Pragmatics is the study of utterance meaning, and it is well known that prosody – or, more informally, 'tone of voice' – can contribute crucially to that meaning. Pragmatic effects in speech are thus the product of both what is said and how it is said, and the two are inextricably linked. However, while many working in pragmatics are well aware of the important contribution of prosody, exactly how these effects are generated is harder to establish. A number of the ways in which prosody plays a pragmatic role are set out in this volume. It aims to give a cross-section of the many different topics and approaches within the field of prosody and its interface with pragmatics.

Levinson (1983), in his textbook on pragmatics, acknowledged that the absence of prosody from his account, particularly intonation, was a serious omission, but justified the omission on two grounds: first that there was as yet no agreement on how to analyse intonation, and second that the area was understudied. Twenty-five years on, the American autosegmental model, captured in the ToBI transcription system, has become the international standard in intonational phonology (e.g. Ladd, 1996; Gussenhoven, 2004), and for typological comparison (Jun, 2005), but other models still continue to have currency, including variations of the British system of holistic contours (fall, rise, fall-rise etc.). None, however, accounts sufficiently for all pragmatic effects of prosody – which also derive from the kind of effects often referred to as paralinguistic.

Levinson’s second caveat, that the area was understudied, is easier to counter. Interest in the contribution of prosody to pragmatic meaning has grown markedly in the intervening decades, albeit in a fragmented way, and from a wide variety of theoretical perspectives. The scope of these developments depends on one’s view of pragmatics. The Anglo-American approach
(cf. Hirschberg, 2004) focuses on the role of prosody in relation to other linguistic systems, for example in resolving syntactic ambiguities, signalling information and discourse structure and identifying speech acts. The European tradition, on the other hand, takes a broader view of pragmatics, including the above phenomena but also adopting a wider cognitive, social and cultural perspective on meaning in context. This volume includes studies across the whole range of work on prosody and meaning, carried out within different theoretical frameworks. Topics range from studies of interaction management and turn sequencing to pragmatic implicatures and cognitive processes. Analyses include both auditory and instrumental approaches, ranging from corpus studies to case studies, and include features of intonation, timing and rhythm, phrasing and voice quality. Each chapter makes a unique contribution to the field from its own theoretical perspective, and together they form a multi-dimensional view of the role of speech prosody in generating speaker meaning.

In this introductory chapter we will prepare the ground by first outlining the prosodic resources available to speakers (Section 1), and then describing past work in the areas touched on in the chapters we have included (Section 2), in order to show each contribution in its methodological and theoretical context. Finally, we summarise the individual contributions themselves in Section 3, and briefly outline how they relate to previous research.

1. Prosodic Resources

We use the term ‘prosody’ in this volume to refer not only to intonation – i.e. pitch movement – but also to other suprasegmental features such as timing, loudness and voice quality. A seminal study in this regard is that of Crystal (1969), and readers are also referred to detailed accounts in the following publications: Bolinger (1989), Cruttenden (1997), Ladd (1996) and Gussenhoven (2004), which deal extensively with intonation; Wichmann (2000) and Wennerstrom (2001), which focus on prosody in discourse; and the edited collections by Couper-Kuhlen and Selting (1996) and Couper-Kuhlen and Ford (2004), which provide a good overview of prosody in conversation.

In the following paragraphs we outline briefly the role that prosodic features may play in the generation of pragmatic meaning.

Pitch (intonation) is the perceptual correlate of fundamental frequency (F0). The main difficulty in analysing intonation is that both paralinguistic and linguistic information are carried in the same acoustic channel. Pragmatic effects can be generated in a number of ways: a single phonological choice (e.g. a falling or rising contour) in a given context can turn a declarative utterance into a question, or a question into a command, or can indicate whether or not a speaker wishes to continue speaking. The choice of contour (or, in the autosegmental system, a pitch target) is generally a linguistic one, while the realisation of an individual contour – e.g. the choice of pitch range – can have a discoursal function and also a paralinguistic one. The realisation of a
whole utterance, e.g. with an extra wide or extra narrow pitch range, may signal the rhetorical relation between the utterance and that which precedes it (e.g. continuation, digression, elaboration), or it can add an element of affect, reflecting the emotional state of the speaker, such as surprise, anger or sadness. Pitch can also play a part in accentual highlighting by creating local prominences. Pitch prominence on an individual syllable may be emphatic, or may indicate the beginning of a new topic. Prominent syllables may also be a guide, in some languages, to information structure – distinguishing between what is psychologically shared or accessible knowledge (‘given’ information) and that which is new. This distinction can be exploited for pragmatic effect.

