1. Intonation and Discourse: Current Views from Within

ELIZABETH COUPER-KUHLEN

Subject Linguistics » Discourse Analysis
DOI: 10.1111/b.9780631205968.2003.00002.x

0 Introduction

In a millennium year we can expect increased stock-taking of the sort: where have we come from? Where are we now? Where do we go from here? The present contribution is an attempt to do this kind of stock-taking with respect to intonation and discourse. It consists of three millennialistic views organized temporally, starting with the view backwards, then the view of today, and finally a view of the future, near and far. Needless to say, all of these temporal viewings have their reference point at the moment of speaking, that is "now." Moreover, they are the author's views: they are anchored deictically to one researcher in the field.[1] Although it is difficult to avoid this natural bias, an adjunct like "from within" can at least recognize it as such.

1 Looking Back

What was the state of the art in the field of intonation and discourse a quarter of a century ago? Actually there was no such field. At that time most linguists felt that it was possible to have language without intonation and therefore to do linguistics without it. In fact, some even thought it imperative to think of intonation, like phonetics, as being outside of language. Not only do we have influential articles, like Bolinger's entitled "Around the edge of language" (1964), to remind us of this; it was (and still is) reflected institutionally in the fact that many renowned British universities had (and have) departments of "Linguistics and Phonetics", the latter subsuming the study of intonation.

Where did this idea come from? First, it was clearly promoted by the bias toward written language which has dominated much of twentieth-century linguistics. The fact that writing works perfectly well without intonation seems to bear out the proposition that we can do without it, and Occam's razor suggests we should. Moreover, the idea found nourishment in the competence-performance dichotomy of the generative paradigm in linguistics. Intonation was easy to relegate to the domain of performance because it only made itself apparent when language was used orally. Finally, *pace* Trager and Smith (1957), intonation did not fit very well into the structuralist mould of thinking anyway. Despite Halliday's (1967) efforts to adduce as much evidence as possible for its distinctive function, there were simply too many occasions when it appeared to be gradient rather than categorical. In fact, this was one of Bolinger's main reasons for saying that it was "around the edge of language," and it was Martinet's (1962) justification for excluding intonation from the functional system of language altogether.

So not only was intonation some thirty years ago a linguistic citizen with dubious credentials, if any at all.[2] Certainly no one had ever thought of combining the notion of intonation with that of discourse. Intonation was the difference between a sentence of written prose and that sentence read aloud. It was what you had when prose was spoken (see also Abercrombie 1965). This surely had nothing to do with discourse – or if it did, the connection was trivial, since discourse was merely a concatenation of sentences and each of these

could be given an intonation on independent grounds.

The change has come slowly but surely. By the 1980s it was beginning to be apparent to some linguists that there might be a discourse function of intonation which would merit investigation (see inter alia Couper-Kuhlen 1986).[3] Brazil, Coulthard, and Johns's *Discourse Intonation and Language Teaching* (1980) was instrumental in bringing about this realization. Significantly the impulse to look at intonation in discourse came from language teachers (or rather, teachers of language teachers). In fact, this was the motivation for most of the early work done on English intonation: Armstrong and Ward's *Handbook of English Intonation* (1926), O'Connor and Arnold's *Intonation of Colloquial English* (1961), and even Halliday's *A Course in Spoken English: Intonation* (1970) are all didacticized texts intended to supplement the teaching of English pronunciation to foreign students. Small wonder then that it was language teachers who, with the turn to communicative skills in language teaching, were among the first to put intonation in this framework.

2 Looking at Now

What is the state of the art today? First, there has been a major paradigm shift with respect to the role of intonation in language. Few if any linguists today would wish to deny the fact that intonation impacts with language. It is hard to identify a single catalyst in this change of paradigm. Perhaps it is best seen as resulting from a slow accumulation of evidence which at some point reached a critical mass. But among those who waxed most persuasive the names of Bolinger, Halliday, Ladd, and Chafe should not be missing.

Three strands of research in the field of intonation in discourse, growing out of three different methodological approaches, may be identified today, in a state of more or less peaceful coexistence.[4] First there is the school of thought which sees intonation as a part of *grammar* broadly speaking.[5] This school actually has quite a tradition. Historically some of the earliest work on intonation tried to establish a correspondence between declarative, interrogative, and exclamatory sentence types and final falling or rising intonation (Couper-Kuhlen and Selting 1996). And there may even be some linguists who still think along these lines. But where speech act theory has been received, those who wish to see intonation as part of grammar will now usually assume that intonations are illocutionary-force-indicating devices and distinctive in the way they pair with different illocutions.

On the American scene, Pierrehumbert's model of intonation nominally belongs in this tradition;[6] it sets up a "grammar" of intonation, with an inventory of six tones or pitch accents, two phrasal tones, and two boundary tones and claims that all well-formed tunes can be generated from this inventory (Pierrehumbert 1980). Recently the intonation-as-grammar approach has addressed the "meaning of intonational contours in the interpretation of discourse" (Pierrehumbert and Hirschberg 1990). The tack taken is to see intonational contours as specifying a relationship between propositional content and the mutual beliefs of participants in the current discourse. One representative study, for instance, attempts to show a context-independent correspondence between a fall-rise pitch accent (L*+H L H%) and a propositional attitude of uncertainty (Ward and Hirschberg 1985; see also Hirschberg and Ward 1992). Here – as in general in the intonation-as-grammar approach – the term "discourse" is used on the grounds that test sentences are read out "in context," as follow-ups to prior sentences which are said to provide a "discourse context" for the interpretation in question.

