1

Towards an interactional perspective on prosody and a prosodic perspective on interaction

ELIZABETH COUPER-KUHLEN
AND MARGRET SELTING

1 Introduction

If prosody is understood to comprise the 'musical' attributes of speech[1] – auditory effects such as melody, dynamics, rhythm, tempo and pause – then it is surely no exaggeration to state that a large part of this field has been left untilled by modern structural linguistics.[2] Only a few scholars at the most have considered prosody, intonation in particular, worthy of their attention. Moreover, when they have done so, it has been primarily in structuralist terms. Speech melodies and rhythms have been pursued as a part of language competence, analysed in minimal pairs as if they were phoneme- or morpheme-like entities with distinctive functions. More socially oriented approaches to the study of language, by contrast, have openly acknowledged the importance of prosodic phenomena in language-in-use but have not put their words into practice or have done so in an intuitive, pre-theoretical way which does not reflect the state of current prosodic research. In this chapter we shall argue that bringing together these two fields of inquiry, speech prosody and language-in-use, and allowing them to cross-fertilize will help to overcome a number of the weaknesses which have become apparent in the current praxis of each.

In retrospect, it is doubtless the overwhelming influence of literacy on thinking about language which has been responsible for the neglect of prosody.[3] Language, perhaps unconsciously, has been equated with 'prose', those linguistic forms put down in writing and intended to be read. Given that prosodic phenomena (i) are not segment-based, referential units, (ii) are gradient rather than

discrete, and (iii) lack systematic codification in writing, it is perhaps not surprising that they are so often ignored.[4] Without such features, prosody becomes one of those aspects of language which, in Silverstein's words, are at the 'limits of awareness' (1976). To be sure, there have been extenuating circumstances: in the early part of this century, when the structuralist model was being developed, spoken language – and with it, prosody – was difficult to 'capture' and make permanent for the purpose of scientific enquiry.[5] However, with the advent of the tape recorder and, more recently, the video camera, the stipulation of permanency can now be met. Auditory and video records are easily obtainable of events in which the spoken language is used. Too many linguists, however, still persist in approaching these 'speech events' as if, like Molière's *bourgeois gentilhomme*, the interlocutors were speaking prose.

David Abercrombie once wrote that 'what in fact linguistics has concerned itself with, up to now, has almost exclusively been – spoken prose' (1965:4).[6] Yet, he points out, spoken prose is an entirely different mode of language use compared, for instance, to conversation – in particular with respect to intonation:

> if you are reading aloud a piece of written prose, you infer from the text what intonations you ought to use, even if, as is almost always the case, you have a choice. The intonation, in other words, adds little information. But if you try to read aloud a piece of written conversation, you can't tell what the intonations should be – or rather what they actually were. Here the intonations contribute more independently to the meaning. (1965:6)

Abercrombie claims that spoken prose has more highly standardized intonation patterns, more evenness of tempo and its pauses are more closely related to grammatical structure. Thus, he concludes, a prose-based approach to intonation is of little use for the analysis of conversation.[7] Yet 'spoken prose' is still an apt characterization of what many modern studies of intonation and prosody end up examining.[8]

The study of conversation *sui generis* has nonetheless made considerable progress since Abercrombie wrote, above all due to the efforts of a small group of American sociologists centred around Harvey Sacks. For Sacks conversation was simply one locus where social order could be discovered:

> conversation analysis . . . seeks to describe methods persons use in doing social life. It is our claim that, although the range of activities this domain describes may be as yet unknown, the mode of description, the way it is cast, is intrinsically stable. (1984:25)

Sacks and his co-workers advocate a special 'analytic mentality' for the discovery of social order in conversation. In their work they emphasize (i) the importance of investigating naturally occurring data, (ii) a view of social interaction as an ongoing, sequentially organized and collaboratively achieved process, and (iii) the necessity for justifying one's analyses by showing the relevance of the categories postulated to the participants themselves. Conversation analysts have called attention to a whole host of phenomena which are often ignored in the study of discourse: hesitation particles, sound stretching, cut-offs, in-breaths, laughter, (micro-)pausing, stressing and intonation. Nominally then, prosody has fared much better in this approach to the study of language as a social phenomenon. Yet, although prosodic events are recorded in the transcripts of interaction, they rarely figure – with the possible exception of pause – in the analyses which conversation analysts have so far offered.

One notable exception to the skirting of prosody by social scientists is to be found in the work of Gumperz (1982, 1992).[9] Gumperz has claimed that participants in verbal interaction employ 'empirically detectable signs' which *cue* conversational interpretation by evoking interpretative schemata or frames. Members of a speech community, he argues, consciously or unconsciously appeal to these frames in drawing inferences about what is being said in interaction or more generally about what is 'going on'. The process whereby participants 'construct' context via such cues in order to make utterances interpretable has come to be known as *contextualization* (see also Auer and di Luzio 1992). Foremost among the means which speakers use to contextualize language are prosodic features; others include code-switching and non-verbal elements such as body position, gesture and gaze (Auer 1992). Gumperz (1992) argues that contextualization cues affect interactants' inferencing at three levels of generalization: (i) conversational management, e.g. in turn-taking and signalling degrees of informational relevance; (ii) sequencing, e.g. in cueing implicatures and disambiguating speaker intent in utterances; (iii) framing, i.e. in generating

expectations about the nature of the interaction and in keying its mood or atmosphere.

The interactional approach to prosody which we advocate here is in effect inspired by the theoretical insights of contextualization theory coupled with the methodological procedures of conversation analysis. However, before filling in the details, we shall discuss some of the problems, as we see them, in the field of prosodic and especially intonation research as currently practised (§2). We shall then outline an interactional perspective, showing how it can resolve some of these difficulties (§3). In the last section of the chapter (§4), we shall argue that not only does prosodic analysis stand to gain from an interactional perspective but that a proper consideration of prosody will also enrich conversational research.

2 Current problems in prosodic research

There are three central questions which lack satisfactory answers in traditional prosodic work. They concern formal, functional and methodological issues respectively.

2.1 Prosodic categories: phonetic or phonological?

Much of prosodic, and especially intonation, research has centred around the concept of the *tone group* or *tone unit.*[10] (We shall use these terms interchangeably.) In this tradition a tone group is defined as a phonological unit which has *one* prominent pitch movement beginning on an accented syllable and optionally extending over any following unaccented syllables.[11] This accented syllable is called the *tonic syllable* or the *nucleus* of the tone group or unit; the pitch movement selected here is known as a *tone* or *nuclear tone*. When there are several accented syllables in a unit, the nucleus is typically the last, most prominent one. The boundaries between phonological tone units are said to be expounded by phonetic cues such as a pause, final syllable lengthening, a rhythmic break and/or a pitch upstep or downstep in unstressed syllables. More recent prosodic models (e.g. Pierrehumbert 1980, Selkirk 1984, Nespor and Vogel 1986) share many of these premises but call the tone group an *intonational phrase* and refer to the pitch movements on prominent syllables as *pitch accents*. Despite acknowledged prob-

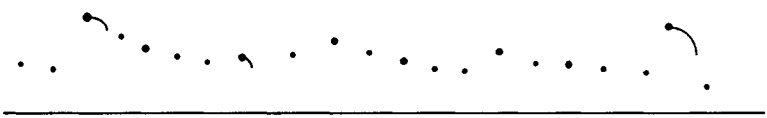
lems in delimiting tone groups and identifying tone-group boundaries (see Brown, Currie and Kenworthy 1980), much recent research has addressed questions concerning the identification and description of nuclei and their functional interpretation (e.g. Brazil, Coulthard and Johns 1980, Ladd 1980, Gussenhoven 1983; for overviews see Couper-Kuhlen 1986, Cruttenden 1986).

The traditional tonetic approach runs into trouble, however, when phonetic and phonological criteria conflict in the identification of tone groups. For although phonetically oriented prosodists have tried hard to narrow the gap between phonology and phonetics, the tone group remains a phonological unit which is not necessarily delineated by phonetic boundary cues. Even though there may be no phonetic evidence of a tone-unit boundary, the analysis of a prominent syllable carrying one of the nuclear tones forces the postulation of a boundary. Similarly, the presence of several nuclear tones in succession usually imposes a tone-unit boundary between them,[12] although there may be no phonetic cues signalling it (see also Selting 1993). In cases like these intonologists will tend to fall back on syntax when pressed to locate the boundary. For instance, assuming pitch accents on *Wales* and *Cardiff* in the sentence under (1), the boundary is likely to be drawn between *Wales* and *is* rather than between *is* and *visiting* or between *visiting* and *Cardiff*, on syntactic grounds:

(1) The Prince of Wàles // is visiting Càrdiff tomorrow
(Cruttenden 1986:38)

Yet such a procedure shows greater affinity with the analysis of a syntactic entity 'S' into units of constituent structure than with 'integral melodic patterns' (Bolinger 1978:474) which are used for communicative and interactional purposes.

If 'zero' realization of tone-unit boundaries creates a problem for intonational phonology, so pauses and hesitation render the phonetics of tone-unit analysis difficult. Consider, for instance, the utterance recorded in Figure 1.1 from a San Francisco radio phone-in programme during the war in the Persian Gulf. Applying the phonetic criteria for tone-unit delimitation mentioned above, we obtain relatively clear boundaries following *down*, *I did*, *today* and *upsets me*.[13] Yet should *and uh*, clearly separated from following talk by a rhythmic break and a pause, be considered part



it was real:ly nice to go down: and be with people who felt the same way that I did (0.4)

and uh (0.4) what I'm seeing today: really upsets me hhh

(Small dots represent unstressed syllables, large dots represent stressed syllables with no audible pitch movement, tadpole-like tails represent pitch glides.)

Figure 1.1

of the tone group *what I'm seeing today*? If not, is it a tone group of its own with, say, a level nucleus on *and*? Similar problems arise when speakers lengthen a pre-pausal segment as a holding device and then resume the pitch and loudness of this segment after the pause. Should these segments be thought of as belonging to two separate tone units although the speaker clearly signals the continuation of one melodically cohesive unit?

The conflict between phonetic and phonological criteria in tone-group analysis can be avoided if prosodic phrasing units are identified in an interactional perspective. To see this, let us draw a parallel with lexico-syntactic structure in language. The units of traditional linguistic analysis are grammatical entities such as sentence, clause, phrase, word. Yet taking a discourse perspective on language, we speak of 'acts' or 'moves' whose basic units are *utterances*. These are units which derive their specific characteristics from their occurrence in turns. By analogy, the basic *prosodic phrase* in speech, when viewed interactively, is likely to be *not* the prosodic counterpart of a grammatical sentence or clause, but rather a unit defined with respect to the utterance as a turn-constructional unit, a 'phonetic chunk' which speakers use to constitute and articulate turns-at-talk. Utterances are subject to con-

straints which derive from properties inherent in interaction: the exchange of (and competition for) turns-at-talk, changing deixis, emergent and negotiable meaning.[14] The same sorts of constraint are likely to apply to the basic prosodic phrase in verbal interaction. That is, its shape will derive from the interactive need of participants to project turn completion. It will be sensitive to interactive requirements such as recipient design and local fit. And it will be subject to repair. Ultimately, interactionally grounded categories for prosodic and intonational analysis may correspond only loosely – or not at all – to traditional tonetic categories.[15]

2.2 *Prosodic function: distinctive or not?*

Henry Sweet, one of the pioneers of modern phonetics, claimed as early as 1877, that, in intonation languages like English, *intonations* or *tones* such as *level, rising, falling, compound rising* and *compound falling*, are functionally *sentence-tones*: 'In all these cases the tones . . . modify the general meaning of the whole sentence' (1877:95). Elsewhere, the sentence-tones are explicitly linked to sentence-types: 'The interesting question now arises, how do such languages express these general ideas (interrogation, affirmation, etc.), which it is the function of the English tones to express?' (*Collected Papers of Henry Sweet*, cited in Henderson 1971:179). Thus, in one of the oldest traditions of British prosodic analysis, the tone group is thought of as corresponding to the grammatical unit of sentence and the primary tones, fall vs rise, as corresponding to the grammatical sentence-types 'affirmative' (*statement*) vs 'interrogative' (*question*).