Some pragmatic effects are generated by aspects of speech **timing**. In English, for example, the length of syllables is related to accentuation: stressed syllables are longer than the same syllable in unstressed position; a lengthened syllable can also be a cue to a phrase boundary (a phenomenon known as ‘final lengthening’). The overall speech or articulation rate of an utterance, including the distribution and length of pauses, can indicate stylistic variables, such as whether the speech is scripted or spontaneous, casual or formal, or – as in the case of parentheses, for instance – how the current information is related to the stretch surrounding it.

**Loudness**, the auditory correlate of intensity, can also make an important contribution: a stressed syllable is usually louder (and longer) than the same syllable in unstressed position. A sudden drop in loudness (together with an increase in tempo) may signal a parenthetical remark, and a ‘raised voice’ may indicate anger, a competitive interruption, or simply that the hearer is some distance away.

The final component, **voice quality**, is difficult to measure and therefore referred to in sometimes quite impressionistic terms (‘breathy’, ‘tense’, ‘harsh’ ‘rasping’). A degree of breathiness, for instance, can indicate closeness or intimacy, while an angry voice is usually also a tense voice. Some voice qualities, such as creak, seem to occur regularly at the end of turns at speaking, and can be a signal of finality. At the same time, however, overall voice quality can also simply be characteristic of an individual.

All these prosodic features – pitch, timing, loudness and voice quality – operate both locally (at the level of the syllable or below) and globally (at the level of part of an utterance, an utterance as whole or a text). The universality of features described here has not yet been fully explored, but all have the potential to contribute to speaker meaning.

### 2. Approaches to Prosody and Meaning

In broad terms, the functions of speech prosody have been studied in relation to referential meaning, discourse and textual meaning, organising and maintaining interaction, and attitudinal and interpersonal meaning. Some studies focus on speaker behaviour; others pay greater
attention to hearer interpretation and inferencing, while still others emphasise sequential aspects of talk.

Not surprisingly, these topics are approached from within very different theoretical frameworks, and are also differently motivated. Early descriptions of the relationship between intonation and meaning (e.g. O’Connor and Arnold, 1961) were pedagogically motivated, a tradition continued more recently by Wells (2006). In the last two decades, work has been largely driven by the needs of speech technology: speech synthesis, automatic speech recognition and latterly interactive dialogue systems. Much of this work has been carried out either experimentally, using laboratory speech, or based on automatic analysis of speech corpora (see, e.g. Botinis, 2000). There have also been considerable advances in the development of intonational phonology (e.g. Ladd, 1996).

A separate development, with very different methods and goals, has been the interest in prosody within the ethnographic framework of Conversation Analysis (CA), an approach in which evidence is drawn from participants’ response in the emergent talk. Work in this tradition was based initially on impressionistic transcriptions of prosodic phenomena (e.g. Sacks et al., 1974; Jefferson, 1985), but later on close auditory and acoustic phonetic analysis, as in ‘Phonology for Conversation’ (e.g. Local et al., 1986; Kelly and Local, 1989; Local and Walker, 2004), and, more recently, Interactional Linguistics (e.g. Couper-Kuhlen and Selting, 1996; Selting and Couper-Kuhlen, 2001a, b; Couper-Kuhlen and Ford, 2004).

In an approach taking specific account of social context, there is growing interest in the role of prosody in politeness (and impoliteness) using a Gricean framework, in which politeness, or other aspects of speaker stance, are seen as prosodically generated implicatures (e.g. Culpeper et al., 2003). Finally, researchers working in Relevance Theory (Sperber and Wilson, 1985/1995), an approach to communication which is embedded in a general view of human cognition, have turned their attention to the role of prosody in the inferential process (Vandepitte, 1989; House, 1990; Fretheim, 2002; Wilson and Wharton, 2006).

2.1. Referential and discourse/textual meaning

Referential meaning can be signalled prosodically in a number of ways, most frequently by means of accent placement or phrasing. An accented syllable is signalled by prosodic features such as additional length, loudness and pitch change, and accent placement can help to clarify anaphoric pronominal reference in cases of ambiguity (see Lakoff, 1971 and Thomas, 1995: 69 for examples).

Much of the research on referential meaning has been carried out on better-known languages. However, in this volume, the chapter by Blythe brings an entirely new dimension to the prosody of reference by describing a language in which person reference is fraught with taboo, and in which the identity of referents is not indicated by local features such as accent placement, but by global prosodic characteristics such as pitch mimicry.
The location and shape of prosodic prominences also play a part in the signalling of information structure, especially in languages such as English with less flexible constituent ordering. Much of the discussion revolves around notions of focus, givenness and accessibility, and tends to assume an almost iconic (inverse) relationship between prosodic prominence and psychological salience. Information status is seen variously as gradient (Hajičová, 1993), or categorical (e.g. Prince 1981; Gundel et al., 1993; Chafe, 1994), but there is no agreement among authors on the exact number of categories. Prosodic attenuation has been seen as a sufficient, but not a necessary condition for givenness (or activeness of the discourse referent; e.g. Lambrecht, 1994: 97), because given information may indeed be accented. A number of authors have thus distinguished between different degrees of givenness and the corresponding prosodic realisation (e.g. Allerton, 1978; Baltazani, 2006; Baumann and Grice, 2006). Calhoun’s contribution to this volume ties in with this kind of research.