In a second and no less lively tradition, intonation is thought of as related not to grammar but to *information flow*, the movement of ideas into and out of active, semi-active and inactive states of consciousness. In Chafe's work (1979, 1980, 1993), for instance, intonation is said to provide a window on consciousness via the establishment of two different types of unit: the intonation unit and the accent unit. The intonation unit encompasses the information that is in the speaker's focus of consciousness at a given moment (1993: 39); the accent units are the domains of activation for new, accessible and/or given information. Also within this tradition, Du Bois et al. (1992, 1993) have elaborated the notion of transitional continuity between one intonation unit and the next, marked by different sorts of terminal pitch contours. The term *transitional continuity* describes the extent to which "the discourse business at hand will be continued or has finished" (1993: 53). Thus, depending on whether some material is segmented into one or, say, two intonation units and on how these intonation units are linked transitionally to one another, claims can be made about its status in consciousness and about whether it is viewed as completed or not.

In contrast to the intonation-as-grammar approach, the intonation-and-information-flow approach has paid

less attention to type of pitch accent and more attention to issues of unit segmentation and inter-unit continuity. Methodologically – also in marked contrast to the intonation-as-grammar school of thought – it has developed out of close observation of real discourse rather than from introspection and constructed examples. At times, the discourse under observation in the intonation-as-information-flow tradition has been prompted by an experimental set-up (for instance, the Pear Story film in Chafe 1979 or an instructional task e.g. in Swerts and Geluykens 1994). And it has tended to be primarily monologic as well as uniform in genre (e.g. oral narration, instructional monologue). In this sense the information-flow approach is different from the third school of thought, which takes a deliberately interactional approach.

The third approach might be called provisionally the intonation-as-contextualization approach, to make it comparable with its contemporaries. It is complementary, rather than contrastive, to the intonation-as-information-flow approach but stands in stark contrast to the intonation-as-grammar school of thought. The idea of contextualization goes back to seminal work by the anthropologist Bateson (1956, 1972). But it was first applied specifically to language and intonation in the second half of the 1970s (Cook-Gumperz and Gumperz 1976). Contextualization refers to the fact that linguistic signs need embedding in a context in order to be fully interpretable. In this sense *all* linguistic signs are indexical, not just a small subset of them. Contexts are not given but are said to be invoked, or made relevant, by participants through so-called contextualization cues. The cues may be verbal or nonverbal in nature: they include such stylistic uses of language as code-switching as well as gestural, proxemic, paralinguistic, and prosodic phenomena which accompany linguistic forms (see also Auer and di Luzio 1992). Contextualization cues function by indexing or evoking interpretive schemas or frames within which inferential understanding can be achieved (Gumperz 1982; Tannen 1993). Intonation – by its very nature nonreferential, gradient, and evocative – is seen as a prime contextualization cue in this approach.

Yet intonation – in the restricted sense of "pitch configuration" – rarely functions alone to cue an interpretive frame. The same frame may be cued by timing and volume as well. In fact, frames are cued best (most reliably) when their signals are multi-faceted and come in clusters (Auer 1992). Pitch, volume, and timing have in common that they are prosodic: syllable-based auditory effects produced by vocal-fold and air-flow manipulations orchestrated in time (Crystal 1969). This is why in the contextualization-cue approach there has been a subtle shift away from the study of "intonation" to the study of *prosody* and discourse. The third school of thought thus actually deserves to be called "prosody-as-contextualization cue."

In this approach contextualization cues, and consequently prosodic phenomena, are not seen as accidental or aleatory, nor as automatic reflexes of cognitive and affective states. They are thought to have their own systematicity, but a systematicity which can only be accessed in a context-sensitive fashion. This is why, methodologically, the contextualization-cue approach advocates situated empirical investigation of naturally occurring spoken data. To complement the intonation-as-information flow approach, it focuses less on monologue and more on interaction. In fact, prosodic contextualization research is *grounded* in verbal interaction. This has important consequences for the type of claim made and for the way in which the claims are warranted.

What do prosodic contextualization cues signal in discourse? Viewed from the perspective of interaction, prosodic phenomena can be thought of as furnishing a *format design* for turns at talk. This format design helps interactants meet two general sorts of requirement, which Goffman (1981) has dubbed "system requirements" and "ritual requirements." "System requirements" refer to "requirements that an interaction system must have, given that the participants have certain anatomical, physiological and information-processing capacities"; "ritual requirements" involve "rules that govern interaction, given that the participants are moral beings who are governed by reciprocally held norms of good or proper conduct" (Kendon 1988: 31f)- In other words, prosodic contextualization cues help interactants make inferences about turn-taking and floor management, on the one hand, and about what actions or activities are being carried out, how they are being carried out, and how this might impinge upon participants' face, on the other.

How does one warrant claims about prosodically cued interactional meaning? Here the groundedness of the contextualization-cue approach affords a built-in methodology. The local display which interactants provide to each other of how they have understood a prior turn and of what action is conditionally (or preferentially) relevant in a next turn can be exploited for warranting claims about prosodic signalling in interaction. That is, by viewing prosody as sequentially embedded in interaction, as occasioned by prior actions and occasioning subsequent actions, both embodied in turns with specific prosodic designs themselves, we can develop grounded hypotheses about what its function is from the interactional data and at the same time

validate these hypotheses in the interactional data. This is the contextualization-cue paradigm for the study of prosody in discourse (see also Couper-Kuhlen and Selting 1996).

3 Looking Ahead

As work in this paradigm is just getting under way, it is only appropriate to place the following remarks under the heading of the future, albeit it should be thought of as the near future. What substantial gains in the study of prosodic contextualization can be anticipated over the next few years? The answer to this question will be influenced by the extent to which new territory can be explored. Some of this new territory lies beyond the intonation phrase, and some lies beyond intonation altogether. In the following, single-case analyses from these new territories will be used to show what kind of discovery can be expected with more systematic investigation.

3.1 Beyond the intonation phrase

As soon as one's perspective switches from the individual intonation phrase and events within it to sequences of intonation phrases – which is what should naturally happen in the study in discourse – then the question becomes: are all intonation units alike, merely juxtaposed in time, or are there differences between them? If there are differences, what is their effect? Do they create global intonational structure?