The grammatical approach to intonation reaches its apex in Armstrong and Ward (1926), where the melodies of English sentences are reduced to two (Fig. 1.2). Tune I (with final falling pitch) is said to be used in 'ordinary' statements, in questions requiring an answer other than 'yes' or 'no', in commands and in exclamations (1926:8ff.), while Tune II (with final rising pitch) is said to be called for in 'indefinite' statements, questions requiring the answer 'yes' or 'no' and requests (1926:19ff.).[16] The procedure in this and other grammar-based approaches to intonation is thus to associate tunes with different grammatical sentence-types and to treat excep-

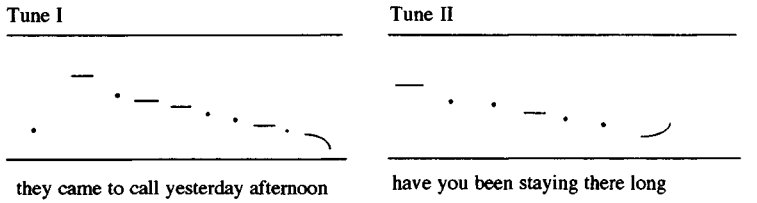


Figure 1.2

tions as conveying attitudinal modifications such as 'indifferent', 'polite', 'surprised', etc.

Some intonologists have preferred to segment melodic tunes like those Armstrong and Ward identify into their component parts, a procedure inspired by phonemic analysis. Here the pitch of each of the small constituent parts of a melodic whole is said to create a distinctive meaning opposition. Pike (1945), for instance, postulates four pitch phonemes for English intonation: extra-high (1), high (2), mid (3) and low (4), which articulate the intonation contour (1945:26). The following minimal pairs illustrate:

(2) (A raised degree mark precedes the first stressed syllable of the contour.)

a He's gone
3- °2-4

b He's gone
3- °2-3

c He's gone
3- °1-4

d He's gone
3- °3-4

In terminal position a falling contour to pitch level 4 as in (a) is claimed to have the meaning of finality, whereas one which falls to pitch level 3 (b) is said to be incomplete and to suggest an addition such as 'but he's coming back' (1945:45). A falling terminal contour from pitch level 1 (c) is claimed to be more intensely contrastive, often insistent, and may have an element of surprise, as compared to a falling terminal contour from level 3 (d), which is said to be much milder and detached (1945:47f.).

Halliday (1967, 1970) is also at pains to demonstrate the distinctiveness of intonational features, although he does not 'segment' intonation to quite the extreme of the American structuralists. He argues, for instance, that presence or absence of a tone-group boundary will change a sentence with one 'information unit' as in (3a) below into a sentence with two 'information units' as in (3b):[17]

(3) a //he washed and brushed his hair//
b //he washed//and brushed his hair//

(adapted from Halliday 1967:36)

Furthermore, according to Halliday, the variable location of tonic prominence will result in different 'information foci', as in:[18]

(4) a //I saw him//
b //I saw him//

(adapted from Halliday 1967:38)

Finally, he states that the substitution of, for instance, rising for falling pitch on a tonic syllable will change a statement as in (5a) into a question (5b):

(5) a //John hàs//
b //John hás//

(adapted from Halliday 1967:41)

The claim is that because such choices are systematically and distinctively related to meaning, intonation is *grammatical* in function. However, as any linguist who has looked seriously at naturally occurring data will know, the problem is that intonational patterns – in contrast to morphemes, with which they have been compared – do not always carry this kind of functional load in language use; in fact they do so so rarely that to treat them as 'intonemes' is arguably rather meaningless.

Two specific difficulties compound the problem of identifying intonational functions based on distinctive oppositions. First, *covarying lexico-syntax*: with very few exceptions, it is hard to prove distinctive function for intonational features because it is nearly impossible to control all variables but one and still have plausible examples. This is why O'Connor and Arnold (1973), for instance, must match their prototypical tone groups with different (non-overlapping) sets of sentences. Compare, for example:

(6) THE SWITCHBACK (unemphatic)
You can ˇ try.

(7) THE JACKKNIFE (unemphatic)
It's riˊˋdiculous.
(adapted from O'Connor and Arnold 1973:42)

When the Switchback is now glossed as 'grudgingly admitting' (170) and the Jackknife as 'censorious' (214), O'Connor and Arnold are left open to the criticism that it is the lexico-syntactic content, and not the intonation, of their sentence+tone-group combinations which invites each interpretation. If lexical content covaries with intonational tune, then in strict structuralist terms, minimal pairs are not possible and no distinctive function can be established.

Second, *iconicity of intonation*: this difficulty can be put in a nutshell by Bolinger's observation that 'intonation has more in common with gesture than with grammar' (1986:xiii). Although an elaborate argument can be constructed to show that a gesture which originates iconically (i.e. is non-arbitrary) can nevertheless become part of a structural system used conventionally, the fact is that intonational iconicity is not just rudimentary and overlaid by other structurally organized features. It is quite ubiquitous and therefore constantly in danger of 'interfering' with rigorous structuralist demonstration. Consider, for instance, the following O'Connor and Arnold examples:

(8) *(It's six o'clock)* ↑Hèavens!

(9) *(You've passed your exam)* What wonderful ↑nèws!
(adapted from O'Connor and Arnold 1973:57, 75)

In cases like these, the high and widely falling tone (represented here as ↑ˋ) is said to convey an emotive element of *surprise*. Yet what is the structural value of high pitch and wide pitch movement as conveyors of surprise when the level of physiological arousal associated with this state quasi-automatically produces a sudden fluctuation in vocal-fold vibration, and consequently in pitch?

In sum, there appears to be, on the one hand, little or no constancy between intonational form and meaning: in the worse case, the same tune 'means' something different with each different

lexico-syntactic carrier. On the other hand, there is no final proof, on a more general level, that pitch ultimately functions by way of nurture (cultural convention) and not by way of nature.[19] As long as intonologists pursue a grammar-based, structuralistic approach, there is little hope of solving such problems.

An interactional perspective, however, offers a way out. Here intonation is linked up to functions which derive from the situated use of language to accomplish interactional goals. Placed in its context of use, language 'means', i.e. suggests interpretations, through a complex interaction of verbal forms with contextual and situational factors. Arguably, the discourse functions of intonation are more likely to relate to this kind of *pragmatic* 'meaning' (situated, inference-based interpretation) rather than to the semantic meanings of decontextualized linguistic forms.[20] In an interactional perspective, analysts are consequently not looking for minimal pairs and distinctive functions. Instead they typically find that intonation and prosody have a *contextualizing* function (Gumperz 1982, Auer 1986, Auer and di Luzio 1992). This means that they cue frames for the situated interpretation of utterances. In other words, they constitute *how* something is said, not what is said, and they ultimately influence only what participants *infer* is the meaning. Prosodic contextualization cues are not referential, but indexical, signs (Silverstein 1992). They stand in a reflexive relationship to language, cueing the context within which it is to be interpreted and at the same time constituting that context. Their iconicity facilitates (rather than interfering with) this function (see Auer 1992). In this sense, the 'problems' referred to above (covarying lexico-syntax, iconicity) are not problems at all for intonational function in an interactional perspective; they are rather merely reflections of the nature of this function.

2.3 *Prosodic methodology: what data? what proof?*

'Discourse' intonologists have been somewhat more successful in avoiding the lexico-syntactic trap described above. Brazil (1981, 1982), for instance, postulates two distinctive tones for English discourse, a 'Referring' tone, which falls and then rises, and a 'Proclaiming' tone, which simply falls. Compare, for example:

(10) a when I've finished \/MIDDlemarch // I shall read Adam \BEde //
b when I've finished \MIDDlemarch // I shall read Adam \/BEde //
(Brazil, Coulthard and Johns 1980:13)

The falling–rising tone in both these examples is said to *refer* (*r*) to a piece of information as part of the common ground, the falling tone, to *proclaim* (*p*) a piece of information as new. In addition, three 'keys' are postulated – High, Mid and Low – which apply to the tones and describe their relative pitch height within the speaker's voice range. Compare:

(11)

a *high* eventually
mid // *r* // *p* we gave it our neighbours // *p*
low the Robinsons//

b *high* eventually Robinsons//
mid // *r* // *p* we gave it our neighbours // *p* the
low

(Brazil, Coulthard and Johns 1980:30)

Low key as in (11a) is said to mark an *equivalence* relation between items in successive tone groups, i.e. here between *neighbours* and *Robinsons*. High key (11b) marks a *contrast* between two items.[21] Thus, both tone and key are claimed to be responsible for (minimally) distinctive oppositions in English discourse. However, the crucial point here is that they cannot be predicted from lexico-syntax: 'Tone choice, we have argued, is not dependent on linguistic features of the message, but rather on the speaker's assessment of the relationship between the message and the audience' (Brazil, Coulthard and Johns 1980:18). The model employed by Brazil, Coulthard and Johns thus has the salutary effect of uncoupling intonation from lexico-syntax. Once intonation is seen as a quasi-independent signalling system, it becomes a powerful means of creating interactional meaning, in alignment or non-alignment with verbal forms.

Yet the analytic interpretations which this model offers have a methodological weakness: they are warranted only by the analyst's own intuitions (see also Local, Kelly and Wells 1986). Although Brazil's theory is based on data from classroom interaction, we are given no independent evidence that interactants do indeed interpret, for instance, low key as marking an equivalence relation and

high key as marking a contrastive one. Instead, Brazil, Coulthard and Johns seem to rely on our implicit recognition of the validity of these claims. Yet without some proof of speakers' 'indigenous' interpretations, the analyst risks assigning interpretative labels 'from without', labels which may lack cross-contextual and cross-linguistic validity.

The same criticism can be levelled at Bolinger (1986, 1989), who has also stressed the independence of intonation as a signalling system. Rather than seeing the major thrust of intonational meaning in relation to information, as do Brazil, Coulthard and Johns, Bolinger emphasizes its emotional contribution: 'Intonation is part of a gestural complex whose primitive and still surviving function is the signaling of emotion' (1986:195). The intonational system is in principle independent of lexico-syntax: 'even when it interacts with such highly conventionalized areas as morphology and syntax, intonation manages to do what it does by continuing to be what it is, primarily a symptom of how we feel about what we say, or how we feel *when* we say' (1989:1).