Meaning is also conveyed by the location of phrase boundaries. Speech comes in smaller chunks – intonation domains – and exactly how a stretch of speech is divided into smaller units is assumed to be relevant for the interpretation of a message. The mapping of syntax to prosody has a long tradition in relation to prosodic phrasing and prominence location (e.g. Downing, 1970; Selkirk, 1986, 1995 and related work; Nespor and Vogel, 1986; Truckenbrodt, 1995, 1999). Generally speaking, syntactic structure has been seen as input for prosodic structure, and syntactic constituent boundaries have been argued to systematically coincide with prosodic constituent boundaries. This is also important, for example, for the realisation and interpretation of utterances that are potentially ambiguous (e.g. Price et al., 1990; Schafer et al., 2000; Warren et al., 2000; Hirschberg, 2004).

However, along with syntactic constituency, several other factors are known to play a role in intonational phrasing. Among these are syntactic length and complexity, prosodic length and weight, and balanced prosodic constituent size. Also included are performance factors such as speech rate and style of speech, focus and contrastive prominence, and semantic coherence (see, e.g. Gee and Grosjean, 1983; Selkirk, 1984, 2000, 2005; Nespor and Vogel, 1986, among many others). In this volume, Degand and Simon address questions related to the factors governing the relation between syntax and prosody, taking discourse genre into account.

Much research has also been devoted to the ways in which speakers use prosody to signal text structure. These observations have a long tradition, going back to notions of ‘paragraph intonation’ and ‘paratones’ (Brown, 1977; Brown and Yule, 1983), relating to global pitch trends across spoken texts. Studies of scripted monologue have found co-occurrences of phonetic features signalling a new topic: for instance a long pause, followed by a high (and sometimes delayed) pitch peak and increased amplitude (Lehiste, 1975; Wichmann et al., 2000). In contrast to the extra-high beginning, a compressed initial accent is a typical cohesive device to indicate close rhetorical relations between consecutive spoken sentences, including ‘elaboration’, ‘background’ and ‘restatement’ (Mann and Thompson, 1988). This has been observed, for
example, in telephone conversations (Douglas-Cowie and Cowie, 1998) and in broadcast news summaries (Wichmann, 2000).

A feature of discourse which has received much attention is the use of discourse markers (see, e.g. Brinton, 1996; Aijmer, 2002; Blakemore, 2002). As grammaticalised particles, their meaning is pragmatic rather than literal. From the perspective of Relevance Theory they are assumed to have ‘procedural’ meaning, i.e. they ‘facilitate the identification of the speaker’s meaning by narrowing the search space for inferential comprehension’ (Wilson and Wharton, 2006). From an interactional perspective they are seen as contextualisation cues: ‘not referential but indexical signs . . . cueing the context within which (an utterance) is to be interpreted’ (Couper-Kuhlen, 1996: 21).

As a phenomenon more typical of speech than writing, discourse markers such as still, anyway, well, so, now, are also of prosodic interest. There have been studies of the prosody of individual discourse markers in a number of languages, among them English: you know (Holmes, 1986), well (Bolinger, 1989; Lam, this volume), now and well (Hirschberg and Litman, 1993), anyway (Ferrara, 1997), so (Local and Walker, 2004) and Swedish: men (Horne et al., 2001).

2.2. Organising and maintaining interaction

Prosody, particularly intonation, is known to contribute to the illocutionary force of an utterance. Many studies have focused particularly on question intonation in different languages (e.g. Bartels, 1999; Grønnum and Tøndering, 2007; Vella, 2007). Both local and global parameters play a role in the identification of speech acts: yes/no questions, for example, are typically thought to display final rising contours (Hedberg et al., 2004). However, Haan et al. (1997) have shown that questions also tend to be overall higher in pitch range. There is as yet little published typological research that would shed light on the universality of such observations: some regional varieties of British English, for example, have very different patterns associated with question and statement (e.g. Wells and Peppé, 1996). Computational studies have expanded the notion of speech act to include a wider variety of functions (‘dialogue acts’) including backchannel responses, ‘agreement’ and ‘evaluation’ (Jurafsky, 2004: 588), with the aim of establishing the cues that will enable automatic identification (e.g. Shriberg et al., 1998; Jurafsky et al., 1998). The function of utterance-final intonation contours, however, is not restricted to signalling the intended speech act. They may, for example, indicate whether or not the speaker intends to continue, or, as in the case of the much maligned ‘uptalk’ (a rising contour on what is apparently a declarative statement), an attempt to elicit feedback from the listener. Final contours may therefore provide the hearer with information about how to interpret the contribution of the utterance to the discourse as a whole (see House, 2006 and this volume).