The groundwork for studying intonational structure beyond the intonation phrase has been laid by Chafe (1988), Schuetze-Coburn et al. (1991), and Du Bois et al. (1993). In particular, the notion of declination unit ('t Hart et al. 1990) – which, as Schuetze-Coburn et al. (1991) show, can be identified in naturally occurring discourse as well as in the laboratory – suggests one answer to the question of global intonational structure. Declination units create structures larger than the intonation unit. When there are several intonation units in a declination unit, they have slightly different shapes, depending on their relative position in the larger structure. The position of a single intonation unit within the larger unit is detectable in its final pitch, but also -importantly – in its initial pitch. It is the way intonation units *begin* which forms one of the new territiories for exploration beyond the intonation phrase.

3.1.1 Onset level

The notion of structure created by intonation phrase beginnings can be operationalized with the category of onset level (Brazil's "key"; see also Couper-Kuhlen 1986). The *onset* of an intonation phrase in English is defined as the first pitch accent in the phrase. If there is only one pitch accent, the onset is identical with the so-called nucleus, usually defined as the last pitch accent of the phrase. Brazil et al. (1980) suggest that at least three different onset levels can be identified in speech: High, Mid, and Low. These are to be thought of as pitch levels relative to that of a nucleus or onset in the prior intonation phrase. In the absence of a prior intonation phrase, they are presumably related to the speaker's default pitch range (which is itself related to that speaker's natural voice range: see below). Brazil has argued that the three different onset levels or keys have distinctive functions in discourse. Yet this statement is based more on introspection and carefully chosen constructed examples than on the analysis of large quantities of naturally occurring data. Whether indeed three levels are relevant in everyday conversational interaction is an empirical question which is still open at this time. Should conversationalists operate with only two, the following fragments suggest that an appropriate labeling might be High and Nonhigh.

In interaction there are two possible domains within which an intonational or a prosodic phenomenon may be relevant: (1) the turn or (2) a sequence of turns. In the first, a prosodic phenomenon makes itself apparent relative to surrounding prosody within a speaker's turn; in the second, a prosodic phenomenon is apparent relative to the prosody of a prior or subsequent turn, i.e. *across* speaker turns. Onset level is deployed in both domains by conversationalists, as the following extract demonstrates:

(1) **Kilimanjaro**

(*Ann and her boyfriend Chuck have returned for a visit to Minnesota and are having supper with Ann's high-school friend, Janet, and her husband Steve. Prior talk has centered on nature trips in the Upper Peninsula (U. P.) of Michigan. Ann is talking here about mountain treks in Scandinavia.*)

1 A: there's some sort of rule though (there)

A: there's some sort of idle though (there)
when- when you're in a cabin,
no (gh) in Sweden
when you're in a cabin and someone comes?
next day you have to leave.
but other-
if no one comes
you can stay there as long as you want to.
(.)
so
it's just (like)
to get-
J: right
to keep the process -
S: yeah
(probably right)
J: going
so someone doesn't have to ski for t(h)en days,
heh heh heh
A: oh ho [ho ho ho
J: [without sleep
looking for the only open cabin,
A: No you end up with a lot of people going camping.
but uh
(.)
J: °mhm°
(.)
J: (acc) yeah that sounds nice.
→ There is a place like that in the U. P.;
uhm
Porcupine Mountains.
but they have cabins:
up the mountain
and you can hike
from one cabin
and the next and
(.)
S: [°yeah°
J: [perhaps this fall
we'll go do that
S: °yeah that'd be nice°
J: °yeah°
A: °in the fall°
°mmm°
J: shouldn't be very crowded then at all
(1) it wasn't crowded when we were there
A: heh heh heh
J: no:
A: mmm
J: nothing: in the U. P.;
(.)

```
→      A: Jane'll be hiking in the KilimanTjaro next week
       J: (1)wo::w
          (.)
   55  A: mhm
          °poor Jane
          should've seen her when she went back°
          (.)
          °she had so: much stuff with he(h)r°
   60  J: yeah,
          (.)
          this is a friend from college
          that was teaching in Du:sseldorf
          for:: how long;
   65     [four years?
```

Focusing on Janet's turn beginning in line 28, we notice that the first intonation phrase *yeah that sounds nice* has fast speech rate and begins relatively low in her pitch range. The low-pitched onset becomes particularly noticeable when it is contrasted with the next intonation phrase in line 29: *There is a place like that in the U. P.* Here the first pitch accent on *place* is noticeably higher than the first accent on *yeah* in the prior intonation phrase. (The high onset is indicated in transcription with a capital letter at the beginning of the line; a line which does not begin with a capital letter consequently lacks high onset.) Line 29 is thus a case of high onset being used within the domain of a turn. We identify the high start in relation to one or more other intonation phrases within that same speaker's turn. In the case at hand, since there is a transition relevance point (TRP) at the end of line 28, we might wish to say that lines 28 and 29 form separate turn-constructional units (TCUs). If so, we could then state that the intonational format of the second TCU lends it a different status compared to the first one.

What is the effect of high onset here? A line-by-line analysis of this fragment reveals that the TCUs in lines 28 and 29 are doing rather different things. Line 28 is responsive to the story Ann has just told about staying in mountain cabins in Sweden; its orientation is clearly backwards. Line 29, on the other hand, is more forward-looking. Despite its anaphoric reference with *that* to the place Ann was talking about, its primary business is to introduce a new topic, only tangentally related to the prior one. It puts this new topic *a place in the U. P.* on the floor and at the same time projects more talk about it. The intonational formatting of line 29 can thus be thought of as one of the ways this TCU is designed to do its work: it cues the introduction of a new topic.

