the user's utterances and to handle errors when these occur. The collateral status of interaction management (Figure 8.2) makes it possible to implement grounding strategies as domain-independent modules of dialogue systems. Traum (1994) presents one of the earliest computational models of grounding. Other approaches that also build up on the theoretical ideas we have discussed in this section are Paek and Horvitz (2000) and Skantze (2005). More details on error handling strategies in dialogue systems and pointers to additional references can be found in Chapter 41.

8.3.2 Alignment

The grounding process, as we have described in the previous section, refers to the collaborative mechanisms used by dialogue participants to achieve shared understanding. As we have seen, these mechanisms rely on the use of feedback as a means for managing the communication. However, as mentioned in Section 8.2.2, when dialogue participants interact they also coordinate in less explicit ways. There is a fair amount of evidence showing that speakers have a strong tendency to align on the perceptual features of the signals they use in conversation. For instance, dialogue participants rapidly converge on the same vocabulary (Brennan 1996), tend to use similar syntactic structures (Branigan et al. 1995), adapt their pronunciation and speech rate to one another (Pardo 2006), and even mimic their interlocutor's gestures (Kimbara 2006). A number of researchers have also found experimental evidence that human users of dialogue systems adapt several features of their language to the productions of the system (Coulston et al. 2002, Branigan et al. 2010).

The causes underlying the observed convergences seem to be diverse.[5] One of the most influential approaches put forward to explain them is the Interactive Alignment model (Pickering and Garrod 2004), which attributes them to *priming* — an unconscious psychological effect according to which exposure to a particular stimulus or "prime" increases the activation of the corresponding internal representations and therefore it also increases the likelihood of producing behaviour that is identical or related to the prime. Priming is related to memory in such a way that the likelihood of producing forms that have been primed by a previous stimulus decreases as the distance from the prime increases. For instance, controlled psychological experiments done in the

lab have shown that if a subject A describes a scene as "Nun giving a girl a book" to subject B, right after that B is more likely to use the description "Sailor giving a clown a hat" than the alternative description "Sailor giving a hat to a clown". Here the prime can be taken to be the syntactic structure used by A's description with two NPs as complements. Subsequent productions are influenced by priming if the syntactic structure of the potential prime is repeated with higher probability than expected the closer they are from this stimulus. Representations at levels other than syntax can act as primes as well, including phonology, morphology, semantics, gestures, etc.

It is easy to see how priming can lead to the dialogue participants converging on their linguistic (and even gestural) choices. The Interactive Alignment model however goes further to claim that mechanistic effects such as priming underlie successful communication in dialogue. According to Pickering and Garrod (2004), communication succeeds when the *situation models* of the dialogue participants become aligned—i.e. when their representations of what is being discussed in the dialogue are the same for all relevant purposes. The model proposes that alignment of situation models (and thus shared understanding) is achieved by automatic priming mechanisms taking place at different interconnected levels of linguistic processing, which ultimately lead to alignment at the conceptual/semantic level, or in other words, to the building up of common ground.

It should be noted that the Interactive Alignment model is not strictly an alternative to the collaborative models of grounding we discussed earlier. The difference between the two types of approaches is mainly one of focus. The models of Allwood and Clark focus on the strategies employed by the interlocutors, while Pickering and Garrod are concerned with more basic processing mechanisms. The models differ however on how

much of these strategies and mechanisms they consider responsible of successful dialogue. Clark and colleagues consider that shared understanding and successful communication are primarily the result of active collaboration by the participants who jointly work on inferring their common ground given the evidence provided during the conversation. In contrast, the interactive alignment model argues that these strategies only play a substantial role when there is need for repair in situations of *misalignment*, but that in the majority of situations speakers rely on low-level and largely automatic mechanisms such as priming. In any case, it seems clear that both explicit collaborative strategies and implicit convergence contribute to shaping dialogue interaction.

Computational approaches to alignment and convergence can be classified into three main kinds. Firstly, we find corpus-based studies that aim to model the priming effects found in dialogue corpora. These studies use several measures to quantify the degree of priming between dialogue participants and then apply statistical modelling techniques to reproduce it (Reitter et al. 2006, Ward and Litman 2007). This methodology has uncovered several interesting features of alignment effects, such as the fact that priming is stronger in task-oriented dialogue and that it is a good predictor of learning in tutorial dialogue. The second kind of approaches are related to *user adaptation*. The focus here is on the implementation of generation systems or conversational agents that are capable of aligning at different levels with their users (Brockmann et al. 2005), such as on the lexical choices made (Janarthanam and Lemon 2010), the level of formality adopted (de Jong et al. 2008), or the gestures produced (Buschmeier et al. 2010). Finally, the third type of computational approach to alignment does not only aim at modelling alignment of external features but also convergence of the underlying semantic systems that are

part of speakers' linguistic knowledge. Relevant work in this area includes research on category formation and emergent vocabularies between interacting robots (Steels and Belpaeme 2005), computational modelling of concept learning between humans and robots (de Greeff et al. 2009, Skočaj et al. 2011), and formal modelling of the semantic and pragmatic mechanisms at play in processes of semantic coordination in human-human dialogue (Cooper and Larsson 2009, Larsson 2010).

8.3.3 Dialogue Dynamics

Dialogues, like text, appear to be coherent and structured. Models developed to explain this coherence and how it comes about as a dialogue progresses exploit the level of abstraction obtained by classifying utterances in terms of dialogue act types. One of the first approaches put forward to account for the coherence of a conversation were Dialogue Grammars (Sinclair and Coulthard 1975, Polanyi and Scha 1984). We have mentioned above that conversations appear to be made up of recurrent patterns of dialogue acts. Dialogue Grammars were developed as a means to model these patterns. They can be implemented as finite-state machines or sets of phrase-structure rules and are intended for parsing the structure of a dialogue in a way akin to how syntactic grammars are used to parse sentences (see Chapters 4 and 23). Dialogue grammars, however, have been criticised on the grounds that they do not allow enough flexibility and that — similarly to the notion of *adjacency pair* we introduced in Section 8.2.1 — they only offer a descriptive account of the sequential dependencies between dialogue acts but fall short of explaining them.