The notion of systematic turn-taking (Sacks et al., 1974) is central to the study of conversational interaction, and the prosodic correlates of turn completion, turn continuation,
as well as prosodic cues to imminent turn completion, have been widely investigated (e.g. Yngve, 1970; Cutler and Pearson, 1986; Couper-Kuhlen and Selting, 1996; Couper-Kuhlen and Ford, 2004; Szczepek Reed, 2004). Methods range from CA-type case studies to interactional-linguistic, quantitative analyses, and a number of varieties of UK English have been studied, including Tyneside (Local et al., 1986) and London Jamaican (Local et al., 1985; Walker, 2004). Discourse-functional approaches have studied the relationship between syntactic, pragmatic and prosodic completion and its implications for speaker change (Ford and Thompson, 1996; Lerner, 1996). The phenomena facilitating the taking of full turns have been studied for some time, but so far less attention has been paid to the organisation of non-floor-claiming responses, ranging from the withholding of responses via continuers to assessments and brief questions. Research on German and English has shown them to be produced at particular places in the ongoing turn, assumed to be marked by syntactic and prosodic cues (e.g. Lerner, 1996; Selting, 2000; Gardner, 2001; Kern, 2007). Barth-Weingarten (this volume) examines the placement and types of responses to extended turns-at-talk in American English.

Despite the various cues, turn-taking is not always smooth, and overlap is common, though not necessarily competitive. The prosody of interruptions, competitive and non-competitive, has been described by, e.g. French and Local (1983), Auer (1996) and Wells and Corrin (2004). In the past, these observations have been described impressionistically, in part because of the difficulty of carrying out instrumental analysis on simultaneous speech, but instrumental and quantitative studies reinforce earlier observations that speakers use pitch to indicate the nature of their interruption (see Kurtic et al., this volume). Even when the alternation of turns occurs smoothly, without overlap, this is not sufficient to construct meaningful interaction. The turns themselves must be meaningful in a given sequence: a question generally requires an answer, and a greeting is generally followed by a greeting. Which action a particular utterance is accomplishing must be somehow contextualised, and prosody can contribute to this. In some contexts, such as phone-in programmes, a greeting can simultaneously function as a request to participate, and interlocutors then have to negotiate whether to re-initiate the greeting-greeting sequence or not (see Szczepek Reed, this volume).

2.3. Style, stance and interpersonal meaning

The attitudinal function of prosody is undisputed, and there is renewed interest in how it works, due to the development of human-machine dialogue systems. This has prompted the desire to generate machine voices that sound ‘friendly’ or ‘polite’, along with the aim to recognise when human customers are irate, frustrated or just plain exasperated (Ang et al., 2002). But the nurturing of human participant relationships, or ‘rapport management’ (Spencer-Oatey, 2000), is also of interest for pragmatics in general.
One element in the maintaining of ‘rapport’ is the vocal expression of emotion. This is of interest to many, including psychologists, neuroscientists, sociolinguists, clinical linguists and speech engineers, and yet it remains difficult to establish clear phonetic correlates between recognisable emotions and speech (see Scherer, 1986; Murray and Arnott, 1993; Abelin, 2004), partly because of the infinite array of labels used. Research has therefore focused mainly on ‘primary’ emotions such as happiness, anger, sadness and fear, and these appear to be typically characterised in terms of pitch (height, range, contour). However, many of the labels used to describe affect are more suitably categorised as ‘attitudinal’ – i.e. indicating a stance or behaviour towards an interlocutor rather than the mental state of the speaker (Couper-Kuhlen, 1986: 186) – and thus fall more clearly within the field of pragmatics. The search for direct acoustic correlates of attitudes, as for emotions, continues. Ito (2003) claims, for example, that a breathy voice quality is an indicator of politeness in Japanese. It is also possible, however, that this and many other perceived attitudes are in fact prosodic inferences that have been generated indirectly. For example, Scherer et al. (1984) and Ladd et al. (1986) show in two early examples of research in this field that attitudes are not inherent to any particular contours or other phonetic patterning. Rather, ‘the pragmatic effect of an utterance depends on its immediate context as well as its suprasegmental structure’ (Ladd et al., 1986: 128) and judgements such as agreeable, polite, courteous cannot reside in the prosody alone, but are meanings inferred from certain juxtapositions of prosody and context (see also Culpeper et al., 2003).