Yet, looking somewhat further in the exchange, line 52 is worth considering. Here Ann appears to be introducing a new topic – there has been no mention of either Jane or Kilimanjaro in the forty minutes of talk preceding this fragment – and yet her onset is *not* noticeably higher than the onset of the surrounding intonation phrases.[7] Is this a counterexample to the postulation that new topics are cued with high onset, or is Ann strategically exploiting the contrast between high and nonhigh onset? The evidence suggests the latter. When examined more closely, Ann's new topic will be seen to be qualitatively rather different from Janet's. For one, it has a different sort of trajectory. Janet's TCU (line 29) introduces an entity into the discourse via a presentative construction with *There* is and an indefinite noun phrase *a place like that in the U. P.*, projecting more information on this entity in subsequent TCUs. Ann's TCU (line 52), on the other hand, treats *Jane* as a discourse entity already introduced and accessible, i.e. as common ground, and predicates something about this entity within the same unit. That is, Ann's TCU is constructed and executed as a complete turn of its own.

Second, notice that Janet's new topic receives uptake from all of the participants active in the conversation, whereas Ann's topic is acknowledged only by Janet. Moreover, the nature of Janet's response in line 53 reveals her to be a partially knowing recipient (Goodwin 1981). Were she unknowing, we would expect a response treating the components of Ann's turn – that Jane is or will be in Tanzania, that she will be hiking and that the hiking will be in the Kilimanjaro the following week – as news. Yet as it happens, Janet treats none of these pieces of information as particularly new or surprising. Instead her low-keyed, lengthened wow is heard as registering mild appreciation of something which was (at least partially) already known. That

Janet knows that Jane has recently gone back to Tanzania is, moreover, implicit in the way Ann's next turn is phrased: *should've seen lier when she went back* (line 57) takes both the fact that she returned and where she returned as given.[8]

Third, Ann's follow-up talk on the new topic (lines 56–9) is delivered – in contrast to Janet's (lines 31–6) – sotto voce. And only one of the several participants responds (line 60). Ann's talk is thus insider talk: it is cued for, and receipted by, only a subset of those participating actively in the conversation. Janet's next move confirms this: she unilaterally begins to fill in the unknowing participants, explaining who Jane is and why she has gone to Tanzania (lines 60ff). The evidence thus conspires to suggest that "Jane" is not a full-fledged official topic for the general floor but an insider topic for a private floor. And the prosody of Ann's TCU introducing this topic –specifically its format *without* high onset – can be reconstructed as cueing its unofficial, insider status.

On a more general level, the above fragment demonstrates how participants use high onset and its absence as a strategic resource for cueing new topics. This does not mean that on other occasions high onset or its absence might not signal something different. The inferencing which the deployment of onset level cues must be expected to be sensitive to the sequential location and the verbal content of the TCU in question.

3.1.2 Register

In addition to onset level, there is another aspect of intonation beyond the intonation phrase which cues inferences in interactional discourse. This is *register*, defined as the relative position of an intonation phrase within a speaker's overall voice range (Cruttenden 1986: 129). The norm for register, according to Cruttenden, is for intonation phrases to be positioned roughly in the lower third of a speaker's voice range. Marked uses of register occur when the whole range of pitch configuration within an intonation phrase is moved to a higher, or within limits to a lower, position in the speaker's voice range.[9] Register is distinct from onset level because it affects all the pitches in a given intonation phrase rather than only that of the first accented syllable.

Just as with onset level, register and register shifts are deployed both within the speaking turn and across speaking turns in interaction. Well-known uses within speaking turns include the use of register shift to mark voicing in reported speech (see e.g. Klewitz and Couper-Kuhlen 1999), and the use of register shift to signal that a stretch of speech is parenthetic with respect to primary talk. But register, and more specifically register shift, may also be deployed across speakers' turns, as the next set of examples will demonstrate.

Let us begin by observing the unmarked case of two speakers using the same register in a sequence of turns. The use of the same (as opposed to a different) register by two different speakers is particularly noticeable if everything else in the two turns is held constant – that is, if one speaker is actually doing a repeat of what another speaker has said. For instance:

(2) **Brain Teaser: Fenella McNally**

(*A Radio Picadilly phone-in program in Manchester, where listeners call in with answers to a riddle. M is the moderator, C the caller*.)

```
    1   M:  It is complete;
            though it seems it isn't.
            what do you reckon.
        C:  Well I think I've got this one;
    5       and I got it as you were reading it ou:t.
→           Is the answer ho:le.
            (0.6)
→       M:  Is the answer ho:le.
        C:  yes.
   10   M:  er: no.
        C:  ↑oh!
```

In auditory terms, judging register here involves (1) determining how high the caller's turn *Is the answer ho:le* is in relation to her voice range, (2) determining how high the moderator's repeat *Is the answer ho:le* is in relation to his voice range, and (3) comparing the two relatively. Register comparison across speakers is particularly difficult when the speakers have naturally different voice ranges, as here. However, the fact that the moderator comes off in line 8 as quoting what his caller has just said in line 6 suggests that his TCU is a good rendition of hers and consequently that the relative heights at which they are speaking are similar. Normalized measurements of fundamental frequency will back up this auditory judgment. Figure 1.1 shows a graph of fundamental frequency readings taken every one-tenth of a second for the two turns in question.