Several theories have been put forward to explain the mechanisms behind the ob-

served conversational patterns and dialogue coherence more broadly. For instance, plan-based approaches developed within Artificial Intelligence during the 1980's appeal to the beliefs, desires and intentions (BDI) underlying the plans of the speakers (Allen and Perrault 1980, Grosz and Sidner 1986, Cohen et al. 1990). According to this line of research, coherence ensues when utterances and the dialogue acts they realise can be understood as motivated by the plans and goals of the dialogue participants.[6] A more general way to model the dynamics of conversations and their cohesion that is prevalent in current dialogue research is to treat dialogue acts as context-change operators or *update functions* — i.e. to define them in terms of how they relate to the current state of the conversation and how they change it or update it.[7] For instance, acknowledgements can be analysed as changing the status of a particular piece of information (say, a proposition introduced by a previous assertion) from ungrounded to being part of the common ground, while questions can be seen as introducing an obligation for the addressee to address the question in the future dialogue.

In general, actions are characterised by changing the world around us. Dialogue acts, however, are special types of actions in that they bring changes to the assumptions (the knowledge, the commitments, and so forth) of the dialogue participants. The term **information state** is commonly used to refer to the context on which dialogue acts operate. Models of the dynamics of dialogue need to make precise what the components and the structure of information states are. A distinction is often made between *private* and *public* or *shared* information. Private information refers to the information that is only available to each individual participant, such as the personal goals and personal beliefs of each speaker. The plan-based theories we have mentioned above appeal mostly to

private mental attitudes that would be part of this component. In contrast, public information represents the common ground of the participants, that is, the information that becomes shared as the dialogue progresses. Information-state theories tend to put their emphasis on this shared component, which reflects the step-by-step actions that are publicly performed by the participants during a conversation. Different models will structure this component differently. For instance, they may distinguish between grounded and ungrounded information or between the latest dialogue act and the previous dialogue history, or they may highlight elements such as the current question(s) under discussion or the current obligations of the dialogue participants. Some of the most influential information-state theories include Bunt's Dynamic Interpretation Theory (Bunt 1994), Ginzburg's KoS (Ginzburg 1996, 2012) and the so-called Poesio-Traum Theory (PTT) (Poesio and Traum 1997, Poesio and Rieser 2010).

These dynamic approaches to dialogue coherence, which as we have seen focus on the update effects of dialogue acts, have underpinned the Information State Update approach to dialogue management, a framework for the development of the dialogue management component of dialogue systems (see Chapter 41) that is intended as a declarative platform for implementing different types of dialogue theories. The framework is succinctly summarised in Larsson and Traum (2001).

8.3.4 Dialogue Act Taxonomies

As we pointed out in Section 8.2.1, dialogue acts can be seen as a generalisation of speech acts. While Searle (1975) distinguishes between five basic types of speech acts (representatives, directives, commissives, expressives, declarations), taxonomies of dialogue acts

aim to cover a broader range of utterance functions and to be effective as tagsets for annotating actual dialogue corpora. Some of the features that make dialogue acts more suitable for analysing actual dialogues than classic speech acts include the following:

- Incorporation of grounding-related acts: Taxonomies of dialogue acts cover not only acts that have to do with the subject matter of the conversation, but also crucially with grounding and the management of the conversational interaction itself. Thus they may include acts such as Reject, Accept, or Clarify.

- Multi-functionality: Proponents of dialogue act schemes recognise that an utterance may perform several actions at once in a dialogue and thus often allow multiple tags to be applied to one utterance.

- Domain dependence: They also acknowledge the fact that the set of utterance functions to be considered depends – to some extent – on the type of conversational exchange, the task at hand, or the domain or subject matter of the dialogue.[8] Although some taxonomies aim at being domain-independent, when annotating particular types of dialogue they are typically complemented with appropriate domain-dependent tags.

A variety of dialogue act taxonomies have been proposed. One of the most influential ones is the DAMSL schema (Dialogue Act Markup using Several Layers) described in Core and Allen (1997). DAMSL, which is motivated by the grounding theories we reviewed in Section 8.3.1, is organised into four parallel layers: *Communicative Status*, *Information Level*, *Forward Looking Functions* (FLF), and *Backward Looking Functions* (BLF). These four layers or dimensions refer to different types of functions an utterance can play simultaneously. A single utterance thus may be labelled with more than

one tag from each of the layers. The layer Communicative Status includes tags such as Abandoned or Uninterpretable, while tags within the Information Level layer indicate whether an utterance directly addresses the Task at hand, deals with Task Management, or with Communication Management. FLFs include initiating tags such as Assert, Info-Request and Offer that code how an utterance changes the context and constrains the development of the dialogue. BLFs code instead how an utterance connects with the current dialogue context, with tags such as Answer, Accept, Reject, Completion, and Signal-non-Understanding. The following short dialogue shows a sample annotation.

(8.8) Utt1.A: How may I help you?

Inf-level:task

FLF:info-request

Utt2.B: I need to book a hotel room.

Inf-level:task

BLF:answer, accept (Utt1.A)

FLF:assert, action-directive

Another comprehensive taxonomy is the HCRC dialogue structure annotation scheme (Carletta and Isard 1996), which was designed to annotate the HCRC Map Task Corpus of task-oriented dialogues. Similarly to DAMSL, the taxonomy distinguishes between Initiating Moves and Response Moves. In addition, the scheme also codes higher level elements of dialogue structure such as *dialogue games* and *transactions*. Carletta et al. define games as follows: "A conversational game is a sequence of moves starting with an initiation and encompassing all moves up until that initiation's purpose is either

fulfilled or abandoned." Transactions correspond to one step in the task (in this case, navigating through different landmarks on a map) and are built up of several dialogue games.