In particular, it seems that any marked prosodic realisation – i.e. one that departs from a general norm – will trigger the search for an implied meaning much in the way that textual implicatures are generated. This is the basis for studies such as Couper-Kuhlen (1996), which describes the attitudinal effect of relative or absolute pitch matching across speaker turns. Relative pitch matching, or pitch concord, has been noted by Brazil (1985) and more recently by Wennerstrom (2001) and Szczepesk Reed (2006): it occurs when one speaker matches the pitch of another by using a corresponding region in their own voice range. In Gricean terms this is ‘co-operative’ behaviour, and Couper-Kuhlen suggests that it is perceived as supportive. However, a close melodic imitation, when a speaker matches another speaker’s pitch exactly (‘absolute register matching’) is perceived as mimicry, and thus as mocking the other speaker. The kind of mismatch that juxtaposes a semantically positive utterance with the semblance of negative affect is frequently used to convey irony. In a situation that clearly requires a negative response, responses such as: Oh great, That’s just what I needed, or Oh wonderful!, will convey irony, and the underlying negative meaning will be reinforced by a semblance of negative affect in the voice (e.g. low pitch, narrow range: see Tepperman et al., 2006 on Yeah right).

Interpersonal relations are not only influenced by prominent, affectively-loaded utterances. They can be equally influenced by more subtle ways of (dis-)affiliation and the production of (dis-)preferred responses to sequence-initiating actions. The notion of preference is a formal, rather than a psychological, notion: for example, the preferred response to a proposal is an
acceptance because it is typically performed without hesitation, delay and/or further elaboration (cf. Pomerantz, 1984; Cameron, 2001: 97). Dispreferred responses, by contrast, tend to be more elaborate (yes please vs. well actually I'd rather not), and may also be prosodically marked by an overall slowness (Plug, this volume). The affiliative or disaffiliative effect of prosody is also observable in the realisation of response particles or ‘continuers’ (Schegloff, 1982: 81), such as m-hm, uh huh, right, yes, OK. These are generally assumed to be supportive (hence ‘continuer’), but in the absence of melodic and rhythmic integration they can also subtly indicate the opposite (Müller, 1996). Studies such as these would suggest an almost iconic relationship between preference or affiliation and timing, but there is other evidence that suggests that this relationship may be more complex. For instance Couper-Kuhlen (this volume) challenges the assumed iconic link between relatedness (or affiliation) and timing of successive turns with the example of abrupt joins and certain types of delayed news deliveries.

Finally, an interesting area of research into spontaneous speech has focused on the prosody and function of apparent disfluencies – hesitations, filled pauses (um, uh etc.), and repair cues (Nakatani and Hirschberg, 1994). These are generally assumed to reflect language-planning difficulties on the part of the speaker. However, Clark (2004), who focuses on conversation as a jointly coordinated activity, sees them equally as hearer-oriented, and Horne (this volume) shows that they may also have a stance-marking function, including epistemic stance, suggesting doubt as to the precision of the utterance, and thus operating as a kind of hedge. In this way, disfluencies may also contribute to the expression of interpersonal meaning in interaction.

3. The Chapters

The chapters in this book represent different approaches and different foci. They are organised according to the broad themes we have discussed above.

- Signalling referential and discourse/textual meaning;
- Organising and maintaining interaction;
- Signalling style, stance and interpersonal meaning.

3.1. Signalling referential and discourse/textual meaning

The chapters in the first section examine prosodic cues to referential and discourse/textual meaning.

Blythe’s chapter, “Prosodic person reference in Murriny Patha reported interaction”, contributes to research on the relation between prosodic means and referential meaning, focusing in particular on person reference. His contribution extends the range of languages discussed to
Murriny Patha, a polysynthetic language spoken in the Northern Territory of Australia. Blythe uses CA techniques to examine person reference by means of prosodic mimicry. As he reports, recent research on the pragmatic roles of prosody within talk-in-interaction has concentrated on areas such as mimicry, affect, reported speech, repair and turn-taking, which can all have a bearing on the interpretation of referents. Speakers ‘do referring’ by means of a number of phenomena, including morphological, syntactic, pragmatic and prosodic devices, as well as gesture and gaze. Murriny Patha is a language in which the use of personal names is avoided, and in reported speech, speakers may employ prosodic features such as mimicry to signal an ‘other’ world of discourse, in order to avoid otherwise taboo referential expressions.