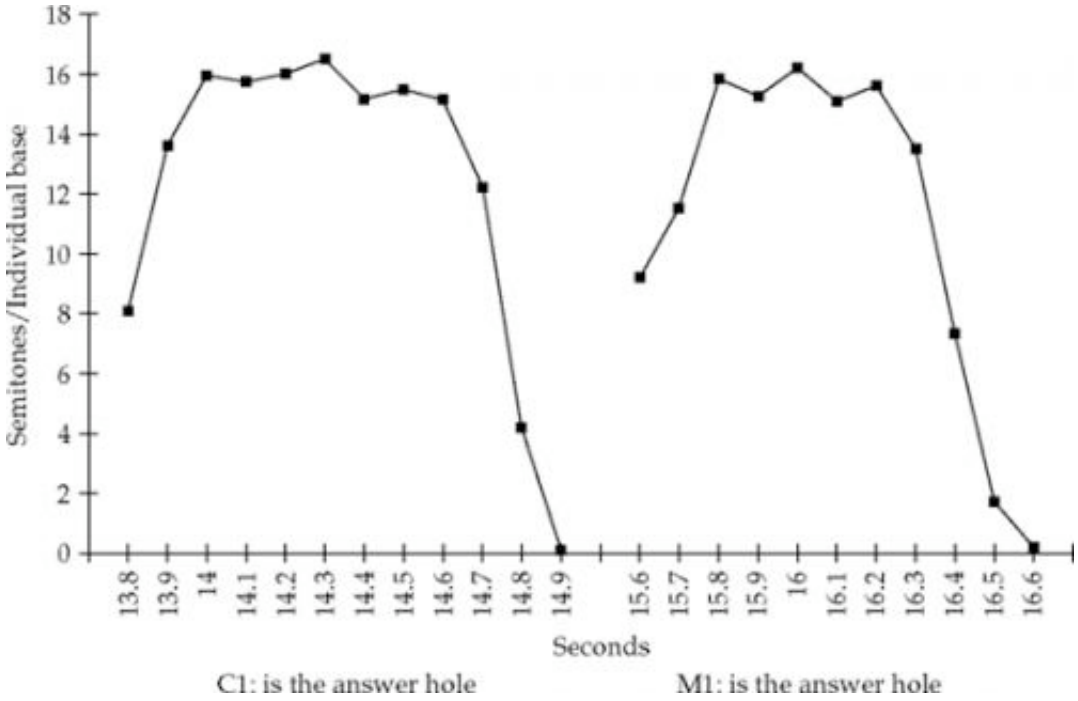

Figure 1.1

In order to normalize the readings and thus make different individual voices comparable, the Hertz values have been expressed here as semitones above the lowest pitch which each speaker is accustomed to use. Seen this way, it is quite obvious that the moderator is speaking at approximately the same height in his voice range as the caller is speaking in her voice range.

Compare now a similar interactional situation where there is a noticeable shift of register in the moderator's repetition of a caller's prior turn:

(3) **Brain Teaser: Julie Salt**

```
1    M:  h you can find reference,
         in any Latin dictionary -
         to a brigade.
     C:  .hh ↑troops!
5        (0.5)
→    M:  {h}↑troops!
         erm
→        {h}↑troops!
         is wrong.
10   C:  oh. hheh
```

Here the fact that the moderator has shifted to an exceptionally high register on *troops* is obvious from comparing it to the prior *you can find reference in any Latin dictionary to a brigade* (lines 1–3) or to the following *erm* (line 7) and *is wrong* (line 9). The normalized f0 curves obtained from acoustic analysis of these turns are shown in figure 1.2.

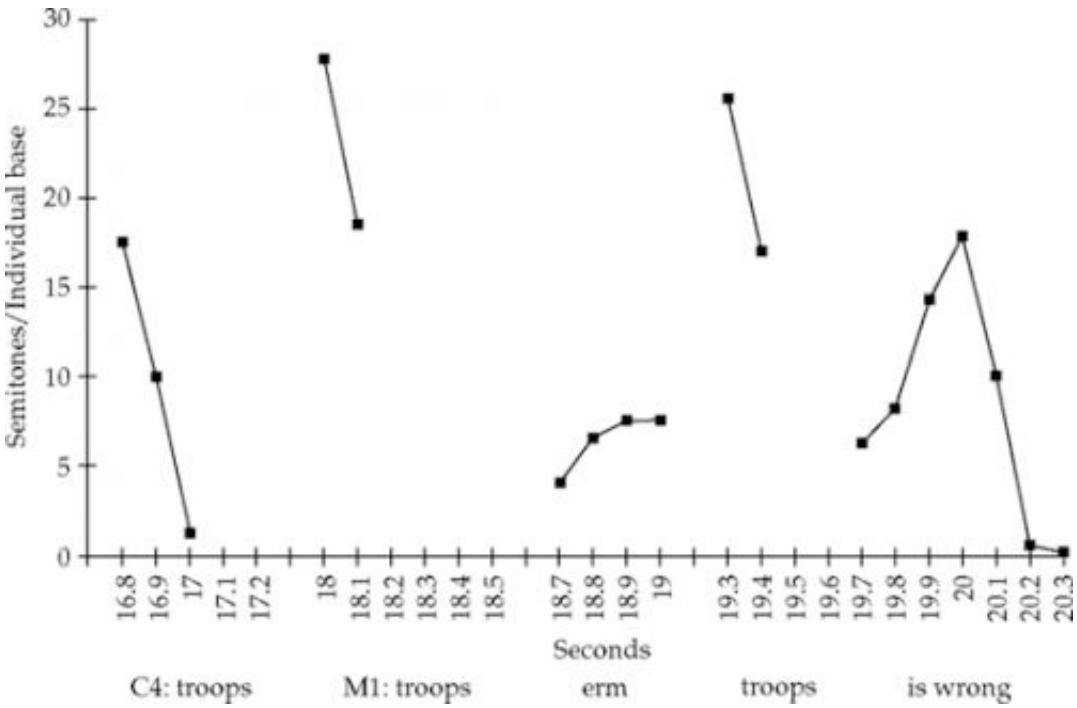

Figure 1.2

Figure 1.2 shows that the moderator is saying *troops* at a point much higher in his voice range than *erm* or *is wrong*. The latter expressions, however, are placed at approximately the same relative height in his voice range as is the caller's *troops* in hers.

What does the moderator cue with this register shift? As argued elsewhere, because he not only shifts his register higher but shifts it to exactly the same *absolute* pitch as his caller, the moderator is heard as mimicking his caller. In doing so, he seems to be subtly (or not so subtly) making a critical comment on the caller's guess – e.g. that it is a silly guess, or that it is delivered in an abnormally high voice (Couper-Kuhlen 1996). Due to the use of absolute pitch, this fragment is thus a special case of register shift. Yet it has in

common with other cases of register shift that it cues special inferences about how talk is being produced and understood.