DAMSL and the HCRC scheme have inspired many subsequent dialogue act taxonomies, including SWBD-DAMSL (Jurafsky et al. 1997) used in the annotation of the Switchboard corpus of two-person telephone conversations, MRDA (Meeting Recorder Dialog Act) designed to annotate the multi-party ICSI Meeting Corpus (Janin et al. 2003, Shriberg et al. 2004), and the dialogue act taxonomy developed for the annotation of the AMI (Augmented Multi-party Interaction) Meeting Corpus (Carletta 2007). DAMSL has also partially inspired DIT++ (Bunt 2011),[9] a very comprehensive and fine-grained taxonomy not tied to any particular corpus that builds on Bunt's Dynamic Interpretation Theory (Bunt 1994).

8.3.5 Fragments

As we saw earlier, utterances in dialogue often have a reduced form that does not correspond to that of a canonical full sentence. According to several corpus studies, around 10% of all utterances in unrestricted dialogue are elliptical fragments (Fernández and Ginzburg 2002, Schlangen and Lascarides 2003).

(8.9) G1: Where are you in relation to the top of the page just now?
F1: About four inches
[HCRC Map Task corpus, dialogue q2nc3]

(8.10) A: It's Ruth birthday.

B: When?

[BNC, file KBW]

Non-sentential utterances such as “About four inches” and “ When?” in the examples above are similar to anaphoric expressions or presuppositions in that, to be interpreted, they require a suitable antecedent in the context. The full message conveyed by these fragments (in this case ‘*I am about four inches from the top of the page just now*’ and ‘*When is Ruth’s birthday?*’, respectively) is recovered by combining their content with salient elements of the dialogue context. As is the case in (8.9) and (8.10), often the material required for resolving the content of the fragment can be found in the latest utterance by the fragment’s addressee, but source and fragment need not be immediately adjacent, as illustrated by the answer “Damp, cold, warm” in our earlier example (8.1) B.4, whose antecedent can be found three turns earlier. Thus, dialogue models that aim at explaining the interpretation of fragments need to make precise the conditions under which antecedents are accessible for fragment resolution in a way akin to discourse models for anaphora resolution (see Chapters 6 and 27).

A particularly interesting kind of fragment are elliptical clarification requests. As discussed earlier, clarification requests are feedback utterances with a meta-communicative function that refer back to acts performed by previous utterances (recall examples (8.6) and (8.7) in Section 8.3.1). This means that in order to account for the interpretation of elliptical clarification requests, we need a highly structured notion of dialogue context that includes not only the content of previous utterances organised in a suitable manner, but also the specific communicative acts performed in the dialogue. Of course the details of what counts as accessible and how resolution takes place vary amongst

dialogue theories. For instance, in Ginzburg's theory, fragments are interpreted by combining their content with the current *question under discussion* or QUD (Ginzburg 2012, Fernández 2006, Ginzburg and Cooper 2004, Purver 2004); in SDRT they are resolved by connecting them to previous dialogue acts by means of the appropriate rhetorical relation (Schlangen 2003, 2004); while Ericsson (2005) proposes a model of fragment interpretation that exploits notions from theories of Information Structure.

We shall finish our discussion of fragmentary utterances in dialogue with a few comments on *collaborative completions* or *split utterances*—utterances that are begun by one speaker and finished by another one. Often, the building parts of a split utterance are fragments, while the overall utterance constitutes a full sentential construction, albeit uttered by different participants in turn, as in example (8.11) from Lerner (1996, p.260):

(8.11) A: Well I do know last week thet=uh Al was certainly very
B: pissed off.

One of the challenges posed by collaborative utterances is that the second part of the split is a guess about how the antecedent utterance is meant to continue. It seems reasonable to assume that in order to complete an utterance, addressees must be able to interpret the ongoing (and possibly partial) utterance they are completing. Furthermore, points of split do not necessarily occur at constituent boundaries but can occur anywhere within an utterance (as evidenced by recent corpus studies: Purver et al. 2009, Howes et al. 2011). To model the ability of speakers to complete each other's utterances thus requires a theory of incremental interpretation—that is, a theory that assigns meaning to utterances progressively as they are being produced, and where the increments that

are being interpreted can be units smaller than constituents. Poesio and Rieser (2010) and Gregoromichelaki et al. (2011) propose formal accounts of collaborative utterances that address these challenges. Computational modelling of collaborative utterances has begun to be explored by researchers working on incremental spoken dialogue systems. Section 4 of Chapter 41 offers more details on this recent line of research.

8.3.6 Models of Disfluencies

In spontaneous dialogue, speakers are not always able to deliver their messages fluently. According to Levelt (1989), disfluencies are the product of the speaker's **self-monitoring**—the online process by which the speaker tries to make sure her speech adheres to her intentions. The production process, like the process of interpretation, takes place incrementally. During this process speakers may stall for time to plan their upcoming speech or revise their ongoing utterance if it is not in accordance with their communicative goals. This gives rise to different types of disfluencies, such as repetitions (or stuttering), corrections, and reformulations.

Regardless of their apparent messy form, speech disfluencies exhibit a fairly regular structure. We already saw an example with annotated disfluencies from the Switchboard Corpus in Section 8.2.1. The following example, also from Switchboard, labels the different elements that can occur in a disfluent utterance using the terminology introduced by Shriberg (1994) (building on Levelt 1983).