Bolinger's 'primary profiles' – *A*, a relatively high pitch on an accented syllable followed by a jump down, and *B*, a pitch jump up to or up from an accented syllable – have abstract meanings that ultimately derive from the metaphors of falling or 'down' vs rising or 'up'. In context they take on more specific meaning by undergoing metaphorical transformations. For instance, 'down' may come to mean termination or demarcation for separate importance. 'Up' extends to meanings of incompletion and 'keyed-upness' (1986:341).

Yet here too we are given no evidence that actual participants in verbal interaction do indeed orient to 'down' as being associated with rest and relaxation and 'up' as being associated with tenseness and effort. As intuitively pleasing as Bolinger's interpretations may be, the final proof has yet to come that they are relevant in this way to interactants in real conversations. Moreover, the data upon which Bolinger's analysis is based are for the most part constructed by himself. While there can be little objection to the sporadic use of artificial examples for the purpose of argumentation, to build one's intonational theory on the basis of these alone is rather risky. As those who have worked with genuine data will know, reality often bears more surprises than fiction. From an interactional perspective,

only genuine real-time communication can provide context-rich data. As we shall demonstrate below, it is participants' own handling of prosodic cues within this context that enables empirical proof procedures for the validation of analytic categories. In other words, the theory relies crucially on the way prosody is deployed in real interaction.

3 Prosody in an interactional perspective

Our argument so far has been that problems of the sort sketched above have prevented a breakthrough in the study of prosody in discourse and that they can be avoided by adopting a different perspective, one we have provisionally dubbed *interactional*. Methodologically, this perspective is inspired by the practices of conversation analysis, a sociologically grounded approach to verbal interaction which developed from the application of ethnomethodological principles (Garfinkel 1963, 1967) to the analysis of everyday conversation.[22] One of the fundamental assumptions of conversation analysis (henceforth CA) is that verbal interaction is structurally organized and that traces of this organization can be found in the interaction itself. Consequently, as Heritage puts it, 'no order of detail can be dismissed, *a priori*, as disorderly, accidental or irrelevant' (1984:241). In fact, CA studies have revealed conversational interaction to be a surprisingly orderly accomplishment, brought about by speakers who rely for its management on the local use of context-free and context-sensitive organizational systems such as turn-taking and repair.

One of the characteristics of CA work is its strictly empirical stance: hypotheses are generated from close analysis of audio- or video-recorded, naturally occurring, conversation. Like ethnomethodologists, conversation analysts attempt to reconstruct social (conversational) interaction as the collaborative endeavour of 'members' (the participants of a culture, group or other social category). They seek to describe and systematically account for the practical methods which members employ in the production and interpretation of conversational activity. These methods are viewed as means or devices designed to resolve specific interactional tasks (Levinson 1983:319). Yet the idea is not to describe conversational sequences in terms of well-formedness but to reconstruct the

unmarked expectations which underlie interlocutors' inferential processes. Deviations from expected sequences may also systematically steer inferences: 'Conversationalists are . . . not so much constrained by rules or sanctions, as caught in a web of inferences' (Levinson 1983:312).

The point that we wish to make here is that prosody can be seen as one of the orderly 'details' of interaction, a resource which interlocutors rely on to accomplish social action and as a means of steering inferential processes. Prosodic features, we suggest, can be reconstructed as *members' devices*, designed for the organization and management of talk in social interaction. They can be shown to function as part of a signalling system which – together with syntax, lexico-semantics, kinesics and other contextualization cues – is used to construct and interpret turn-constructional units and turns-at-talk (Selting 1992a, 1995). Viewed as a member's device, prosody is amenable to CA methodology. That is, some of the same techniques which have been used to expose the orderliness of conversational organization can be applied to the analysis of prosody in conversation. In the following we shall show how a selection of conversation-analytic research 'maxims' can be adapted to prosodic research.

3.1 Give priority to the analysis of naturally occurring talk

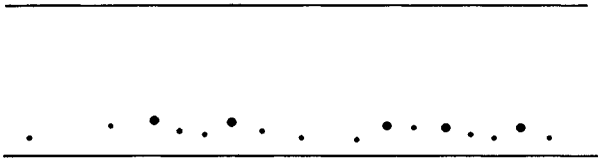
CA's rigorously empirical stance for the study of conversation transfers quite naturally to prosodic investigation. Close inspection of prosody in large quantities of everyday conversational data should make it possible to reconstruct members' prosodic devices for achieving their conversational goals. Ethnomethodological experiments of course need not be ruled out.[23] Conceivably, the controlled use of 'deviant' prosodic features by well-trained speakers in everyday conversational encounters can help to generate working hypotheses. Additional insight into members' commonsense reasoning about prosody can be gained from interviews with experimental subjects.[24]

3.2 Treat the data as an integral part of the context in which it occurs

Conversation analysts have emphasized that speaker's utterances in interaction are context-sensitive, i.e. that they are especially constructed to fit the particular location and occasion of their use. Transferred to the study of prosody, context-sensitivity means recognizing that prosodic features may be selected with regard to the particular verbal forms which 'carry' them. The prosody associated with a given interactive function, for instance, may be sensitive to whether the accompanying syntax is complete or not (Selting 1995; Auer, this volume). Or it may be sensitive to whether the carrier is a syntactic adjunct ('post-completer') (Local, Kelly and Wells 1986: 423).

Viewing prosody context-sensitively also means paying attention to the fact that its carrier is a turn at talk (or part of one) which itself has a sequential location. Just as Schegloff states that a syntax-for-conversation must 'recognize that its sentences will be in turns and will be subject to the organization of turns and their exigencies' (1979a:281), so a prosody-for-conversation must recognize that its basic unit – whether expounded by intonational, rhythmic, pausal or dynamic means – co-occurs with, and will be interpreted in relation to, a turn-constructional unit. Similar to turn-constructional units, which may be situated parts of larger 'projects', prosodic units will be sensitive to their location in a series. And similar to turns, which occur in sequence, prosodic units can be expected to demonstrate sensitivity to their sequential location (cf. Müller, this volume). As Schegloff concludes for syntax: 'all the types and orders of organization that operate in and on turns in conversation can operate on the sentence' (1979a:282). Substituting 'basic prosodic unit' for 'sentence', the same applies to prosody: all the types and orders of organization that operate in and on turns in conversation can operate on the basic prosodic unit.

By way of illustration, consider prosody in the *decontextualized* English utterance in Figure 1.3. Considered out of sequential context, the combination of low and suppressed pitch range, low volume and relatively fast tempo plus a carrier which refers to an event as *upsetting* might evoke a functional interpretation of, say, 'repressed concern'. Nor is this interpretation necessarily wholly



<uh (tʃi:) that guy I just listened (to)> that really really upsets me <p, acc>

(Deviations from normal volume are marked in angled brackets as *ff, f, p* or *pp*; deviations from moderate tempo are indicated with *acc* or *dec*; *high* or *low* to the left of the interlinear transcription indicates that the whole intonational clef is at an extreme in the speaker's voice range.)

Figure 1.3

wrong. Yet considered in its actual sequential location, the prosody appears to be doing more:

(12) *A San Francisco radio phone-in programme during the Persian Gulf war; callers have been encouraged to make their opinions known and share their feelings about the war. There has just been an on-the-spot report from an anti-war demonstration, the reporter having described how he is momentarily boxed in by angry demonstrators and policemen about to use tear gas.*

(This fragment is transcribed in conventional CA notation, with the addition of /;/ for terminal falling pitch ending at a mid level in the speaker's voice range.)

L: we continue to take calls=Debbie on the line from San Jose, you're on the Giant 68 KNBR. I'm Leo Laporte.
D: hi Leo.
L: hi Debbie.
→ D: uhm.(0.3) uh [tʃi:] *that guy I just listened (to) that really really upsets me.*
L: [°why°]
D: [uhm;] (0.4) well the reason why I called; is I was uh (0.1) in San Jose, on Monday, downtown at the the uh (0.4) demonstration that was going on then, and it (0.5) was such a different feeling; than from what I'm seeing these last two days. it was so peaceful, (0.4) uh (0.1) you know; people were down there to really support each other. (0.3) and wha(h)t I'm seeing now; I I think you know is just uh; (0.3) goes against what these people are trying to demonstrate for.

From conversation analytic work on telephone calls (Schegloff and Sacks 1973, Schegloff 1979b), we know that one of the main concerns of a caller is placing some 'mentionable' in a way that it

will be recognized as the 'reason for the call'. In private telephone conversations, it may be in the participant's interest to delay such an announcement, especially if the reason is to request a favour. In radio talk, however – as here – the constraints on time and topic are such that callers can generally be expected to introduce the reason for the call with few or no preliminaries.[25] Yet the first complete utterance by the caller in (12) (lines 5f.), although it is preceded by an *uhm* which begins high and loud as if to announce an upcoming reason, is not hearable as the reason for the call. Instead it comes across as an 'aside' directed to an aspect of the *hic et nunc* of this telephone call: *that guy I just listened (to)* manifestly refers at this point in talk to the endangered on-the-spot reporter. The fact that this comment is not the reason for Debbie's call is made explicit in her next turn, where she verbally labels the concern presented there as the reason for calling (lines 8ff.). Thus, retrospectively, the unit *that guy I just listened (to) that really really upsets me* is confirmed as not being the reason for the call.

Yet the 'aside' character of Debbie's turn in line 5 is not produced only retrospectively. Debbie's turn bears in it prosodic cues signalling its subsidiary status: low pitch, soft dynamics and fast tempo. This prosody stands in marked contrast to the high, loud and slow prosody of subsequent utterances such as *it was such a different feeling* (lines 10ff.) or *it was so peaceful* (line 12). The latter are verbally labelled and treated as relevant and topical. Thus it stands to reason that the markedly contrasting prosody in *that guy I just listened (to)* is contextualizing its turn as less relevant and off-topic. In addition to its emotional overtones, the prosody of Debbie's utterance in lines 5f. can be thought of as a feature of the design and fit of her turn in its particular context. As such, it has a substantial contribution to make to the situated interactional meaning of talk.