Calhoun’s chapter, entitled “What makes a word contrastive? Prosodic, semantic and pragmatic perspectives”, is a contribution to the work on how prosodic prominence plays a part in the signalling of the information status of constituents. It considers the power of accentual prominence in the indication of contrasts. The notion that all foci are theoretically contrastive is now widely accepted in discourse semantics theory (after Rooth, 1992). Focus is usually taken to be marked in English by a pitch accent on the focused word. Calhoun (2006) shows, however, that the position of focus should be seen as a strong constraint on the alignment of words with metrical prosodic structure. Foci want to align with the position of nuclear prominence. However, this account raises many questions about what leads to words being perceived as contrastive, in the pragmatic sense of actively contrasting with something else in the context. This chapter therefore looks at the question of what makes a word contrastive from prosodic, semantic and pragmatic perspectives. The study is based on data from the Switchboard corpus, already annotated for discourse features (see Carletta et al., 2004), with new layers of annotation added for kontrast (contextually contrastive) and prosody. Results of statistical (multiple regression) analyses show an expected strong correspondence between nuclear accents and kontrast. More interestingly, they suggest that a word is more likely to be kontrastive if it is more prominent than expected given its information status and syntactic features. In terms of pragmatic meaning, it has long been assumed that whether a word is interpreted as contrastive is related to its prosodic prominence. However, this study shows that this is mediated by all the other features of the word. Firstly, the perception of prominence itself is dependent on metrical structure. Secondly, the interpretation of the prominence of a word depends on the other properties of the word, including its length and syntactic status. Lastly, it depends on the plausibility of a contrastive interpretation in the context.

In their chapter “Mapping prosody and syntax as discourse strategies: How Basic Discourse Units vary across genres”, Degand and Simon focus on the relation between syntactic structure (using the framework of dependency syntax) and prosodic structure (the chunking of speech into prosodic units). Examining the speech of four discourse genres (political address, radio news, conference talk and conversational narration), they find considerable variability depending on genre. Degand and Simon describe four types of what they term Basic Discourse Units (BDU): congruent
(one-to-one mapping), syntax-bound (one Dependency Clause divided into several Major Prosodic Units), prosody-bound (one Major Prosodic Unit incorporating several Dependency Clauses) and regulatory (one discourse marker or adjunct with prosodic autonomy). They propose that each type of BDU represents a discourse strategy, and suggest that this segmentation method should be used in the analysis of discourse. They claim that there are minimal constraints but that speakers can freely choose how to group the syntactic constituents so as to fulfil their pragmatic goals, and that these strategies depend on the degree of formality or speaking style.

Lam’s chapter “What a difference the prosody makes: The role of prosody in the study of discourse particles” forms a link with previous research on the role of prosody in the study of discourse particles. These have the function of offering hearers a cue to interpreting the utterance that follows, or, in the terminology of Relevance Theory, they have procedural meaning (see Section 3.2). However, discourse markers are derived from lexical expressions through the process of grammaticalisation, and often co-exist synchronically with their fully lexical counterparts. This is a problem for automatic speech recognition. Based on data drawn from the Hong Kong Corpus of Spoken English (HKCSE prosodic) and corresponding annotations carried out in the discourse intonation framework of Brazil (1985, 1997), Lam examines the realisation of a large number of occurrences of well. She shows that intonation serves as an important cue to the disambiguation of discourse particles from their non-grammaticalised counterparts. Studying the different prosodic realisations of well as a discourse marker, she shows that prosodic independence, prosodic phrasing and accentuation all play a part in signalling the different functions of well.

3.2. Interaction management

The second section of this volume is devoted to the role played by prosody in the negotiation of speaker change in conversational interaction.

In the first chapter, “Prosody and context selection: A procedural approach”, House addresses turn-taking from a Relevance-theoretic perspective. She discusses how speakers’ organisation of utterances into coherent intonation phrases, the distribution of high and low pitch accents, the creation of a rising or falling pitch contour and the use of a wide or narrow pitch span, all play a role in guiding the hearer to the most relevant interpretation of an utterance. The hearers’ task, on the other hand, involves a combination of linguistic decoding and inferencing procedures to derive a set of hypotheses about the explicit and implicit content of the utterance, from which they will recover the interpretation that yields optimal relevance for minimum processing effort. It has been claimed (House, 2006) that intonational form cuts down on processing effort by imposing constraints on inferencing procedures, and that procedural instructions, as distinct from conceptual meanings, may be encoded in aspects of the intonational form itself. In the framework of Relevance Theory (Sperber and Wilson, 1985/1995), the notion of procedural coding has been
applied to discourse connectives such as but or however, which similarly constrain the inferencing process (Blakemore, 1987, 2002). In dialogue, the cognitive environment, or context, shared by participants is continuously negotiated and updated. House’s close analysis of map task data reveals the contribution of prosodic properties to the construction through inferencing procedures of the shared cognitive context. The use of turn-final rising or falling tones is particularly suggestive as a marker of the status of the linguistic material. Observed systematic regularities over turn sequences could be thought of as prosodic characteristics of the specific discourse type; however, House argues that construction of these discourse patterns is a by-product of the turn-by-turn updating of the cognitive environment.