(8.12) you get people from [other countries + {E I mean} other parts] of the state

start *reparandum* *editing terms* *alteration* *continuation*

The '+' symbol marks the so-called *moment of interruption* (Levelt 1983). Disfluencies may not contain all the elements we see in (8.12). The presence or absence of the different elements and the relations that hold between them can be used as a basis for classifying disfluencies into different types. The following are examples of some disfluency types considered by Heeman and Allen (1999):

(8.13) a. Abridged repair (only editing terms present, in this case filled pauses):
"I like the idea of, {F uh,} being, {F uh,} a mandatory thing for welfare"

b. Modification repair (reparandum and alteration are present):
"a TI experiment to see how [talk, + Texans talk] to other people"

c. Fresh start (no start, reparandum present; the alteration re-starts the utterance):
"[We were + I was] lucky too that I only have one brother"

The regular patterns of disfluencies can be exploited to automatically detect and filter them away before or along parsing (Heeman and Allen 1999). Some recent computational approaches, however, have started to exploit disfluencies rather than eliminate them. For instance, Schlangen et al. (2009) used disfluencies as predictive features in a machine learning approach to reference resolution in collaborative reference tasks, while some researchers have proposed to generate disfluencies in order to increase the naturalness of conversational systems (Callaway 2003, Skantze and Hjalmarsson 2010).[10] From a more theoretical perspective, the work of Ginzburg et al. (2007, 2013) offers a treatment

of disfluencies that integrates them within a theory of dialogue semantics, building on the similarities between disfluencies due to self-repair mechanisms and other forms of inter-participant repair such as clarification requests—an idea that originated within Conversation Analysis (Schegloff et al. 1977).

8.3.7 Models of Turn Taking

That participants in dialogue take turns in talking is one of the most obvious observations one can make about how conversations are organised. At any given point in a dialogue, one participant holds the *conversational floor*, i.e. has the right to address the other dialogue participants, and that right gets transferred seamlessly back and forth from one participant to the other. Although this is a somewhat idealised picture of conversation, turn changes are indeed accomplished very smoothly, with overlapping speech and long pauses being the exception across cultures (Stivers et al. 2009). How do interlocutors achieve such a systematic distribution of turns? A possible view, suggested by psychologists in the early 70's (see the references cited by Levinson 1983, p. 302), is to assume that the current speaker signals when her turn is over by different means (for instance, by stopping speaking and/or by looking at the addressee) and that the other participants recognise such signals as indication that they can take the turn. An approach along these lines is implemented by some dialogue systems, where the system only takes the turn once the user has explicitly released it. There is however clear evidence that natural turn taking does not proceed in accordance to this view. Pauses at speaker switches are very short, with participants starting to speak just a few hundred milliseconds after the previous speaker has finished the turn.[11] Such precise timing cannot be achieved

by reacting to signals given at turn completion. Thus, models of turn taking need to explain not only the systematic allocation of turns to the participants, but also the fact that speakers are able to predict points where a turn may end before actually reaching those points. An adequate model of turn taking thus needs to be *projective* rather than reactive.

Sacks et al. (1974) argued for precisely such a model in a seminal paper which laid the theoretical foundations for most research on turn taking to date. According to this model, turns consist of *turn constructional units*. The precise nature of these units is left vague by the authors, but their key feature is that they end at **transition relevance places** (TRPs)—points at which speakers may switch. According to Sacks and colleagues, these points are projectable, i.e. they can be predicted online from different surface features of the ongoing turn, such as syntax and intonation. Who takes the floor when a TRP has been reached is governed by a set of ordered rules, which can be summarised as follows: (i) the current speaker may induce a speaker switch or "select" the next speaker by addressing a particular participant directly using a first part of an adjacency pair, such as a question or a greeting;[12] if so, the selected participant is expected to take the turn; (ii) if no particular next speaker is selected in this manner, TRPs offer an opportunity for any participant other than the previous speaker to take the floor, or (iii) for the previous speaker to continue if no one else does.

Such a model, simple as it may seem, makes the right predictions. It predicts that turn changes will occur fast given the projectability of TRPs, that generally only one participant will be speaking at a time, and that overlap, if it occurs, will mostly take place at predictable points: For instance, when more than one speaker compete for grab-

bing the turn in case (ii) above, or when TRPs have been wrongly (but systematically) predicted, as in the following examples from Sacks et al. (1974):[13]

(8.14) a. A: Well if you knew my argument why did you bother to

a: sk

B: Because I'd like to defend my argument

b. Desk: What is your last name Lorraine

Caller: Dinnis

Overlap seems indeed to be rare in dialogue, ranging from 5% reported by some early experimental studies (see review in Levinson 1983, p. 296) to around 12% found across two-party and multi-party dialogue corpora (Çetin and Shriberg 2006). However, as Clark (1996) points out, utterances dealing with the management of the interaction, most prominently acknowledgements and backchannels such as "`uh huh`", are not meant and not perceived as attempts to take the floor and are frequently produced in overlap.[14]

(8.15) `A: Move the train...`

`B: Uhu.`

`A: ...from Avon...`

`B: Yeah.`

`A: ... to Danville.`

From this discussion, we can identify three main aspects of turn taking in dialogue that computational models need to work out in detail. One of them is of course the prediction of TRPs: what kind of cues can be used to reliably predict the end of a turn

as it is being produced? A second aspect concerns who should speak next once a TRP is reached. As we will see in the next section, this is an issue mostly in multi-party conversations involving several candidate speakers besides the current one. Finally, a third aspect concerns the right placement of feedback utterances such as backchannels which, as mentioned, are not subject to the same kind of turn-taking constraints as other utterances devoted to the "official business" of the dialogue.

All these aspects have been studied computationally. For clues that help in predicting TRPs, see amongst others Thórisson (2002), Schlangen (2006), Atterer et al. (2008), Raux and Eskenazi (2009) and Gravano and Hirschberg (2011). Traum and Rickel (2002), Kronlid (2006), Selfridge and Heeman (2010) and Bohus and Horvitz (2011) describe computational models of turn taking in multiparty dialogue, while Cathcart et al. (2003) and Fujie et al. (2005) offer models of backchannel placing.