3.3 Treat the data as emergent in the real time of ongoing interaction

In contrast to much text and discourse analysis, which tends to regard spoken data *ex post factum* as a finished product or the 'behavioral realization of a preplanned cognitive unit' (Schegloff

1982:71), conversation analysts stress that discourse must be treated as being accomplished over time. In Schegloff's words: 'Good analysis retains a sense of the actual as an achievement from among possibilities: it retains a lively sense of the contingency of real things' (1982:89). The notion of emergence also transfers to prosody: just as a multi-unit turn is, in the default case, not produced as a monolithic, preconceived whole but grows incrementally

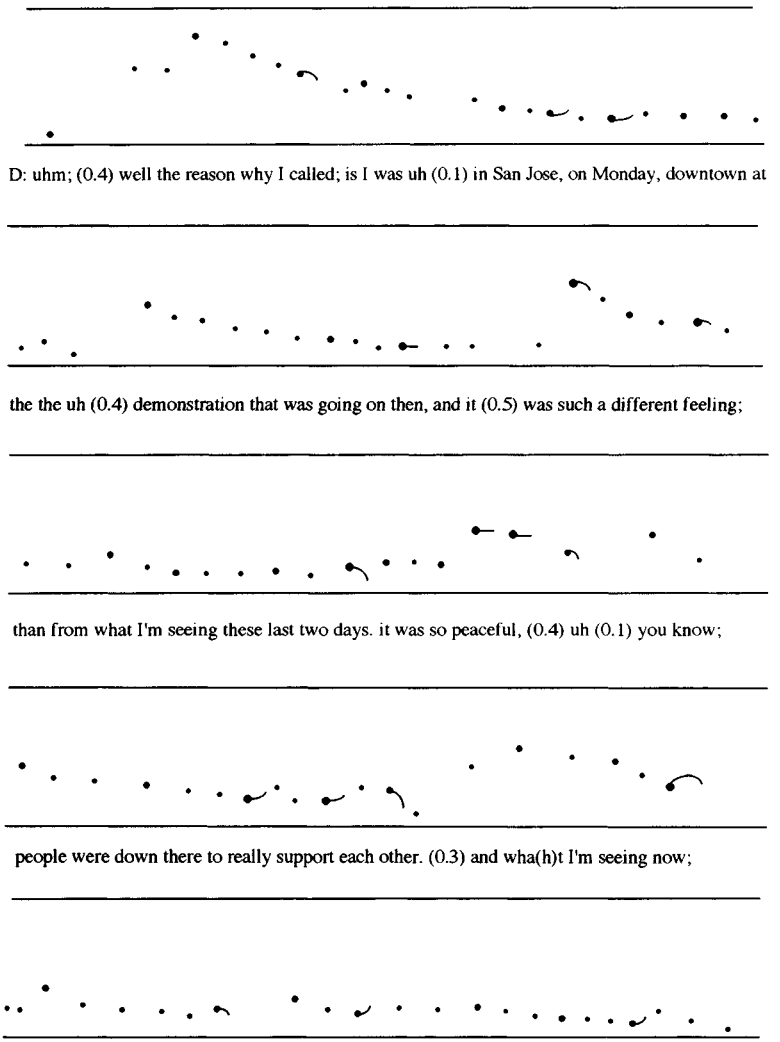


Figure 1.4

and anew at each transition relevance place (TRP) as the result of an interactive achievement, so prosodic units carried by turns or turn-constructional units may emerge as a contingency of the moment-to-moment decisions speakers make to continue or not and, if so, *how* to continue.

By way of illustration, consider Debbie's multi-unit turn in (12), reproduced here in full with prosodic notation (see Fig. 1.4).

If we are to regard this turn as emergent discourse, then we must think of it as an incremental product achieved by both participants collaboratively. This means realizing that at each and every TRP which arises, Leo has the opportunity to take over the floor himself. When he passes up this option, Debbie's continuation is as much a product of Leo's pass as of her own speaking plan. Assuming, hypothetically, that TRPs in this variety of English are signalled by a combination of complete syntactic structure (whether word, phrase or clause) and semantic (topical) unit, in conjunction with a fall-to-low pitch and an isochronous rhythmic pattern,[26] the first place at which a floor switch could occur would be following *people were down there to really support each other*.[27] At this point, Debbie has completed an idea: she has established a contrast between the demonstration in San Jose and the deictically given one in San Francisco, between the *then* of the former and *these last two days* of the latter, between the *so peaceful* nature of the former and the deictically implied nature of the latter. As the pause after *support each other* indicates, Debbie *stops* at this point, long enough to allow Leo time to come in.[28] However, Leo does not take up the option, whereupon the floor reverts to Debbie. Debbie now has the option to take over the floor again or to pass it back to Leo. If she does the latter, a lapse would presumably result, since Leo has just indicated no desire for the floor himself. Debbie continues instead. Yet her 'continuation' is not a prosodic continuation (Local 1992): there is a break in rhythm between the regular beats on *really*, *support* and *other* and the next stress on *what*. There is also a declination reset, a shift up to a high onset on *what*, from which subsequent stresses descend anew.[29] Debbie thus designs the prosody of her turn to contextualize it as a new contribution to talk (not a continuation of her former turn), and in doing so she retrospectively displays that the prior stretch of talk was complete. A prosodically 'informed' transcription of this passage as emergent discourse would be as follows:

(13) *Revised transcription of (12) as emergent discourse*

1 D: uhm; (0.4) well the reason why I called; is I was uh (0.1) in San Jose, on Monday, downtown at the the uh (0.4) demonstration that was going on then, and it (0.5) was such a different feeling; than from what I'm seeing these last two days. it was so peaceful,
5. (0.4) uh (0.1) you know; people were down there to really support each other.
L: (0.3)
D: and wha(h)t I'm seeing now; I I think you know is just uh; (0.3) goes against what these people are trying to demonstrate for.

An emergent perspective on the prosody of this multi-unit turn suggests that it was not conceived as a whole with two prosodic 'paragraphs' but that it grew into its ultimate shape as a result of on-the-spot local decisions made by participants in the process of negotiating talk.

3.4 Ground analytic categories in the data itself

Sacks, Schegloff and Jefferson have stressed that the sequential organization of conversational interaction serves as a means for interactants to display their understanding of prior talk to each other. At the same time, they point out, this mutual display of understanding can be used by analysts as a 'search procedure' in order to discover 'what a turn's talk is occupied with' (1974:728f.). Applied now to prosody, it should be possible to refer to the sequential organization of conversation in order to discover what a turn's prosody is doing. Wootton (1989) has pointed out that conversation analysts can be shown to rely on several different types of evidence for the discovery of members' devices: (i) relationship to just prior turns, (ii) co-occurring evidence within the turn, (iii) subsequent treatment in interaction, (iv) discriminability, and (v) deviant cases.[30] In fact, each of these types of evidence can be brought to bear on prosodic analysis.

(i) The relationship of the device to just prior turns. Prosodic categories can be discovered by examining the relationship between a prior turn and/or its prosody and the prosody at the beginning of a new turn. To illustrate: Local (1992) shows that a speaker may signal that a new turn is actually a continuation of a foregoing interrupted or discontinued turn by gauging pitch level, loudness

and tempo at the beginning of this turn to match those at the end of the turn broken off. Thus:

(14) *Prosodic continuation (spelling standardized)*

```
1     LOTTIE: where'd you get the turkey.
      EMMA:   up at the Balbo Market
      LOTTIE: oh y[eah. ]
      EMMA:       [°hhh]  it had a little bla:ck spot though on the
5             white skin I wonder if it was brui::sed. h
                                  (.)
      LOTTIE: [ye:ap]
→     EMMA:   [blood]
                                  (.)
10    LOTTIE: probably [wa::s,
      EMMA:            [°hh
→     EMMA:   blood coagulated but uh
```

(Local 1992:283)

According to Local, there is nothing in the prosody of Emma's talk at drop-out (line 8) which projects continuation, and yet, following Lottie's *probably was* (line 10), Emma's resumption with *blood coagulated* (line 12) has the same pitch and loudness characteristics as her first *blood* (line 8). Local argues that matching the prosody of prior interrupted talk in this way is a means conversationalists have of 'doing continuation'.

A second example of this type of evidence being used in prosodic analysis can be found in Couper-Kuhlen and Auer (1991), where it is shown that in English conversation the timing of the first prosodic prominence in a new turn with respect to the last two prosodic prominences of a prior turn may display the degree of cohesion between turns – and metaphorically, the degree of momentary speaker affiliation or disaffiliation with an interlocutor. Consider, for example:

(15) *Lack of rhythmic cohesion*

(Left-hand slashes precede prosodic prominences; vertical alignment indicates that the prominences come at perceptually regular intervals in time.)

```
1     DJ:   /whereabouts in/
            /Bolton do you /
            /work.
→     G:    (0.5) eh -I
```

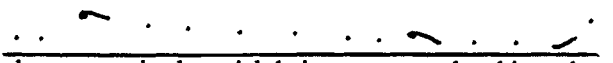
5 /don't; I'm unem-/
/ployed – well a /
/student; /
/part-time.

(1991:11)

Here DJ sets up a clear 'rhythm' at the end of his turn by timing the last prosodic prominences to come at quasi-equal intervals in time. G, however, 'misses his cue' by not picking up the established rhythm and coming in on the next beat. Instead, the first prosodic prominence of G's turn is late with respect to the pulse, although he establishes a rhythm of his own thereafter. Given the fact that DJ's question (lines 1–3) appears to be making an unwarranted assumption about G (see below), it can be argued that the timing of G's subsequent incoming (line 4) is a quasi-iconic 'document' of the gap in understanding between himself and his interlocutor.

(ii) Co-occurring evidence within the turn. The interpretation of prosodic categories may be discoverable from verbal evidence in the turn or in surrounding talk. Thus, in cases like the above, when cues are used cumulatively, the interpretation of prosody is suggested by the wording of the turns concerned. In (14), for instance, Emma repeats her prior turn, this repetition being consistent with the interpretation that the prosodic cues signal 'resumption' and 'continuation'. In (15) G formulates an answer which explicitly contradicts the presupposition of DJ's question; the content of his turn thus suggests the interpretation of the rhythmic break as a sign of disaffiliation.

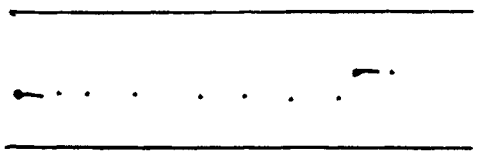
(iii) Subsequent treatment of the interactional device in question. Prosodic categories may be discoverable by inspecting following talk to see how a turn-constructional unit or a turn and its prosody are treated there. Selting (1995), for instance, shows that in standard German when speakers use mid-level pitch on the final accented syllable of a turn-constructional unit, this is treated by recipients as a signal that current speaker has not finished the turn. Potential next speakers hold off their entries at this point, as the example in Figure 1.5 illustrates. Here I's answer reaches points of potential syntactic/semantic completion after each individual turn-constructional unit. Yet the absence of next-speaker incomings – despite pauses – following *because everyone needs an education,* and *and I'm not going to worry about it any more*, and the



N: aber KUNST is aber nich kein gutes ANgebot hier oder
but there's not much offered in art here is there

I: (0.5) ES GEHT NEE:: (0.3) NICH so SONderlich GUT
it's alright no not so very good
(0.5)

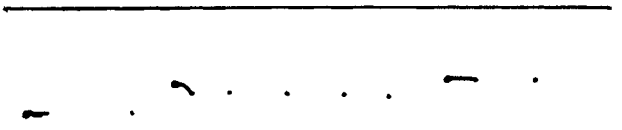
N: mhm
(1.0)



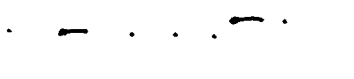
-> I: a:ber ich mach das jetzt hier zuENde (0.7)
but I'm going to finish this now here
<f>

I: WEIL: eine ausbildung BRAUCH der mensch (1.4)
because everyone needs an education

I: aso s HAB ich mir jetzt so geSA:GT (0.2)
or so I've said to myself now

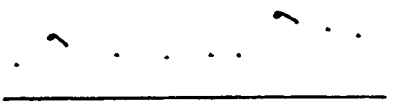


I: und: (0.2) ich KÜMmer mich da nich WEIter drum (0.7)
and I'm not going to worry about it any more



-> I: ich MACH das hier zuENde (0.7)
I'm going to finish this here

I: un mal SEHN was DANN kommt (1.0)
and I'll see what happens then



N: in WELchem semester BIS du denn
what semester are you in anyway

(Selting 1995; **CAPS** denote primary, CAPS secondary, accented syllables.)
Figure 1.5

two occurrences of *I'm going to finish this here* suggests that mid-level pitch on a final accented syllable in a turn-constructional unit is functioning as a prosodic turn-holding device here.