The exploitation of register across speaking turns is not restricted to guessing sequences nor to shifts to high. Here is a case on the same quiz show where a register shift to low is deployed by the moderator in quite a different context:

(4) **Brain Teaser: Sexy Sharon**

```
    1   M:  then we go to Hardwick. (.)
            and there we get -
            (.) h sexy Sharon.
            ↓hi!
    5   C:  (0.4) °hello° -
→       M:  (1) °hello° -
            how are you Sharon -
        C:  °all right [thanks°
        M:             [oh: ↑cheer up dear,
   10   C:  he hh
        M:  Cheer up;
            for goodness sake;
            don't- don't put me in a bad mood;
            at (.) one o'clock;
```

Focusing on the register of line 6, it will be observed that the moderator's *hello* is noticeably lower than his *sexy Sharon* in line 3. But it is at approximately the same relative height as Sharon's prior *hello* in line 5. This is a case of register shift to low which becomes noticeable across speaking turns by the same speaker. The moderator appears to be shifting to a register closer to that of his caller, as is evident from figure 1.3.

What does this register shift to low cue? Here too the moderator is heard as mimicking his caller and thereby making a critical comment on her turn. But in contrast to the prior example, where one of the messages was "Your voice is so high!," the message now seems to be "Your voice is so low!" This moderator has very definite expectations about his callers' register, especially his female callers. The upwards tendency in the register of his next TCU (*how are you Sharon*), visible in figure 1.3, may be another, more subtle hint to the caller to "raise her voice." If so, this would account nicely for why – when the strategy fails and Sharon continues with low pitch on *all right* (see figure 1.3) – he becomes more explicit in subsequent talk: *cheer up dear* (line 9) and *Cheer up for goodness sake* (lines 11–12).

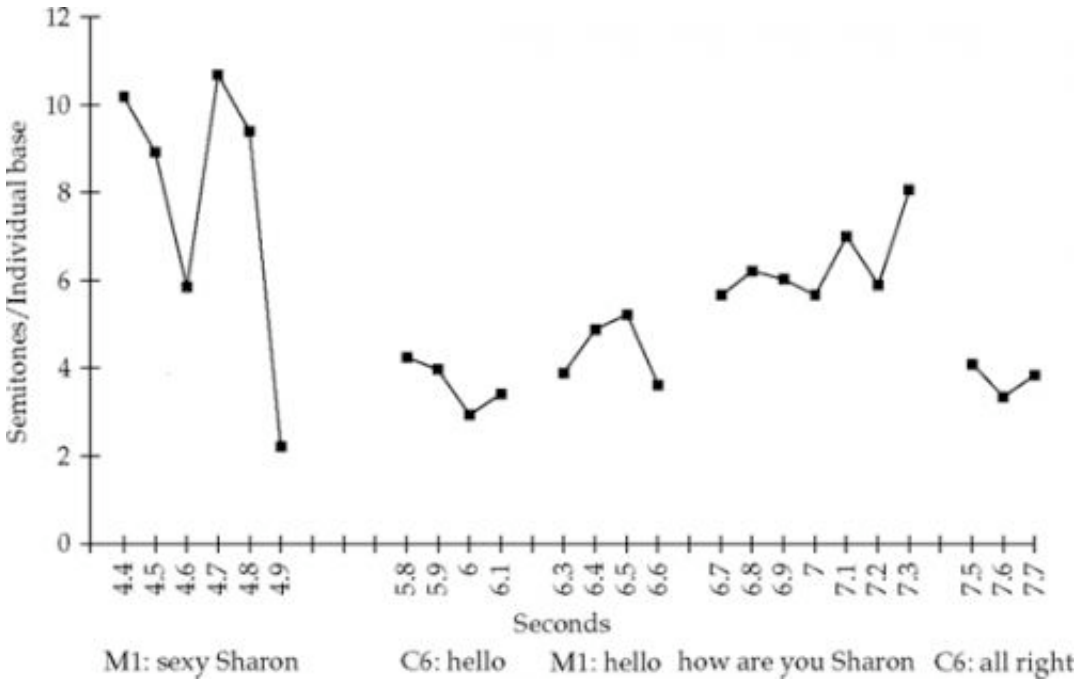

Figure 1.3

On a more general level, the above fragment provides a particularly clear demonstration of the fact that to make sense of what participants do in interaction, it is crucial to take the prosodie design of talk into consideration. Yet if we try to reconstruct why the moderator admonishes Sharon to cheer up, we will discover that there is more than just her pitch that is amiss: the volume and timing of her turn in line 5 are also off. This suggests that to fully understand the contextualization process the perspective must be broadened to include other prosodie phenomena.

3.2 Beyond intonation

A second type of new territory in the field of interactional prosody is that beyond pitch or intonation altogether. The focus here will be on *timing*. Needless to say, all spoken discourse unfolds in time. Moreover, our scientific tradition provides us with objective ways of dividing up time neatly and of measuring it precisely. Yet it is doubtful whether lay speakers *experience* time in interaction in terms of units measured objectively in minutes and seconds. To speak meaningfully about timing in interaction, the metric which is behind participants' subjective judgment of time must be identified. It is this metric which enables them to determine that “now” is the right time for some word or for a turn, and that someone has departed from this right time by pausing or by coming in too early or too late. Erickson and Shultz (1982) have proposed that subjective judgments of experienced time in interaction are made with reference to rhythmic cycles which organize the verbal and nonverbal behavior of participants. And, as Pike (1945), Halliday (1970), and others have pointed out, the basis for rhythm in English is the regular recurrence of accented syllables in time. Thus the hypothesis that *speech rhythm* provides a metric for timing in English interaction seems rather compelling (see also Couper-Kuhlen 1993).

Rhythm in the interactional sense refers to a regular beat which establishes itself in talk through the even placement of accented syllables in time (see Auer et al. 1999). The distance between two, typically adjacent accented syllables creates a temporal interval.[10] When two or more successive temporal intervals are perceived to be approximately equal in duration, the speaker (or speakers) can be said to be speaking rhythmically. Isochronously timed accents create the impression of a regular rhythmic beat in speech. Observation suggests that speakers use the rhythmic delivery of within-turn talk for a variety of structural and rhetorical purposes. And it appears to be the maintenance of a common rhythmic beat across turns at talk which counts as the well-timed option for turn transition in English conversation.