8.3.8 Multiparty Dialogue

Traditionally, formal and computational studies of dialogue have focused on two-person conversations. However, research on multiparty dialogue – dialogue amongst three or more participants – has gained importance in recent years and has by now become common place. Several aspects related to dialogue interaction become more complex when moving to a multi-party scenario (Traum 2004). For instance, while in two-party dialogue the conversational roles played by the dialogue participants are limited to speaker and addressee, dialogues with multiple agents may involve different types of listeners, such as *side-participants* or *overhearers*. Clark (1996) gives a taxonomy of participant roles based on Goffman (1981). Conversational roles are important for interaction be-

cause they determine who the speaker takes into account when planning a particular utterance, who has responsibility for replying to the speaker's contributions, and more generally who is engaged in the conversation besides having access to it. The grounding process is affected by the increased complexity of a multi-party situation. For instance, we may wonder whether the speaker's utterances should be considered grounded when any of the other dialogue participants has acknowledged them, or whether evidence of understanding is required from every listener. Similarly, turn management becomes more complex as the number of participants increases because more participants are available to take the turn. In addition, the structure of multi-party conversations tends to be more intricate than in two-party dialogue since it is easier to keep several topics open in parallel when there are multiple participants. This makes dialogue segmentation more difficult in the multi-party case.

To investigate these and other issues related to multi-party interaction, several multi-party dialogue corpora have been collected in recent years. In particular, corpora of multi-party meetings such as the ICSI Meeting Corpus (Janin et al. 2003) and the more recent AMI Meeting Corpus (Carletta 2007), which contains multi-modal data and rich annotations, have stimulated much research on multi-party dialogue processing. Some of the tasks that have been addressed include speech recognition in meetings, addressee identification, dialogue segmentation, meeting summarisation, and automatic detection of agreements and disagreements. Renals (2011) gives an overview of research carried out using the AMI Meeting Corpus and provides many references to other studies of multi-party meetings.

Further Reading and Relevant Resources

Within Computational Linguistics, dialogue is still a relatively novel area of research and comprehensive surveys are yet to appear. Nevertheless there are a few resources that are worth mentioning. An excellent short overview of dialogue modelling is given by Schlangen (2005). The chapter on *Dialogue and Conversational Agents* from Jurafsky and Martin (2009) surveys the main features of human dialogue as well as the main approaches to dialogue systems. Ginzburg and Fernández (2010) provide a summary of Ginzburg's theory and point to its connection with dialogue management more broadly. Chapter 6 from Levinson (1983), *Conversational Structure*, gives an extensive and critical overview of Conversation Analysis notions, while Schegloff (2007) provides a more recent review by one of the main CA practitioners. Clark (1996) remains one of the most inspiring texts on language use and dialogue interaction.

SIGdial (the conference of the Special Interest Group on discourse and dialogue of the Association for Computational Linguistics)[15] and SemDial (the workshop series on the semantics and pragmatics of dialogue)[16] are the two main yearly venues where the latest research in the field is presented.

Notes

[1]Note that this definition of utterance, which is prevalent in dialogue research (see, e.g., Traum and Heeman 1997), is different from the one typically used within the speech community, where an utterance—or a *talk-spurt* (Brady 1968)—is simply a unit of speech by one speaker bounded by that speaker's silence, which is closer to our informal definition of turn above.

[2]The example is taken from Lerner (1996, p. 241), who studies split utterances within the framework of Conversation Analysis. The equality symbol (=) indicates that there is no pause between turns.

[3]Other denominations are *communicative act* (Allwood 1978) or *dialogue move* (Power 1979).

[4]In dialogue research, *grounding* refers to the process of reaching mutual understanding which results in adding information to the *common ground* — a notion originally introduced by Stalnaker (1978). See Chapter 7 of this Handbook.

[5]See, for instance, the discussion in Haywood et al. (2003).

[6]See the discussion on modelling context, Gricean pragmatics, and speech acts in Chapter 7 of this Handbook.

[7]The starting point of this view are the dynamic approaches to meaning in philosophy of language (Stalnaker 1978, Lewis 1979) and formal semantics (Kamp and Reyle 1993, Heim 1982, Groenendijk and Stokhof 1991). Chapter 7 and Chapter 5 of this Handbook elaborate on these approaches.

[8]A point that echoes Wittgenstein's (1958) idea of *language games*, according to which utterances are only explicable in relation to the activities in which they play a role.

[9]See also `http://dit.uvt.nl/`.

[10]See also Section 4.2 of Chapter 41.

[11]In a study involving 10 languages from 5 different continents, Stivers et al. (2009) found

that most speaker transitions in question-reponse pairs occur between 0 and 200 milliseconds cross-linguistically.

[12]Recall that first parts of adjacency pairs expect a second part contributed by a different participant; see end of Section 8.2.1.

[13]The colon in "a: sk" (8.14-a) indicates the elongation of the vowel. What seems to be going on in this example is that B had rightly predicted that A's turn would end after "ask" but had not expected the elongation of the vowel, which results in a brief overlap. Similarly in (8.14-b), the address term "Lorraine" had not been projected.

[14]Backchannels are also called *continuers* as often they are used by the addressee to encourage the speaker to go on with her turn.

[15]`http://www.sigdial.org/`

[16]`http://www.illc.uva.nl/semdial/`

Bibliography

Allen, James F. and C. Raymond Perrault (1980). 'Analyzing intention in utterances'. *Artificial Intelligence*, 15(3), 143–178.

Allwood, Jens (1978). 'On the analysis of communicative action'. In Brenner, editor, *The Structure of Action*, 168–191. Basil Blackwell.

Allwood, Jens (1995). 'An activity-based approach to pragmatics'. Gothenburg Papers of Theoretical Linguistics 76, Göteborg University, Sweden.

Atterer, Michaela, Timo Baumann, and David Schlangen (2008). 'Towards incremental end-of-utterance detection in dialogue systems'. In *Proceedings of COLING*. Manchester, UK.

Austin, John L. (1962). *How to do things with words*. Oxford University Press.

Bohus, Dan and Eric Horvitz (2011). 'Multiparty turn taking in situated dialog: Study, lessons, and directions'. In *Proceedings of the SIGDIAL 2011 Conference*, 98–109. Association for Computational Linguistics, Portland, Oregon.

Brady, Paul T. (1968). 'A statistical analysis of on-off patterns in 16 conversations'. *The Bell System Technical Journal*, 47, 73–91.