(iv) Discriminability of the interactional device. Systematic comparison of the prosodic configuration in question with other configurations may help to reveal the specific role which a prosodic device plays. For instance, the mid-level terminal pitch in standard German referred to above can be compared with final falling or final rising terminal pitch configurations in similar contexts, which do not routinely block entries by next speakers.

Similarly, with respect to the notion of prosodic continuation illustrated in example (14) above, it can be shown that speakers sometimes use different prosody on a new turn following an interrupted or discontinued one:

(16) *Prosodic re-starting*

1	W:	... this was about ten o'clock at night on the London Underground and there was a whole carriage in stitches
→	LT:	I must [say y[ou do sound a bit
5	W:	[°hhh [do you have any idea [what a- * * *
	PR:	[* * *
		(.)
→	LT:	you do sou:n:d disconcertingly like an alternative theatre group

(Local 1992:285)

According to Local, the prosody of LT's repeat *you do sound* (line 8) is higher in pitch than the corresponding part of his prior turn (line 4). Thus he can be heard as not continuing but 'starting anew'.

By contrast with (15) above, it can be shown that the timing of an answer to an out-of-the-blue question like *whereabouts do you work* or *what do you do in life* need not be late and arhythmic:

(17) *Rhythmic cohesion*

1	DJ:	what d'you
		/do in life /
		/John? /
→	J:	/uh `well I'm /
5		/off` sick at /
		/present, /

(Couper-Kuhlen and Auer 1991:14)

Although here too DJ's question (lines 1–3) appears to be making an unwarranted assumption about his interlocutor and the second speaker J formulates an answer which is not the preferred or expected one, in doing so he picks up DJ's rhythm by placing the first prosodic prominence of his turn such that it coincides temporally with the next pulse (line 4). Thus, the timing of J's turn onset is rhythmically well-coordinated with the prior turn and the sequence as a whole comes off cohesively.

(v) Deviant cases in the use of the device. Using evidence such as that cited above, conversation analysts typically proceed on a case-

to-case basis assembling an 'analytically coherent set of cases' (Schegloff 1982:80). With a set of like cases, systematic features in interaction can be more reliably identified. Of course deviant cases may appear in which the hypothesized regularity seems to break down. However, close analysis of such deviant cases often reveals that what appears to be a contravention of the 'rule' at first sight is actually a confirmation of it on a deeper level.

It was precisely such deviant-case argumentation which we employed above to underline the context-sensitivity of prosody in the Gulf War phone-in example (12).[31] Recall that Debbie's *that guy I just listened (to)* initially appeared to deviate from the routine placement of the reason for the call in telephone conversations on radio phone-ins. Without taking prosody into consideration, this example would constitute a deviant case. However, with proper consideration of its prosody, example (12) can be shown to confirm a deeper generalization: routinely, on-topic turns in this sequential position are reasons for the call, a status confirmed by corresponding prosodic treatment (non-low pitch register and volume, non-fast tempo). Off-topic turns in the same position, on the other hand, are prosodically contextualized as 'misplaced'. Their contrasting prosodic treatment thus indirectly confirms the routine understanding of similarly positioned (and appropriately contextualized) turns.

There is one well-known study in which deviant cases, or 'hitches' (Levinson 1983:319), have been used to discover prosodic regularity. Beattie, Cutler and Pearson (1982) set out to investigate turn delimitation, and in particular the numerous interruptions found to occur in formal interviews with the then Prime Minister Margaret Thatcher. Based on close analysis of one interview, they find that turn-taking hitches – in which the interviewer was invariably heard to be interrupting his interviewee – were regularly accompanied by current speaker's (i.e. Thatcher's) use of rapid or steep pitch falls which reached a mid point in her voice range. By contrast, in clear speaker switches in the same interview, current speaker used steep falls but ones which regularly ended quite low in the voice range. The authors conclude that the turn-taking hitches in this interview were due to a conflicting set of cues: a pitch fall to non-low typically found in turn-medial utterances but a steep gradient as found in turn-final utterances. These conclusions are substantiated via traditional social-scientific methods such as

polling independent judges for judgements of finality/non-finality and statistical tests. However, it was the study of hitches which originally led to hypothesizing the regularity behind prosodic turn delimitation in these interviews.

3.5 Validate analytical categories by demonstrating participants' orientation to them

As Sacks, Schegloff and Jefferson point out, the fundamental sequential organization of conversational interaction can also be appealed to for the validation of analyses (1974: 729). That is, via an inspection of turns at talk directed at a prior turn's talk, analysts can access participants' displays of understanding directly, the latter serving as warrants for analytic decisions. Here too, there is a straightforward application to the analysis of prosody. One study which demonstrates this is Kelly and Local's (1989b) investigation of test-word repetition by interviewees in the Tyneside Linguistic Survey. Their hypothesis is that speakers' use of a rapidly falling pitch contour (from high to low) with marked loudness on a repeat following interviewers' proffering of a word for recognition counts as an understanding check, designed to verify that the word has been heard correctly. For example:

(18) *Understanding check* (McN is the interviewer, Eir the interviewee)

1	MCN:	er (0.2) varnigh
		(3.5)
→	EIR:	varnigh (high falling pitch to low, marked loudness)
		(0.4)
5	MCN:	aye (1.5) you know for nearly
		(1.5)
	EIR:	w I've never heard it I've heard me mother use it

(adapted from Kelly and Local 1989b:267)

Kelly and Local's argument in support of this hypothesis is based on sequential evidence: the following turn by the interviewer routinely contains some acknowledgement or confirmation token (here *aye* in line 5), and the interviewee does not offer any sign of recognition (as here in line 7) until this acknowledgement is given. Moreover, when the hypothesized prosodic cues are not present, i.e. when speaker's falling pitch stops mid-way, or does not begin high and is realized with a breathy, soft voice, the

sequential treatment is different. In the latter case, rather than immediately confirming or disconfirming the speaker's rendition of the word, the interviewer routinely withholds a turn until the speaker has offered some appreciation of it. Kelly and Local suggest that this kind of prosody signals that speakers are 'mulling over' the word. Compare:

(19) *Mulling over* (McN is the interviewer, GSh the interviewee)
1 MCN: er (0.8) varnigh
(0.9)
GSh: varnigh (flat falling pitch from mid, quiet voice)
(1.0)
5 oh yes I've sometimes said varnigh
MCN: aye (.) uh (.) yeah
(adapted from Kelly and Local 1989b:274)

Thus the interactional and prosodic categories which this study proposes are warranted by demonstrating that participants to the interaction orient to them in predictable ways.[32]

4 Interaction from a prosodic perspective

The preceding discussion has demonstrated how the methodology of conversation analysis can be profitably transferred to the study of prosody within an interactional framework. Ultimately, we believe, this kind of prosodic analysis will also enrich the study of conversation. Among current approaches to spoken interaction, three come to mind which acknowledge in one way or another the prosodic make-up of talk: standard conversation analysis,[33] the 'information-flow' model (Chafe 1980, 1988) and contextualization theory (Gumperz 1982; Auer and di Luzio 1992).[34] These approaches differ widely with respect to the importance they attribute to prosodic form and function. Yet all stand to benefit from advances in the field of interactional prosody, we shall argue, in two ways: systems for the transcription of prosody can be made more sensitive to interactionally relevant features, and analyses of conversational discourse can be made more revealing of participants' methods by incorporating interactional prosodic functions.

4.1 The transcription of prosody

With one or two minor exceptions, the conventions for prosodic transcription in current approaches hark back to a single model, that developed over a number of years by Gail Jefferson (see the surveys in Schenkein 1979, Atkinson and Heritage 1984, Psathas and Anderson 1990). Newer systems, e.g. DuBois, Schuetze-Coburn, Cumming and Paolino (1993) and Gumperz and Berenz (1993), may advocate slightly different symbols in order to facilitate computerized word-processing, but the categories themselves tend to remain on the whole the same. The following discussion takes CA conventions as its point of departure, as these are most widely known. However, our remarks apply as well to similar systems of prosodic transcription.

Traditionally, conversation analysts have used a largely ad hoc notation system for the representation of segmental-phonetic and prosodic detail. These conventions are justified with respect to the aim of making transcripts 'accessible to linguistically unsophisticated readers' (Sacks, Schegloff and Jefferson 1974:734). High priority is given to the readability of transcripts by non-linguists. Dialectal or other noticeable deviations from standard pronunciation are typically indicated by adopting English spelling conventions in a kind of 'folk-phonetic' representation 'to get as much of the actual sound as possible into our transcripts' (Sacks, Schegloff and Jefferson 1974:734). Such transcripts inevitably rely on the reader's familiarity not only with (implicit) spelling conventions but, more generally, with the speech variety at hand (cf. also Gumperz and Berenz 1993).

In the domain of prosody and intonation, a minimal notation system has been proposed which, in one of the more recent compilations, provides for the use of a period /. / for a 'stopping fall in tone', a comma /, / for 'continuing intonation', a question mark /?/ for a 'rising inflection', a combined question mark/comma /;/ for 'rising intonation weaker than that indicated by the question mark', upward and downard arrows /↑,↓/ for 'rising or falling shifts in intonation immediately prior to the rise or fall' and exclamation point /!/ for 'animated tone' (Psathas and Anderson 1990:94). Pitch change within the word is to be notated by the combination of underscoring for stress in relation to the prolongation marker /: /.

The stress-mark (underscoring) is placed on the 'first letter' of a stressed (and prolonged) syllable if there is no change in pitch: /venee:r/ (*sic*); on the vowel immediately preceding the colon in case of a pitch drop: /venee:r/; and on the prolongation in case of a rise: /venee:r/ (Psathas and Anderson 1990:94f.). As for timing, it is recommended that the onset of overlapping talk be represented with double obliques (//) or right-hand brackets, that 'no interval' between turns be indicated with an equal sign (=), and that silent intervals in the stream of talk (within and between utterances) be timed in tenths of a second and indicated in parentheses (Psathas and Anderson 1990:93).

Yet current notation conventions will be found lacking with respect to two general principles of design for systems of discourse representation (see also Edwards 1993). First, prosodic symbols should allow readers without access to the original data to recapture important features unequivocally. However, the use of punctuation marks, for instance, for the representation of phrase-final intonation depends on stereotypical associations between sentence-type and intonation which are misleading, especially for the non-linguist. Clearly, not all questions have rising intonation, nor do all utterances with rising intonation function as questions. Although most current transcription systems explicitly warn users against such simplistic equations, nonetheless the use of written-language devices for the transcription of spoken language invites misunderstanding.[35] This could be avoided with the use of less functionally loaded and more iconic symbols.

Furthermore, the notation of pitch change within the word by variable underlining in the graphic word is problematic because of the frequent lack of correspondence between phonics and graphics in language. In addition, to mark a pitch change (or lack of it) by underlining the syllable onset or coda is unfortunate because it suggests optically that it is consonants which carry pitch. (Phonetically speaking, presence or absence of pitch change should be marked on the syllable peak.) Moreover, when a pitch change is marked on the syllable coda, this suggests, by virtue of the implicit iconicity between space and time, that there is a delay in the temporal alignment of pitch peak and syllable core. Such delays do occur in intonation but are generally recognized as producing different tones (e.g. a rise–fall instead of a fall; cf. Ladd 1983).