In the following chapter, “When to say something – Some observations on prosodic-phonetic cues to the placement and kinds of responses in multi-unit turns”, the focus is not on full turns, but on non-floor claiming and often minimal responses to extended turns. Barth-Weingarten considers the range of prosodic-phonetic cues to the placement and types of responses to extended turns-at-talk in American English telephone conversation. Four detailed case-studies illustrate her findings: based on data from the CallHome corpus and using auditory and acoustic methods, she has established that – in the vicinity of points of (preliminary) syntactic-pragmatic completion – not only certain phenomena of pitch but also of loudness, breathing and voice quality cluster to indicate the relevance of response, while others form a response-blocking set of features. Moreover, the type of response made relevant appears to be regularly indicated by the degree of pragmatic completeness in correlation with a certain, possibly iconic, extension of the prosodic-phonetic features at hand. On the basis of her analyses of the placement of response tokens and the prosodic-phonetic features preceding them in their sequential context, Barth-Weingarten shows that the organisation of responses to extended turns is apparently systematic in a similar way to turn-taking. Cues to turn-taking and response-giving depend on a close interrelationship between syntax, lexis, pragmatics and prosody-phonetics.

Prosody plays a role not only in the achievement of smooth speaker change, but is also relevant to the negotiation of competing turns and overlapping speech. In their chapter “Fundamental frequency height as a resource for the management of overlap in talk-in-interaction”, Kurtic, Brown and Wells report a study of how participants in multi-party conversations utilise prosody to manage simultaneous talk. On the basis of previous research, they assume that the function of overlapping speech in conversations can be twofold: on the one hand, overlap can provide a means of competing for a turn. On the other hand, it can occur non-competitively as a by-product of turn-taking in conversation, or emerge from the collaboration between speakers in turn-taking. They examine whether there are differences in how speakers use and orient to prosody in turn-competitive overlaps, as opposed to non-competitive ones. Their approach combines detailed CA-based analyses of overlapping speech sequences with acoustic measurements of prosodic parameters at the overlapping sequence in its local context. The analyses are carried out on a set of overlaps extracted from the ICSI Meeting Corpus (Janin et al., 2003), which contains transcribed
recordings of face-to-face, multi-party research meetings. Sequences of overlapping talk were identified in the data and annotated as competitive or non-competitive, and these were then subjected to an instrumental analysis of prosodic events during the overlap (fundamental frequency, intensity and speech rate). Their findings support the observations made in previous studies, which are based on impressionistic data analysis (e.g. French and Local, 1983), showing that participants in talk-in-interaction systematically manipulate pitch height when competing for a turn.

Szczepk Reed's contribution "FIRST or SECOND: Establishing sequential roles in radio phone-in programs through prosody" examines the way that turns at talk are contextualised, basing her analysis on turns in the opening sequences of radio phone-in programmes. In most two-party telephone conversations, the roles of caller and called are given, which allows for the respective sequential pair parts of first and second (summons-response followed by greeting-greeting) to be distributed without negotiation. Other conversational environments are less conducive to such clear sequential distribution of turns. One such environment is that of opening sequences in radio phone-in programs. Opening sequences on the radio are more varied in form than non-broadcast calls, as the primary activities such as summons/answer and self- and other-identification are initially accomplished between the caller and studio personnel off the air, previous to the encounter between caller and radio presenter. This allows the radio host, who always initiates the interaction with the caller, a small choice of first conversational activities, while certain obligatory activity constraints, such as initiation of contact, caller-identification (ten Have, 1999) and selection of caller as next speaker, still hold. Radio presenters typically have their own individual formats for initiating interactions with callers, from caller name and residence only, to a potential routine including caller name and residence identification, selection of caller as next speaker (you're on), radio station identification, host self-identification, greeting, how-are-you and thank-you-for-the-call. While one may assume that the role of the host's turn as initiating the interaction and selecting the next speaker determines its sequential status as a first, the data in this chapter show that in spite of these overriding conversational activities the local sequential status of the host's turn as first or second is negotiable. In this negotiation the participants' use of prosody and their awareness of each other's prosodic patterns play a crucial role.

3.3. Signalling style, stance and interpersonal meaning

The final section of the volume focuses on various aspects of interpersonal meaning and how they are conveyed.