Branigan, Holly P., Martin J. Pickering, Simon P. Liversedge, Andrew J. Stewart, and Thomas P. Urbach (1995). 'Syntactic priming: Investigating the mental representation of language'. *Journal of Psycholinguistic Research*, 24(6), 489–506. ISSN 0090-6905.

Branigan, Holly P., Martin J. Pickering, Jamie Pearson, and Janet F. McLean (2010). 'Linguistic alignment between people and computers'. *Journal of Pragmatics*, 42(9), 2355–2368. ISSN 0378-2166.

Brennan, S. E. and Calion B. Lockridge (2006). 'Computer-mediated communication: A cognitive science approach'. In Brown, K., editor, *Encyclopedia of Language and Linguistics*, 775–780. Elsevier, 2nd edition.

Brennan, Susan E. (1996). 'Lexical entrainment in spontaneous dialog'. In *Proc. of the International Symposium on Spoken Dialogue*, 41–44.

Brockmann, Carsten, Amy Isard, Jon Oberlander, and Michael White (2005). 'Modelling alignment for affective dialogue'. In *Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling (UM-05)*. Edinburgh, UK.

Bunt, Harry (1979). 'Conversational principles in question-answer dialogues'. In Kallmann, D. and G. Stickel, editors, *Zur Theorie der Frage*, 119–141. Narr Verlag, Tübingen.

Bunt, Harry (1994). 'Context and dialogue control'. *THINK Quarterly*, 3(1), 19–31.

Bunt, Harry (2011). 'Multifunctionality in dialogue'. *Computer Speech & Language*, 25(2), 222–245.

Burnard, Lou (2000). *Reference Guide for the British National Corpus (World Edition).* Oxford University Computing Services.

Buschmeier, Hendrik, Kirsten Bergmann, and Stefan Kopp (2010). 'Adaptive expressiveness – Virtual conversational agents that can align to their interaction partner'. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, 91–98. Toronto, Canada.

Callaway, Charles (2003). 'Do we need deep generation of disfluent dialogue?' In *Proceedings of the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, 6–11. Palo Alto, CA.

Carletta, Jean (2007). 'Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus'. *Language Resources and Evaluation*, 41(2), 181–190.

Carletta, Jean and Amy Isard (1996). 'Hcrc dialogue structure coding manual'. Technical report, Centre, University of Edinburgh.

Cathcart, Nicola, Jean Carletta, and Ewan Klein (2003). 'A shallow model of backchannel continuers in spoken dialogue'. In *Proceedings of the tenth conference of the European chapter of the Association for Computational Linguistics*, 51–58.

Çetin, Özgür and Elizabeth Shriberg (2006). 'Overlap in Meetings: ASR Effects and Analysis by Dialog Factors, Speakers, and Collection Site'. In Renals, S., S. Bengio, and J. G. Fiscus, editors, *The Third International Workshop on Machine Learning for Multimodal Interaction (MLMI). Revised Selected Papers*, volume 4299 of *Lecture Notes in Computer Science*, 212–224. Springer.

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

Clark, Herbert H. and Edward F. Schaefer (1989). 'Contributing to discourse'. *Cognitive Science*, 13(2), 259–294.

Cohen, Philip, Jerry Morgan, and Martha Pollack, editors (1990). *Intentions in Communication*. The MIT Press.

Cooper, Robin and Staffan Larsson (2009). 'Compositional and ontological semantics in learning from corrective feedback and explicit definition'. In *Proc. of SemDial 2009: 13th Workshop on the Semantics and Pragmatics of Dialogue*, 59–66.

Core, Mark and James Allen (1997). 'Coding dialogues with the DAMSL annotation scheme'. In Traum, D., editor, *Proceedings of the 1997 AAAI Fall Symposium on Communicative Action in Humans and Machines*.

Coulston, Rachel, Sharon Oviatt, and Courtney Darves (2002). 'Amplitude convergence in childrens conversational speech with animated personas'. In *Proceedings of the 7th International Conference on Spoken Language Processing*, volume 4, 2689–2692.

de Greeff, Joachim, Frederic Delaunay, and Tony Belpaeme (2009). 'Human-robot interaction in concept acquisition: a computational model'. In Triesch, J. and Z. Zhang, editors, *Proceedings of the IEEE 8th International Conference on Development and Learning*, 1–6.

de Jong, Markus, Mariët Theune, and Dennis Hofs (2008). 'Politeness and alignment in dialogues with a virtual guide'. In *Proceedings of the 7th international joint conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, 207–214.

Ericsson, Stina (2005). *Information Enriched Constituents in Dialogue*. Ph.D. thesis, Göteborg University.

Fernández, Raquel (2006). *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. Ph.D. thesis, King's College London, University of London.

Fernández, Raquel and Jonathan Ginzburg (2002). 'Non-sentential utterances: A corpus study'. *Traitement automatique des languages. Dialogue*, 43(2), 13–42.

Fujie, Shinya, Kenta Fukushima, and Tetsunori Kobayashi (2005). 'Back-channel feedback generation using linguistic and nonlinguistic information and its application to spoken dialogue system'. In *Ninth European Conference on Speech Communication and Technology (Interspeech)*, 889–892.

Ginzburg, Jonathan (1996). 'Interrogatives: Questions, facts and dialogue'. In Lappin, S., editor, *The Handbook of Contemporary Semantic Theory*, 385–422. Blackwell.

Ginzburg, Jonathan (2012). *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Ginzburg, Jonathan and Robin Cooper (2004). 'Clarification, Ellipsis, and the Nature of Contextual Updates'. *Linguistics and Philosophy*, 27(3), 297–366.

Ginzburg, Jonathan and Raquel Fernández (2010). 'Computational models of dialogue'. In Clark, Fox, and Lappin, editors, *Handbook of Linguistics and Natural Language Processing*. Blackwell.