Of course the exact wording of the transcription conventions may not warrant these implications, but the point is that the pitch and stress conventions invite an iconic reading where none is called for. In actual fact, the symbols would benefit from being made more mnemonic.

Third, as notation conventions stand now, it is unclear what the domain of the pitch changes recorded with punctuation marks is. A plausible guess would be that they extend from the last stressed syllable over any and all following unstressed syllables. Yet this assumption is not made explicit, nor is stressing systematically marked in all phrases.[36] A related problem concerns where exactly the pitch of terminal intonation changes. Does a final question mark refer to a rising inflection on the last stressed syllable or on the last (unstressed) syllable? If the former, how can a change in pitch direction on an unstressed tailing syllable following the last stressed syllable be captured? In traditional intonation analysis, such a pattern would correspond to a falling–rising (nuclear) tone as opposed to a simple falling one, a contrast which has been attributed a distinctive function in English.[37] In Brazil's discourse-analytic approach to English intonation, the difference is fundamental (recall the Referring vs Proclaiming tone). Local (1992) provides conversational evidence that falling–rising pitch at the end of a turn is projective of more talk to come (1992:275). Arguably then, an additional category for falling–rising phrase-final intonation must be introduced if it can be shown to be functionally different from a simple rise.

This brings us to a second and all-important design principle with respect to systems of prosodic notation. They should make it possible to represent especially those prosodic features which are relevant to conversationalists. Here too, current systems are often lacking. For instance, the provision for two categories of rise as opposed to one for falls suggests that the *kind* of fall used is irrelevant for the organization of talk. Yet Lehiste (1975), Yule (1980) and others have shown that speakers of English use differing degrees of falling intonation to signal something akin to 'paragraph' non-finality vs 'paragraph' finality. Kelly and Local (1989b) find that speakers of Tyneside English use low falling pitch for finality, non-low falling pitch (i.e. fall to mid) for non-finality in turn construction. Since how low a contour falls may have both a discourse-structural and an

interactive function, it can be argued that transcription conventions should provide for two kinds of fall: low (full) vs non-low (truncated). Furthermore, there is a certain amount of evidence accumulating that the relative pitch level at which a speaker *begins* an intonation phrase (called *onset* or *head* in tonetic research and *key* in discourse-analysis studies) may have a discourse-structural function (Brazil, Coulthard and Johns 1980, Couper-Kuhlen 1983, Wichman 1991). High onsets, for instance, typically accompany topic initiations. In oral narratives they may be used to signal a return to the foreground following excursions into the background or evaluative commentary. Low onsets tend to be associated with intonational parentheses. Yet current transcription conventions make no systematic provision for notation of onset level.

Whereas tonal categories in current systems tend to be underdeveloped, conventions for the notation of timing are often overly 'exact'. With respect to the transcription of pauses, for instance, they imply an objectivity which is unwarranted in two senses. First, in relation to accuracy of measurement: Jefferson (1989) reports originally having used a stopwatch to time breaks exceeding approximately 0.2 sec. Yet the method now recommended is slow and regular counting: 'no one thousand, one one thousand, two one thousand . . . ' (Jefferson 1989). Psathas and Anderson (1990) also recommend this method over clock-time accuracy, provided the transcriptionist is internally consistent. However, the problem is that pauses measured this way continue to be expressed in tenths of a second. This not only incorrectly implies real-time objectivity; it also makes cross-transcription comparison hazardous and prevents meaningful generalization.[38]

The objectivity implied by timing conventions is also unwarranted with respect to pause identification. According to pause-perception research, observed pauses do not always correspond to actual silences and, vice versa, actual silences need not be heard as pauses (see e.g. Duez 1985, Carpenter and O'Connell 1988). Consequently, when analysts set out to transcribe pauses as accurately as possible, they are not making an objective record of silence (absence of phonation over some specified interval of time) but of what *they* consciously *perceive* as silence. These (transcriber) perceptions are not necessarily the same as what interactional participants hear and/or orient to.

This is by no means to say that the lay notions of long and short pause or of coming in 'too late' or 'too early' are irrelevant for interaction. On the contrary, they are highly relevant, but the point is that they are not based on absolute time. Empirical studies of naturally occurring English conversation suggest, for instance, that it is the *rhythm* and *tempo* of surrounding talk which determine whether some absolute length of silence is perceived as long or short. As for turn-taking, the timing of (minimally) the last two prominent syllables in a speaker's turn appears to serve as a metric for the next speaker's entry (Couper-Kuhlen 1991, 1993). Yet in current transcription practice, no provision is made for indicating this perceptually more relevant *rhythmic* timing. By virtue of recording 'objective' time, transcribers misleadingly imply potential relevance for it. Moreover, its (pseudo-)objectivity suggests parallels where none are called for and obscures perceptually relevant likenesses.

As they stand then, current notation conventions are not easily readable nor precise enough to retrieve crucial aspects of the original – and, most importantly, they do not necessarily capture those prosodic categories which are of potential interactive relevance for conversational participants. In place of ad hoc categories, we would advocate the introduction of interactionally grounded ones for the transcription of prosody in conversation. By interactionally grounded categories, we mean ones which have been discovered and warranted according to the methodological principles outlined above. What exactly these categories are must be worked out in interactional studies. We thus take an analytic stance similar to that of Gumperz and Berenz (1993), who advocate a *functional* approach to the recording of prosodic phenomena, one in which transcription becomes an integral part of the analysis: 'In rendering prosodic, intonational and paralinguistic phenomena, we . . . concentrate on just those features of pitch, tune, and accent that can be shown to affect situated interpretation at the interactive or relational level as well as at the level of content' (1993:92). The goal of prosodic transcription in our view is not an objectivistic, etic notation of the original,[39] but a record of those prosodic details which may be of relevance for the inferential work of participants. One sets down on paper 'all those perceptual cues that past research and on-going analyses show participants rely on in their online

processing of conversational management signs' (Gumperz and Berenz 1993:92). To a certain extent then, adequate prosodic notation requires prior analysis.

On the other hand, we recognize that some guidelines must be available in performing this preliminary prosodic analysis. Here a plea can be made for detailed *phonetic* observation of an 'impressionistic' sort (Kelly and Local 1989a). ('Impressionistic' is a phonetician's term used to refer to the record of all auditory sound distinctions without reference to their function.) In Kelly and Local's words, 'at the beginning of work on language material we can't, in any interesting sense, know beforehand what is going to be important. Consequently we must attend to and reflect everything we can discriminate' (1989b:26). Behind the notion of impressionistically recording everything one hears is a healthy 'fear' of preestablished notation systems, which may bias and restrict researchers' perception and analysis. In this sense, Kelly and Local are implicitly warning against premature phonologization of categories. Presumably, however, once impressionistic records have been functionally interpreted and the relevant prosodic categories extracted, the latter can be generalized for a given speech community. Then, for the study of conversation as social interaction, only those functional categories shown in interactional analysis to be relevant for conversational inferencing need be retained in the final transcript. Thus both phonetic *and* functional approaches to prosody are relevant to the analysis of conversation, each in its own way.

4.2 *The contribution of prosody to interactive meaning*

Current approaches to spoken interaction range from nominally attributing some relevance to prosody to assigning prosody a central role in the production and understanding of interactive meaning. Conversation analysts, for instance, have openly attested to the relevance of intonation for turn construction. The basic unit of a turn, the so-called turn-constructional unit, they point out, cannot be fully described in syntactic terms, as 'any word can be made into a "one-word" unit-type, . . . via intonation' (Sacks, Schegloff and Jefferson 1974:721f.). This formulation suggests that speakers and recipients make use of intonation or better, prosody, to delimit the basic unit of conversational interaction.

However, on the whole conversation analysts pay little attention to prosodic phrasing. In transcription, the end of something similar to a phrasing unit is presumably represented by a punctuation mark. Yet this only implicitly suggests that the material between two punctuation marks is indeed an integral prosodic phrase. In fact, CA attention has focussed more on the sequential aspects of conversational organization, key notions being 'conditional relevance' and the 'sequential implications' that a first turn creates for the next. The prosody of a turn, however, has tended to be taken for granted: how it contributes – along with sequential location – to steering conversationalists' interpretations is not made explicit. Despite an impressive arsenal of transcription symbols for the representation of prosody, it figures surprisingly little in CA reconstructions of speakers' methods for doing conversation.

In the work of Chafe and his associates, by contrast, the intonation phrase is treated as the fundamental unit of discourse production (Chafe 1979, 1980). It is also therefore a unit in terms of which major research questions are formulated. In Chafe's model the intonation unit has a cognitive counterpart: it corresponds to a focus of consciousness (originally called an 'idea unit'), or the amount of information to which one devotes central attention at a given time. Its phonetic description has evolved over time (cf. Chafe 1980, 1988, 1993). However, major effort is now being invested to specify its characteristics and make explicit strategies for its identification (DuBois, Cumming and Schuetze-Coburn 1988; DuBois, Schuetze-Coburn, Paolino and Cumming 1992; DuBois, Schuetze-Coburn, Cumming and Paolino 1993).

Yet for all the importance attributed to this unit of intonational phrasing in Chafe's approach, surprisingly little attention is paid to other prosodic phenomena, although provision is made for noting stress, loudness and speech rate in transcription. As for intonation, the main functional distinction is between final (or 'period'), continuing ('comma') and appeal ('question mark') transitions between intonation units (so-called *transitional continuity*). Reminiscent of the American structuralists' terminal juncture, transitional continuity is said to signal 'whether the discourse business at hand will be continued, or has finished' (DuBois, Schuetze-Coburn, Paolino and Cumming 1992:28). The functional classes of transitional continuity may have different phonetic contour realizations cross-

linguistically, but presumably have a constant realization within one speech community. The distinction between continuing and finished intonation plays a crucial role in Chafe's theory of levels of linguistic processing. A 'finished' intonation contour signals the end of a spoken 'sentence' and is expected when a thought or centre of interest is complete.

In contrast to Chafe's rather limited focus on intonation unit, pause and terminal continuity, Gumperz envisages a much broader range of prosodic features which enter into speaker inferencing. His attention has focussed on stress, pitch direction and register, timing, and loudness, as well as on paralinguistic and kinesic features as situated cues to the (mis-)understanding of speakers' interactional meanings in (intercultural) communication (1982, 1992).[40] Yet such analyses can only be carried out against the background of speakers' unmarked prosodic expectations in discourse, and, given the present stage of research, these can often only be guessed at.

On the whole then, both with respect to turn construction and to activity-related inferencing, the contribution of prosody in the organization of conversational interaction and in the negotiation of interactional meaning needs more explicit recognition and detailed attention. The interactional approach to prosody outlined above will benefit current conversational and discourse research by drawing attention to a wide range of interactive functions in which prosody is implicated.