Plug's contribution "On tempo in dispreferred turns: A recurrent pattern in a Dutch corpus" is concerned with indicating the type of turn sequences from the point of view of preference organisation. He reports on the temporal characteristics of turns which implement a dispreferred action: in particular, turns which express disagreement, and turns which provide a problematic
answer to a prior question. Previous research has suggested that one recurrent feature of dispreferred turn types is that they are delayed with respect to the prior turn (Sacks et al., 1974; Pomerantz, 1984) – that is, dispreferred turns are characterised by a slow speaker transition. A relevant question is whether this slowness persists throughout the turn, or whether 'normal' tempo is resumed following the delayed start. Observations by Plug (2005) on dispreferred turns with the discourse marker eigenlijk 'actually' in Dutch suggest the former. This chapter presents a more extensive study of a collection of dispreferred turns taken from a corpus of Dutch conversation, and pays special attention to the relationship between the tempo of the turn component which implements the dispreferred action and that of the prior turn. The study confirms that in a majority of cases, the dispreferred turn component is produced at a lower tempo than that of the turn to which it responds. This suggests that the dispreferred turn types considered are not only characterised by delay, but also by slowness.

Couper-Kuhlen's chapter "Relatedness and timing in talk-in-interaction" considers the role of timing in interaction from a wider perspective. Connecting back to a number of issues raised in the previous contributions, she builds on a number of conversation-analytic findings concerning the timing of turn-constructual units and turns at talk and reflects upon their consequences for an appreciation of relatedness in interaction. There is, first, a sequence-organisational aspect to the timing of successive turns at talk in that contiguity, or close timing, is found within action sequences (Sacks, 1995), but gapping is often encountered at sequence and topic closure (Couper-Kuhlen, 1993, 2004). Second, there is a preference-related aspect to the timing of turns, with structurally preferred responses exhibiting well-timedness in relation to the prior, and structurally dispreferred ones exhibiting delays (Sacks, [1973]/1987; Pomerantz, 1984; Plug, this volume). Prima facie observation in both cases suggests a relatively iconic relation between the relatedness (or affiliation) of an action to a prior and the temporal proximity holding between them: if two actions are (preferentially) related, they tend to come close together in time. However, there are some more or less acknowledged exceptions to this regularity. For instance, there are abrupt-joins, where two turn-constructual units are latched, although it is their unrelatedness which is at issue (Local and Walker, 2004). And there are preferred responses to certain types of news delivery which are delayed rather than well-timed (Couper-Kuhlen, 1993; Wilkinson and Kitzinger, 2006). Couper-Kuhlen argues that such findings force us to rethink the putative iconic link between timing and relatedness in talk-in-interaction.

Horne's chapter, "Creaky fillers and speaker attitude: Data from Swedish", addresses topics in the area of apparent disfluencies in spontaneous speech and their hedging function. Horne focuses on the important discourse role of filled pauses, particularly with a marked voice quality. Filled pauses have most often been characterised as 'disfluencies' due to the fact that they constitute delays in the flow of speech associated with referential meaning (Shriberg, 2001). However, filled pauses can also have important pragmatic functions. Apart from contextualising dispreference, these include signalling an upcoming focused word (Bruce, 1998) in spoken discourse, or the
need on the part of the speaker to plan or code their speech and, at the same time, a desire to hold
the floor (see, e.g. Allwood, 1994). Indeed, Allwood, and Clark and Fox Tree (2002) refer to filled
pauses as ‘words’ since they can be associated with quite specific, albeit pragmatic meanings.
Using spontaneous data from Swedish, Home’s findings suggest that ‘creaky voice’ associated
with the filler EH ‘UH’ is related to the realisation of the speaker attitude ‘uncertainty’, in the
sense of ‘indecisiveness’. This conclusion is based on the assumption of the existence of an
‘accessibility hierarchy’ (Ariel, 1988), which constrains the way information is coded in relation
to the assumed background knowledge of conversational partners. Here, the speaker appears to be
uncertain as to the most appropriate or relevant way in which to encode the content of a
proposition.

4. Final Remarks

The present volume is a response to the growing interest in the prosody-pragmatics interface
observed in recent years (cf. Wichmann and Blakemore, 2006). It presents a collection of chapters
exploring a wide range of issues from both fields of research and the interrelatedness of prosody
and pragmatics. At the same time it demonstrates the variety of approaches that can be employed
in this undertaking as well as the diversity of databases available for analysis. Last but not least,
each individual chapter shows the richness of what happens when two disciplines come together.
It is at these points of intersection that new insights often emerge, and we hope the book will
encourage the spirit of collaborative research and also offer a glimpse of the possible rewards to
be gained from stepping beyond one’s original field and adopting a new perspective.

References

Allerton, D. J. (1978). “The notion of ‘givenness’ and its relation to presupposition and theme”. Lingua 44:
133–168.
grammatik och språkförändring. Lund: Department of Nordic Languages, Lund University, 3–13.
of annoyance and frustration in human-computer dialog”, Proceedings of ICSLP-2002, Denver, CO,
2037–2040.