Consider the case of smooth interactional timing, i.e. where turn transition is wholly unremarkable. For instance:

(5) **Brain Teaser: Fenella McNally**

```
    1  M:  let's see how we do in Staleybridge,
           Fenella McNally;
           hi.
→      F:  hello!
→   5  M:  hello: Fenella,
→      F:  hello;
           we spoke last night.
           hehn
```

The first thing to notice about this opening is the fact that the moderator's accents on *see, Staled/bridge, Fenella* and *hi* are timed regularly at the end of his first turn. The rhythmic beat which this timing establishes can be represented notationally as follows:[11]

(5′) Rhythmic analysis of **Fenella McNally** opening

```
1  M:  let's /'see how we do in /
           /'Staleybridge, Fe-/
           /'nella McNally;   /
           /'hi.
```

Fenella now picks up the moderator's rhythmic beat in the next turn by timing her accent on *hello* accordingly. Moreover, the moderator adjusts the timing of his next turn to synchronize with this beat:

(5″) Rhythmic analysis of **Fenella McNally** opening

```
1  M:  let's /'see how we do in /
           /'Staleybridge, Fe-/
           /'nella McNally;   /
           /'hi.
5  F:                    hel-/
           /'lo!
   M:                    hel-/
           /'lo:
```

Moderator and caller collaborate here in the production of a common rhythm which they maintain across speaking turns by picking up in each new turn the beat established in the prior turn.

Now observe what happens in the continuation of line 5 in the orthographic transcript. The moderator shifts the rhythm slightly by placing an accent on *Fenella* which comes sooner than the next expected beat. This creates a number of rhythmic options for the timing of the next turn. (For instance, a next speaker could simply ignore the syncopation and continue according to prior timing. Or a next speaker could miss the next beat altogether, perhaps causing the rhythm to break down.) What this caller opts for, however, is to create a

new, faster rhythmic pattern based on the timing of the moderator's accents on *hello* and *Fenella* by placing her next accents on *hello, spoke* and *night* accordingly. In rhythmic notation this can be shown as follows:

(5″) Rhythmic analysis of **Fenella McNally** opening

```
1   M:  let's /'see how we do in /
              /'Staleybridge, Fe-/
              /'nella McNally;   /
              /'hi.
5   F:                       hel-/
              /'lo!
    M:                       hel-/
              /'lo: Fe-          /  (faster)
              /'nella,
10  F:                       hel-/
              /'lo; we           /
              /'spoke last/
              /'night. hehn
```

The transitions in this exchange can thus be reconstructed as smooth due to the fact that each turn onset is rhythmically well-timed with respect to the prior turn.

Rhythmic coordination of this sort requires a fine sensing of timing on the part of participants. Unaccented syllables before the first accent of a new turn must be timed so that the first accent falls on the beat. Sometimes just a fraction of a second delay is necessary between turns in order to make the synchronization work. In fact, there are tiny micropauses at each of the transitions here, which suggests that speakers are timing their turn onsets rhythmically. In other words, they are not coming in at the earliest possible moment in time but at the earliest possible *rhythmic* moment in time. The micropauses are scarcely noticeable because they help maintain the regular rhythm rather than destroy it.

Now examine a case where transition timing is less successful:

(6) **Brain Teaser: Sexy Nora**

```
    1   M:  so I think we'll kick off;
            with er -
            sexy Nora;
            who lives in Heaton Chapel.
    5       hi!
→       N:  (0.7) hi.
        M:  hi!
            how are you Nora?
        N:  oh hello. heh
    10  M:  he- hello,
        N:  hello!
        M:  hello!
            you're on the radio!
        N:  well that was a surprise.
    15  M:  surprise surprise.
```

In this opening the moderator also provides his caller with a clear rhythmic beat at the end of his first turn by regularly timing his accents on *sexy, lives*, and *hi*.[12] But she misses his cue. Her *hi* in line 6 is too late to coincide with the beat he has established:

(6′) Rhythmic analysis of **Sexy Nora** opening

```
1  M:  so I 'think we'll kick off;
       with er -
       /'sexy Nora; who          /
       /'lives in 'Heaton 'Chapel./
5      /'hi!
   N:  (0.7) 'hi.        (late)
```

As the subsequent development of talk here shows, the fact that Nora misses the moderator's cue creates a minor interactional "incident": the greeting sequence gets recycled twice, and accounts are offered on both sides for what has happened – *you're on the radio* (line 13) and *well that was a surprise* (line 14). Thus the hitch in turn transition in (6) can be reconstructed as rhythmic ill-timing: the caller's return of greeting is late with respect to the rhythm and timing established in prior talk.[13]

An appreciation of how crucial minor timing mishaps in turn transition can be for the order of interaction now casts a new light on what happened in fragment (4):

(4) **Brain Teaser: Sexy Sharon**

```
    1   M:  then we go to Hardwick. (.)
            and there we get -
            (.) h sexy Sharon.
            ↓hi!
    5   C:  (0.4) °hello° -
→       M:  (1) °hello° -
            how are you Sharon -
        C:  °all right [thanks°
        M:             [oh: ↑cheer up dear,
    10  C:  he hh
        M:  Cheer up;
            for goodness sake;
            don't- don't put me in a bad mood;
            at (.) one o'clock;
```

A rhythmic analysis of this opening reveals that Sharon too misses the timing cues in the moderator's first turn. He sets up a well-defined rhythm with accents on *sexy, Sharon*, and *hi*, but she comes in too late:

(4′) Rhythmic analysis of **Sexy Sharon** opening

```
1  M:  and there we get -
       (.) h /'sexy    /
             /'Sharon. /
             /'↓hi!
5  C:  (0.4) °hel'lo° - (late)
```

In sum, it is the fact that transition timing is off as much as the fact that Sharon's pitch is perceived as low which cues the moderator's inference that she is not cheerful. This fragment thus provides a concrete example of how prosodie contextualization cues cluster and jointly make interpretive frames relevant.