5 Conclusion

In proposing an interactional perspective on prosody as a 'remedy' for formal, functional and methodological problems in current research, we have taken a doubly empirical stance: (i) the approach advocated takes empirical data as its object of study, and (ii) it seeks empirical evidence for the validation of its analyses. While the analysis of empirical data is arguably not new in prosodic studies, the kind of empirical validation suggested here is often absent from other approaches. It is true that the term 'discourse' is beginning to appear in mainstream work on prosodic and intonational phonology.[41] Nevertheless, this kind of work is not oriented towards the empirical analysis of natural data. Other recent studies clearly set out to investigate intonation in natural settings.[42] This attests to the

growing awareness that the study of prosody should be based on empirical data from natural interaction. What is absent, however, is evidence that participants in real interactions do indeed interpret prosody and/or intonation in the way the researcher models it. In contrast to most other empirical approaches, an interactional perspective emphasizes the necessity of empirically warranting prosodic categories and descriptions. And in contrast to structuralist approaches and their offspring, an interactional perspective does not simply aim at pattern recognition and description, but at the reconstruction of patterns as cognitively and interactionally relevant categories which real-life interactants can be shown to orient to.

In suggesting ways to improve the formal and functional analysis of conversation by incorporating a prosodic perspective, we have taken a step towards the reconstruction of what Silverstein (1992) calls the 'interactional text'. The interactional text is 'laid down in realtime discursive interaction' (1992:58); it contains not only all the referential sign-forms which participants in interaction use (these comprise the 'denotational text') but also all those indexical ones which contribute to meaning and inference in the widest sense. If the interactional text is strategic and maximally transparent, i.e. gives a plausible reconstruction of 'moves' in interaction based on participant understandings, it can be 'studied transcriptionally in vitro with confidence that the in vivo reality is close to hand' (1992:74). It is our contention that prosody furnishes some of the indexical sign-forms which contribute to the establishment of interactional text.

Notes

1 In defining prosody this way, we align ourselves within the Firthian tradition, where prosody is understood to comprise the syntagmatic properties of syllables, words and phrases as opposed to the paradigmatic, phonematic properties of single sounds (Firth 1957/1969).
2 See Auer, Couper-Kuhlen and Müller (to appear) for a discussion of how Saussurian linguistics has 'detemporalized' language.
3 For the same reason repair phenomena have been excised from the sentences with which linguists and syntacticians have concerned themselves (Schegloff 1979a; also Müller 1993).
4 *Context* shares a number of these features and has likewise been overlooked in modern, especially formal, linguistics (Goodwin and Duranti 1992).

5 See Auer 1993 for a historical survey of technical recording and speech analysis aids and the way linguists have made use of them.
6 *Prose* is defined as 'essentially language organized for visual presentation' (Abercrombie 1965:3).
7 We might add that the emergent nature of conversation (see below) makes it radically different from prose, a finished product.
8 E.g. Liberman and Sag 1974 or Pierrehumbert and Hirschberg 1990.
9 Others who deserve mention here include folklorists working in the Tedlock/Hymes 'ethnopoetic' tradition (see, e.g., Sherzer and Woodbury 1987) and sociologically oriented literary critics in the Bakhtin circle (see, e.g., Vološinov 1976).
10 Halliday 1967, 1970 and Crystal 1969 represent two well-known British schools.
11 Under certain conditions a tone unit may have *two* prominent pitch movements in Crystal's model, but only provided the directions of these pitch movements are complementary and there is no rhythmic or melodic break between them. Halliday allows for similar 'compound tones', stipulating that there can be no pretonic segment in between.
12 Barring (in Crystal's model) tonal subordination.
13 The falling pitch movements on the two occurrences of *really* would presumably be treated as superordinate to those on *down* and *upsets* respectively in Crystal's system.
14 See also Local and Kelly (1986:185), who make this point cogently.
15 See French and Local (1983), Local and Kelly (1986), Local, Kelly and Wells (1986) and Local, Wells and Sebba (1985) for a selection of phonetic and prosodic parameters which have proved to be relevant for the analysis of speech as verbal interaction.
16 Armstrong and Ward list as a fourth use for Tune II 'incomplete groups', a first step towards the recognition that tunes need not coincide with sentences.
17 The problem with the notion of 'information unit' is one of circularity: there are no means independent of intonation for identifying this unit.
18 Recent work on focus assignment and realization has shown that the relation between 'sentence accent' and focussed constituent is much more complex than Halliday's description suggests (cf. e.g. Gussenhoven 1983 and Uhmann 1991).
19 The case for nature appears to be substantiated by universally valid patterns of intonation such as those reported in Bolinger 1978. The case for nurture, on the other hand, is strengthened by language-dependent rules of e.g. focus realization (Uhmann 1991).
20 As Gumperz (1992) points out, contextualization cues channel or constrain interpretations, which then must be negotiated and confirmed interactively. Meaning in interaction is never a 'one-shot go'.
21 The meaning of (11b) would then be approximately 'the Robinsons, of all people'.

22 See, e.g., Sacks 1992, Schegloff and Sacks 1973, Sacks, Schegloff and Jefferson 1974, Schegloff, Jefferson and Sacks 1977, for early, influential work in the field of conversation analysis.
23 Ethnomethodologists have been known to conduct 'breaching experiments' by e.g. purposefully adopting alienating methods in order to study the routine grounds of everyday interaction (cf. Garfinkel 1963, 1967).
24 See, e.g., Erickson and Shultz (1982), who use a combination of CA-inspired microanalysis of selected interactional sequences as well as post-hoc viewings of videotapes of these sequences with participants to generate and validate hypotheses.
25 Radio phone-in programmes appear to have individual codes of conduct; some provide for an initial exchange of greetings between callers and anchor-person, while others do not; in the former case routine greetings may be one-way or two-way, and may allow for a *how are you* sequence or not. On this programme, callers regularly exchange only greetings with Leo and then proceed immediately to the reason for the call.
26 Strictly speaking, this assumption would first have to be proved. We make it here hypothetically for expository purposes.
27 Arguably the mid-high sustained pitch on *peaceful* (line 12) is a prosodic turn-holding device, although Debbie does pause here long enough for some recipient signal from Leo.
28 According to a rhythm-based metric for turn-taking (Couper-Kuhlen 1991, 1993), an unmarked floor entry at this point would be timed so that its first stress coincided with the regular beat established by the stresses on *really*, *support* and *other*. The fact that the next beat is silent indicates that Leo has not made use of his option to become the next speaker.
29 In other words, the 'declination line' of prior talk does not continue here.
30 Wootton works these out based on CA investigations of tokens such as *oh* and *uh-huh*.
31 Thanks to John Gumperz for drawing our attention to this.
32 For other studies in which this kind of argumentation is used, see Local, Wells and Sebba (1985), Local, Kelly and Wells (1986) and Selting (1987a b; 1992a, b).
33 See, e.g., the contributory volumes by Schenkein 1979, Atkinson and Heritage 1984, Psathas 1979, Button and Lee 1987, Zimmerman and West 1980, Button, Drew and Heritage 1986 and Boden and Zimmerman 1991; also Heritage 1989.
34 This list is not intended to be exhaustive.
35 As Edwards herself states (albeit arguing for the opposite point): 'strategies based on reading habits are not necessarily subject to conscious awareness and may be difficult to suspend when reading a transcript, even if it is desirable to do so' (1993:6).

36 'Emphasis' tends to be recorded more systematically, using italics and/or large letters to signal higher degrees of stress (Psathas and Anderson 1990:94).
37 Cf. *She has a lovely vòice* vs *She has a lovely vŏice* ('but I don't think much of her as an actress') (O'Connor and Arnold 1973:68ff).
38 O'Connell and Kowal (1990a, b) make a plea for the use of physical time as measured acoustically in the transcription of timing despite (or perhaps because of) its non-participant-oriented, alienating effect.
39 *Pace* early work on CA (e.g. Sacks, Schegloff and Jefferson 1974), which advocates 'special care' in order to maximize the detail and precision of transcription. Psathas and Anderson (1990) take a more moderate stance, demanding merely that the transcript 'captures/displays those features of the interaction that are of analytic interest' (1990:76).
40 Cf. Tannen 1984, 1989 for applications of contextualization theory to the analysis of American conversational discourse.
41 See, e.g., Pierrehumbert and Hirschberg 1990.
42 McLemore 1991, for example, gives a 'pragmatic interpretation of English intonation' based on naturally occurring sorority speech.

References

Abercrombie, D. 1965. *Studies in Phonetics and Linguistics*. Third impression, 1971. London: Oxford University Press.

Armstrong, L. E. and I. S. Ward 1926. *Handbook of English Intonation*. Leipzig: Teubner.

Atkinson, J. M. and J. Heritage (eds.) 1984. *Structures of Social Action. Studies in conversation analysis*. Cambridge University Press.

Auer, P. 1986. Kontextualisierung. *Studium Linguistik*, 19: 22–47.

1992. Introduction: John Gumperz' approach to contextualization. In Auer and di Luzio 1992, pp. 1–38.

1993. Über ↰. *Zeitschrift für Literaturwissenschaft und Linguistik*, 90/91: 104–138.

Auer, P. and A. di Luzio (eds.) 1992. *The Contextualization of Language*. Amsterdam and Philadelphia: Benjamins.

Auer, P., E. Couper-Kuhlen and F. Müller (to appear). *Language in Time: the rhythm and tempo of spoken interaction*.

Beattie, G. W., A. Cutler and M. Pearson 1982. Why is Mrs Thatcher interrupted so often? *Nature*, 300: 744–747.

Boden, D. and D. H. Zimmerman (eds.) 1991. *Talk and Social Structure. Studies in ethnomethodology and conversation analysis*. Cambridge: Polity Press.

Bolinger, D. L. 1978. Intonation across languages. In J. P. Greenberg, C. A. Ferguson and E. A. Moravcsik (eds.) *Universals of Human Language*, Vol. II *Phonology*. Stanford University Press, pp. 471–524.

1986. *Intonation and its Parts. Melody in spoken English*. Stanford University Press.

1989. *Intonation and its Uses. Melody in grammar and discourse*. London: Edward Arnold.

Brazil, D. 1981. The place of intonation in a discourse model. In M. Coulthard and M. Montgomery (eds.) *Studies in Discourse Analysis*. London: Routledge & Kegan Paul, pp. 146–157.

1982. The place of intonation in the description of interaction. In D. Tannen (ed.) *Analyzing Discourse: Text and talk*. Washington, D.C.: Georgetown University Round Table on Language and Linguistics, pp. 94–112.

Brazil, D., M. Coulthard and C. Johns 1980. *Discourse Intonation and Language Teaching*. London: Longman.

Brown, G., K. L. Currie and J. Kenworthy 1980. *Questions of Intonation*. London: Croom Helm.

Button, G. and J. R. E. Lee (eds.) 1987. *Talk and Social Organisation*. Clevedon: Multilingual Matters.

Button, G., P. Drew and J. Heritage (eds.) 1986. Interaction and Language Use. Special double issue, *Human Studies*, 9(2–3).

Carpenter, S. and D. C. O'Connell 1988. More than meets the ear: some variables affecting pause reports. *Language & Communication*, 8: 17–27.

Chafe, W. L. 1979. The flow of thought and the flow of language. In T. Givón (ed.) *Discourse and Syntax*. New York: Academic Press, pp. 159–181.

1980. The deployment of consciousness in the production of a narrative. In W. L. Chafe (ed.) *The Pear Stories*. Norwood, N.J.: Ablex, pp. 9–50.

1988. Linking intonation units in spoken English. In J. Haiman and S. A. Thompson (eds.) *Clause Combining in Grammar and Discourse*. Amsterdam and Philadelphia: Benjamins, pp. 1–27.