Ginzburg, Jonathan, Raquel Fernández, and David Schlangen (2007). 'Unifying self– and

other–repair'. In *Proceedings of SemDial 2007 (Dekalog) the 11th Workshop on the Formal Semantics and Pragmatics of Dialogue*, 57–63. University of Trento, Rovereto.

Ginzburg, Jonathan, Raquel Fernández, and David Schlangen (2013). 'Disfluencies as intra-utterance dialogue moves'. *Semantics & Pragmatics*. Forthcoming.

Godfrey, John J., Edward C. Holliman, and Jane McDaniel (1992). 'SWITCHBOARD: Telephone Speech Corpus for Research and Devlopment'. In *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, 517–520. San Francisco, USA.

Goffman, Erving (1981). *Forms of Talk*. University of Pennsylvania Press.

Gravano, Agustín and Julia Hirschberg (2011). 'Turn-taking cues in task-oriented dialogue'. *Computer Speech & Language*, 25(3), 601–634.

Gregoromichelaki, Eleni, Ruth Kempson, Matthew Purver, Gregory J. Mills, Ronnie Cann, Wilfried Meyer-Viol, and Patrick G. T. Healey (2011). 'Incrementality and intention-recognition in utterance processing'. *Dialogue & Discourse*, 2(1), 199–233. ISSN 2152-9620.

Groenendijk, Jeroen and Martin Stokhof (1991). 'Dynamic Predicate Logic'. *Linguistics and Philosophy*, 14(1), 39–100.

Grosz, Barbara J. and Candace L. Sidner (1986). 'Attention, intentions, and the structure of discourse'. *Computational Linguistics*, 12(3), 175–204.

Haywood, Sarah, Martin Pickering, and Holly Branigan (2003). 'Co-operation and co-ordination in the production of noun phrases'. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, 533–538.

Heeman, Peter A. and James F. Allen (1999). 'Speech repairs, intonational phrases, and discourse markers: modeling speakers' utterances in spoken dialogue'. *Computational Linguistics*, 25(4), 527–571.

Heim, Irene (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts at Amherst.

Howes, Christine, Matthew Purver, Patrick G. T. Healey, Gregory J. Mills, and Eleni Gregoromichelaki (2011). 'On incrementality in dialogue: Evidence from compound contributions'. *Dialogue & Discourse*, 2(1), 279–311. ISSN 2152-9620.

Janarthanam, Srinivasan and Oliver Lemon (2010). 'Learning to adapt to unknown users: Referring expression generation in spoken dialogue systems'. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 69–78.

Janin, Adam, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elisabeth Shriberg, Andreas Stolcke, and Chuck Wooters (2003). 'The ICSI Meeting Corpus'. In *Proceedings of ICASSP'03, the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 364–367.

Jurafsky, Dan, Elizabeth Shriberg, and Debra Biasca (1997). 'Switchboard swbd-damsl shallow-discourse-function-annotation coder's manual, draft 13.' Technical Report TR 97-02, Institute for Cognitive Science, University of Colorado at Boulder.

Jurafsky, Daniel and James H. Martin (2009). *Speech and Language Processing*. Prentice Hall, 2nd edition.

Kamp, Hans and Uwe Reyle (1993). *From Discourse to Logic*. Kluwer Academic Publishers.

Kimbara, Irene (2006). 'On gestural mimicry'. *Gesture*, 6(1), 39–61. ISSN 1568-1475.

Kronlid, Fredrik (2006). 'Turn taking for artificial conversational agents'. In *Cooperative Information Agents X*, volume 4149 of *Lecture Notes in Computer Science*, 81–95. Springer.

Larsson, Staffan (2010). 'Accommodating innovative meaning in dialogue'. In *Proc. of SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, 83–90.

Larsson, Staffan and David Traum (2001). 'Information state and dialogue management in the TRINDI dialogue move engine toolkit'. *Natural Language Engineering*, 6(3&4), 323–340.

Lerner, Gene H. (1996). 'On the semi-permeable character of grammatical units in conversation: Conditional entry into the turn space of another speaker'. In Ochs, E., E.A. Schegloff, and S.A. Thompson, editors, *Interaction and grammar*, 238–276. Cambridge University Press.

Levelt, Willem J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.

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

Levinson, Stephen C. (1983). *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Lewis, David (1979). 'Score keeping in a language game'. *Journal of Philosophical Logic*, 8, 339–359.

Paek, Tim and Eric Horvitz (2000). 'Conversation as action under uncertainty'. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, 455–464.

Pardo, Jennifer S. (2006). 'On phonetic convergence during conversational interaction'. *The Journal of the Acoustical Society of America*, 119, 2382–2393.

Pickering, Martin J. and Simon Garrod (2004). 'Toward a mechanistic psychology of dialogue'. *Behavioral and Brain Sciences*, 27(02), 169–190. ISSN 1469-1825.

Poesio, Massimo and Hannes Rieser (2010). 'Completions, Coordination, and Alignment in Dialogue'. *Dialogue & Discourse*, 1(1), 1–89.

Poesio, Massimo and David Traum (1997). 'Conversational actions and discourse situations'. *Computational Intelligence*, 13(3), 309–347.

Polanyi, Livia and Remco Scha (1984). 'A syntactic approach to discourse semantics'. In *Proceedings of the 10th international conference on Computational Linguistics*, 413–419. Association for Computational Linguistics.

Power, Richard J. (1979). 'The organisation of purposeful dialogues'. *Linguistics*, 17, 107–152.

Purver, Matthew (2004). *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, King's College, University of London.

Purver, Matthew, Christine Howes, Eleni Gregoromichelaki, and Patrick G. T. Healey (2009). 'Split utterances in dialogue: a corpus study'. In *Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue (SIGDIAL 2009 Conference)*, 262–271. Association for Computational Linguistics, London, UK. ISBN 978-1-932432-64-0.

Raux, Antoine and Maxine Eskenazi (2009). 'A finite-state turn-taking model for spoken dialog systems'. In *Proceedings of The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 629–637. Association for Computational Linguistics.