1993. Prosodic and functional units of language. In J. A. Edwards and M. D. Lampert (eds.) *Talking Data: Transcription and coding in discourse research*. Hillsdale, N.J.: Lawrence Erlbaum, pp. 33–43.

Couper-Kuhlen, E. 1983. Intonatorische Kohäsion. Eine makroprosodische Untersuchung. *Zeitschrift für Literaturwissenschaft und Linguistik*, 49: 74–100.

1986. *An Introduction to English Prosody*. London: Edward Arnold and Tübingen: Niemeyer.

1991. A rhythm-based metric for turn-taking. *Proceedings, 12th International Congress of Phonetic Sciences*, Vol. I. Aix-en-Provence: Université de Provence, pp. 275–278.

1993. *English Speech Rhythm. Form and function in everyday verbal interaction*. Amsterdam and Philadelphia: Benjamins.

Couper-Kuhlen, E. and P. Auer 1991. On the contextualizing function of speech rhythm in conversation: question–answer sequences. In J.

Verschueren (ed.) *Levels of Linguistic Adaptation*, Vol. II. Amsterdam: Benjamins, pp. 1–18.

Cruttenden, A. 1986. *Intonation*. Cambridge University Press.

Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge University Press.

DuBois, J. W., S. Cumming and S. Schuetze-Coburn 1988. Discourse transcription. In *Santa Barbara Papers in Linguistics*, Vol. II, *Discourse and Grammar*. Santa Barbara: Department of Linguistics, University of California, pp. 1–71.

DuBois, J. W., S. Schuetze-Coburn, S. Cumming and D. Paolino 1993. Outline of discourse transcription. In J. A. Edwards and M. D. Lampert (eds.) *Talking Data: transcription and coding in discourse research*. Hillsdale, N.J.: Lawrence Erlbaum, pp. 45–89.

DuBois, J. W., S. Schuetze-Coburn, D. Paolino and S. Cumming 1992. *Santa Barbara Papers in Linguistics*, Vol. IV, *Discourse Transcription*. Santa Barbara: Department of Linguistics, University of California.

Duez, D. 1985. Perception of silent pauses in continuous speech. *Language & Speech*, 28: 377–389.

Edwards, J. A. 1993. Principles and contrasting systems of discourse transcription. In J. A. Edwards and M. D. Lampert (eds.) *Talking Data: transcription and coding in discourse research*. Hillsdale, N.J.: Lawrence Erlbaum, pp. 3–31.

Erickson, F. and J. Shultz 1982. *The Counselor as Gatekeeper. Social interaction in interviews*. New York: Academic Press.

Firth, J. R. 1957. Sounds and prosodies. In Firth 1969, *Papers in Linguistics 1934–1951*. London: Oxford University Press, pp. 121–138.

French, P. and J. K. Local 1983. Turn-competitive incomings. *Journal of Pragmatics*, 7: 17–38.

Garfinkel, H. 1963. A conception of, and experiments with, 'trust' as a condition of stable concerted actions. In O. J. Harvey (ed.) *Motivation and Social Interaction*. New York: Ronald Press, pp. 187–238.

1967. *Studies in Ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.

Goodwin, C. and A. Duranti 1992. Rethinking context: an introduction. In A. Duranti and C. Goodwin (eds.) *Rethinking Context. Language as an interactive phenomenon*. Cambridge University Press, pp. 1–42.

Gumperz, J. J. 1982. *Discourse Strategies*. Cambridge University Press.

1992. Contextualization revisited. In Auer and di Luzio 1992, pp. 39–53.

Gumperz, J. J. and N. Berenz 1993. Transcribing conversational exchanges. In J. A. Edwards and M. D. Lampert (eds.) *Talking Data: transcription and coding in discourse research*. Hillsdale, N.J.: Lawrence Erlbaum, pp. 91–121.

Gussenhoven, C. 1983. *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Halliday, M. A. K. 1967. *Intonation and Grammar in British English*. The Hague: Mouton.

1970. *A Course in Spoken Intonation*. London: Oxford University Press.

Henderson, E. J. A. (ed.) 1971. *The Indispensable Foundation. A selection from the writings of Henry Sweet*. London: Oxford University Press.

Heritage, J. 1984. *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.

1989. Current developments in conversation analysis. In D. Roger and P. Bull (eds.) *Conversation. An interdisciplinary perspective*. Clevedon: Multilingual Matters, pp. 21–47.

Jefferson, G. 1989. Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In D. Roger and P. Bull (eds.) *Conversation. An interdisciplinary perspective*. Clevedon: Multilingual Matters, pp. 166–196.

Kelly, J. and J. K. Local 1989a. On the use of general phonetic techniques in handling conversational material. In D. Roger and P. Bull (eds.) *Conversation. An interdisciplinary perspective*. Clevedon: Multilingual Matters, pp. 197–212.

1989b. *Doing Phonology: observing, recording, interpreting*. Manchester University Press.

Ladd, D. R. 1980. *The Structure of Intonational Meaning. Evidence from English*. Bloomington: Indiana University Press.

1983. Phonological features of intonational peaks. *Language*, 59: 721–759.

Lehiste, I. 1975. The phonetic structure of paragraphs. In A. Cohen and S. G. Nooteboom (eds.) *Structure and Process in Speech Perception*. Berlin: Springer, pp. 195–203.

Levinson, S. C. 1983. *Pragmatics*. Cambridge University Press.

Liberman, M. and I. Sag. 1974. Prosodic form and discourse function. *Chicago Linguistic Society*, 10: 416–427.

Local, J. K. 1992. Continuing and restarting. In Auer and di Luzio 1992, pp. 272–296.

Local, J. K. and J. Kelly 1986. Projection and 'silences': notes on phonetic and conversational structure. *Human Studies*, 9: 185–204.

Local, J. K., J. Kelly and W. H. G. Wells 1986. Towards a phonology of conversation: turn-taking in Tyneside English. *Journal of Linguistics*, 22: 411–437.

Local, J. K., W. H. G. Wells and M. Sebba 1985. Phonology for conversation: phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics*, 9: 309–330.

McLemore, C. 1991. *The pragmatic interpretation of English intonation: sorority speech*. Ph.D. diss., University of Texas at Austin.

Müller, F. E. 1993. Das audiovisuelle Defizit der cartesianischen Linguistik. *Zeitschift für Literaturwissenschaft und Linguistik*, 90/91: 157–177.

Nespor, M. and I. Vogel 1986. *Prosodic Phonology*. Dordrecht: Foris.

O'Connell, D. C. and S. Kowal 1990a. A note on time, timing, and transcriptions thereof. *Georgetown Journal of Languages & Linguistics*, 1(2): 203–208.

1990b. Some sources of error in the transcription of real time in spoken discourse. *Georgetown Journal of Languages & Linguistics*, 1.4: 453–466.

O'Connor, J. D. and G. F. Arnold 1973. *Intonation of Colloquial English* (2nd edn). London: Longman.

Pierrehumbert, J. B. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. diss., Cambridge, Mass.: MIT.

Pierrehumbert, J. B. and J. Hirschberg 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen and J. Morgan (eds.) *Intentions in Communication*. Cambridge, Mass.: MIT Press, pp. 271–311.

Pike, K. L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Publications.

Psathas, G. (ed.) 1979. *Everyday Language. Studies in ethnomethodology*. New York: Irvington.

Psathas, G. and T. Anderson 1990. The 'practices' of transcription in conversation analysis. *Semiotica*, 78: 75–99.

Sacks, H. 1984. Notes on methodology. In Atkinson and Heritage 1984, pp. 21–27.

1992. *Lectures on Conversation*. 2 vols., ed. E. A. Schegloff and G. Jefferson. Oxford: Blackwell.

Sacks, H., E. A. Schegloff and G. Jefferson 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50: 696–735.

Schegloff, E. A. 1979a. The relevance of repair to syntax-for-conversation. In T. Givón (ed.) *Discourse and Syntax*. New York: Academic Press, pp. 261–296.

1979b. Identification and recognition in telephone conversation openings. In G. Psathas 1979, pp. 23–78.

1982. Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. In D. Tannen (ed.) *Analyzing Discourse: Text and talk*. Georgetown University Round Table on Languages and Linguistics 1981. Washington, D.C.: Georgetown University Round Table on Language and Linguistics, pp. 71–93.

Schegloff, E. A. and H. Sacks 1973. Opening up closings. *Semiotica*, 8: 289–327.

Schegloff, E. A., G. Jefferson and H. Sacks 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53: 361–382.

Schenkein, J. (ed.) 1979. *Studies in the Organization of Conversational Interaction*, New York: Academic Press.

Selkirk, E. O. 1984. *Phonology and Syntax. The relation between sound and structure*. Cambridge, Mass.: MIT Press.

Selting, M. 1987a. *Verständigungsprobleme. Eine empirische Analyse am Beispiel der Bürger-Verwaltungs-Kommunikation*. Tübingen: Niemeyer.

1987b. Reparaturen und lokale Verstehensprobleme – oder: zur Binnenstruktur von Reparatursequenzen. *Linguistische Berichte*, 108: 128–149.

1992a. Prosody in conversational questions. *Journal of Pragmatics*, 17: 315–345.

1992b. Intonation as a contextualization device: case studies on the role of prosody, especially intonation, in contextualizing storytelling in conversation. In Auer and di Luzio 1992, pp. 233–258.

1993. Phonologie der Intonation. Probleme bisheriger Modelle und Konsequenzen einer neuen interpretativ-phonologischen Analyse. *Zeitschrift für Sprachwissenschaft*, 11(1): 99–138.

1995. *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer.

Sherzer, J. and A. C. Woodbury (eds.) 1987. *Native American Discourse. Poetics and rhetoric*. Cambridge University Press.

Silverstein, M. 1976. Shifters, linguistic categories, and cultural description. In K. H. Basso and H. A. Selby (eds.) *Meaning in Anthropology*. Albuquerque: University of New Mexico Press, pp. 11–55.

1992. The indeterminacy of contextualization: when is enough enough? In Auer and di Luzio 1992, pp. 55–76.

Sweet, H. 1877. *A Handbook of Phonetics*. Oxford: Clarendon.

Tannen, D. 1984. *Conversational Style*. Norwood, N.J.: Ablex.

1989. *Talking Voices. Repetition, dialogue, and imagery in conversational discourse*. Cambridge University Press.

Uhmann, S. 1991. *Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie*. Tübingen: Niemeyer.

Vološinov, V. N. 1976. Discourse in life and discourse in art (concerning sociological poetics). In *Freudianism. A marxist critique*, transl. by I. R. Titunik. New York: Academic Press, pp. 93–116.

Wichman, A. 1991. A study of up-arrows in the Lancaster/IBM spoken English corpus. In S. Johansson and A.-B. Stenström (eds.) *English Computer Corpora. Selected papers and research guide*. Berlin: Mouton de Gruyter, pp. 165–178.

Wootton, A. J. 1989. Remarks on the methodology of conversation analysis. In D. Roger and P. Bull (eds.) *Conversation: An interdisciplinary perspective*. Clevedon: Multilingual Matters, pp. 238–258.

Yule, G. 1980. Speakers' topics and major paratones. *Lingua*, 52: 33–47.

Zimmerman, D. H. and C. West (eds.) 1980. Language and Social Interaction. Special double issue, *Sociological Inquiry*, 50/3–4.