Reitter, David, Frank Keller, and Johanna D. Moore (2006). 'Computational modelling of structural priming in dialogue'. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume*, 121–124.

Renals, Steve (2011). 'Automatic analysis of multiparty meetings'. *SADHANA - Academy Proceedings in Engineering Sciences*, 36(5), 917–932.

Rodríguez, Kepa and David Schlangen (2004). 'Form, intonation and function of clarification requests in German task oriented spoken dialogues'. In *Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue (Catalog)*.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson (1974). 'A simplest systematics for the organization of turn-taking for conversation'. *Language*, 50(4), 696–735.

Schegloff, Emanuel A. (1968). 'Sequencing in conversational openings'. *American anthropologist*, 70(6), 1075–1095.

Schegloff, Emanuel A. (2007). *Sequence organization in interaction: A primer in Conversation Analysis I*, volume 1. Cambridge Univ Pr.

Schegloff, Emanuel A., Gail Jefferson, and Harvey Sacks (1977). 'The preference for self-correction in the organization of repair in conversation'. *Language*, 53(2), 361–382.

Schegloff, Emanuel A. and Harvey Sacks (1973). 'Opening up Closings'. *Semiotica*, 4(7), 289–327.

Schlangen, David (2003). *A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue*. Ph.D. thesis, University of Edinburgh.

Schlangen, David (2004). 'Causes and strategies for requesting clarification in dialogue'. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, 136–143. Association for Computational Linguistics, Cambridge, Massachusetts, USA.

Schlangen, David (2005). 'Modelling dialogue: Challenges and approaches'. *Künstliche Intelligenz*, 3, 23–28.

Schlangen, David (2006). 'From reaction to prediction: Experiments with computational models of turn-taking'. In *Proceedings of Interspeech*.

Schlangen, David, Timo Baumann, and Michaela Atterer (2009). 'Incremental reference resolution: The task, metrics for evaluation, and a Bayesian filtering model that is sensitive to disfluencies'. In *Proceedings of the SIGDIAL 2009 Conference*, 30–37. Association for Computational Linguistics, London, UK.

Schlangen, David and Alex Lascarides (2003). 'The interpretation of non-sentential utterances in dialogue'. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*.

Searle, John R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Searle, John R. (1975). 'Indirect speech acts'. In Cole, P. and J. Morgan, editors, *Syntax and Semantics 3: Speech Acts*, 59–82. Academic Press, New York.

Selfridge, Ethan O. and Peter A. Heeman (2010). 'Importance-driven turn-bidding for spoken dialogue systems'. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 177–185. Association for Computational Linguistics.

Shriberg, Elizabeth, Raj Dhillon, Sonali Bhagat, Jeremy Ang, and Hannah Carvey (2004). 'The icsi meeting recorder dialog act (mrda) corpus'. In Strube, Michael and Candy Sidner, editors, *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, 97–100. Association for Computational Linguistics, Cambridge, Massachusetts, USA.

Shriberg, Elizabeth E. (1994). *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of California at Berkeley, Berkeley, USA.

Sinclair, John McHardy and Malcolm Coulthard (1975). *Towards an analysis of discourse: The English used by Teachers and Pupils*. Oxford University Press London.

Skantze, Gabriel (2005). 'Exploring human error recovery strategies: Implications for spoken dialogue systems'. *Speech Communication*, 45(3), 325–341.

Skantze, Gabriel and Anna Hjalmarsson (2010). 'Towards incremental speech generation in dialogue systems'. In *Proceedings of the 11th Annual SIGdial Meeting on Discourse and Dialogue*, 1–8. Association for Computational Linguistics, Tokyo, Japan.

Skočaj, Danijel, Matej Kristan, Alen Vrečko, Marko Mahnič, Miroslav Janíček, Geert-Jan M. Kruijff, Marc Hanheide, Nick Hawes, Thomas Keller, Michael Zillich, and Kai

Zhou (2011). 'A system for interactive learning in dialogue with a tutor'. In *IEEE/RSJ International Conference on Intelligent Robots and Systems IROS 2011*. San Francisco, CA, USA.

Stalnaker, Robert (1978). 'Assertion'. *Syntax and Semantics*, 9, 315–332. New York Academic Press.

Steels, Luc and Tony Belpaeme (2005). 'Coordinating perceptually grounded categories through language: A case study for colour'. *Behavioral and Brain Sciences*, 28(4), 469–489.

Stivers, Tanya, Nicholas J Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter De Ruiter, Kyung-Eun Yoon, et al. (2009). 'Universals and cultural variation in turn-taking in conversation'. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592.

Thórisson, Kristinn R. (2002). 'Natural turn-taking needs no manual: Computational theory and model, from perception to action'. In *Multimodality in language and speech systems*, 173–207. Kluwer Academic Publishers.

Traum, David (1994). *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, University of Rochester.

Traum, David (2004). 'Issues in multi-party dialogue'. In Dignum, F., editor, *Advances in Agent Communication*, volume 2922, 201–211. Springer-Verlag.

Traum, David and Peter Heeman (1997). 'Utterance units in spoken dialogue'. In *Dialogue processing in spoken language systems*, 125–140. Springer.

Traum, David and Jeff Rickel (2002). 'Embodied agents for multi-party dialogue in immersive virtual worlds'. In *Proceedings of the first international joint conference on Autonomous Agents and Multiagent Systems*, 766–773. ACM.

Wahlster, Wolfgang, editor (2006). *SmartKom: Foundations of Multimodal Dialogue Systems*. Cognitive Technologies. Springer.

Ward, Arthur and Diane Litman (2007). 'Dialog convergence and learning'. In *Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, 262–269. IOS Press.

Whittaker, Steve (2003). 'Theories and methods in mediated communication'. In Graesser, A., M. Gernsbacher, and S. Goldman, editors, *The handbook of discourse processes*, 243–286. Erlbaum.

Wittgenstein, Ludwig (1958). *Philosophical Investigations*. Blackwell, Oxford.