What provisional conclusions can be drawn about the way prosodie contextualization cues – here: onset, register, and rhythm – work in discourse? Onset and register have in common that they work to create a rudimentary sort of global structure: both are ways to format a TCU such that it will be heard as either prosodically matching or prosodically contrasting with surrounding TCUs. If matching, this may be interprétable structurally as, roughly speaking, continuing something that has already been started; if contrasting, it may be interprétable as doing something which is disconnected from what has gone before. Where the shift is to high, the structural inference may be that something new is beginning; where it is to low, that something is being subordinated. (On occasion, where sequential location and verbal content make a particular register or onset formatting expectable for a given TCU, the strategic avoidance of that format will cue the opposite interpretation.) Rhythm on the other hand is more of an equalizer: it pulls together units of different sizes and scope in an integrative fashion and sets them off from parts of surrounding talk which are rhythmically nonintegrated or which are patterned differently. What all three prosodie contextualization cues appear to have in common, however, is that they can have a structural (i.e. "system"-related) or an actional (i.e. "ritual"-related) interpretation, depending on the sequential context in which they occur and the syntactic-semantic content of the TCUs they are designed for.

4 Looking Far Ahead

To conclude, what are some of the directions prosodie research might take in the more distant future?

First, as the analysis of fragment (4) above suggests, *volume* needs to be looked at more closely. It will very likely turn out to be a prosodie contextualization cue like intonation and timing which is locally invoked and strategically deployed both within and across speaking turns. Just as with pitch, where the declination unit defines upper and lower gridlines within which pitch events are located, so a loudness declination unit will arguably need to be postulated within which loudness events are located (see also Pittenger et al. 1960 and Laver 1994). Whether loudness declination is coextensive with pitch declination is an open question. Moreover, how loudness declination is handled across turns requires investigation: Goldberg (1978) suggests that amplitude may shift or reset at structural points in discourse organization just as pitch has been shown to do.

Second and more significantly, paralinguistic voice-quality effects require investigation (see also Pike 1945; Trager 1958; Pittenger et al. 1960). This step of course goes not only beyond the intonation phrase and beyond intonation but beyond prosody altogether. Yet it is a logical step if one's goal is to reconstruct the vocal cues which contextualize language. Just as the same interpretive frame can be cued by pitch and timing at once, so it can also be cued by paralinguistic voice quality. *Voice quality* has often been thought of as resulting from the natural or habitual setting of laryngeal and supralaryngeal musculature in the vocal tract (Laver 1980). Yet speakers can and do assume different voice qualities at will. Some of those which appear to be deployed strategically in everyday English conversation are nasal voice, breathy voice, creaky voice, “smiley” voice, whisper, and falsetto. Others can and surely will be found on closer investigation. Here too the question must be: what resources do speakers have at their disposal? And how are these resources deployed in cueing interaction? The answers must be sensitive to possible sociolinguistic and sociocultural variation, but above all grounded in conversational interaction.

TRANSCRIPTION CONVENTIONS

One line	One intonation phrase
First word capitalized	High onset (= full declination reset)
[Line	
[Line	Overlapped utterances
Line=	
=Line	Latched utterances
Line.	Final pitch falling to low
Line!	Final pitch falling to low from high starting point
Line;	Final pitch falling slightly
Line –	Final level pitch
Line,	Final pitch rising slightly
Line?	Final pitch rising to high
{l}	Low register
{h}	High register
{*acc*}	Accelerando
{*dec*}	Decelerando
↑Word	Noticeable step-up in pitch
↓Word	Noticeable step-down in pitch
Wo::rd	Lengthened sound or syllable
Word–	Cut-off sound or syllable
WORD	Loud volume
°word°	Soft volume
′word	Accent or stress
/′word	/

/ˈword	/
/ˈword	Rhythmic patterning of accents
(h)	Breathiness
(gh)	Gutteralness
.hhh	Inbreath
hhh	Outbreath
(word)	Unsure transcription
(.)	Brief pause
(1.0)	Measured pause

1 I am grateful nonetheless to Wally Chafe, Jack Du Bois, and Sandy Thompson for listening to an early version of this chapter at the Linguistics Colloquium, University of California at Santa Barbara, and talking through the ideas with me. I bear full responsibility for not taking their advice when I should have.

2 Outside of linguistics, on the other hand, it was generally acknowledged as a prime metacommunicative device in face-to-face interaction. See e.g. Bateson et al. (n.d.) and Pittenger et al. (1960) for two early attempts to capture it on paper and describe its import.

3 Menn and Boyce (1982) was an early attempt to link quantified measurements of voice pitch with discourse structure.

4 Excluded from this survey are corpus-linguistic studies of discourse, many of which take intonation into consideration without making it the focus of investigation.

5 “Grammar” being understood loosely enough to include speech acts.

6 As does a fortiori Steedman (1991).

7 Nor is Ann's onset in line 52 as high as in line 23, where she is perceived as starting high.

8 Subsequent talk confirms that Janet knows not only that Jane has recently gone back to Tanzania but also why.

9 In addition, some analysts recognize the narrowing or widening of a speaker's register as significant departures from the norm (see Pittenger et al. 1960).

10 Occasionally nonadjacent accented syllables also mark off rhythmic intervals; see (6) below for an example of this.

11 Left-hand slashes are placed before the accented syllables creating a rhythmic beat and are aligned underneath one another on the page to indicate regular timing. Right-hand slashes give a rough indication of tempo, or how close together/far apart the beats come in time.

12 Notice that the accents on *Heaton Chapel* are disregarded in the interest of a higher-level rhythmic pattern created by the regular timing of accents on *sexy, lives*, and *hi*.

13 It is true that Fenella was probably on hold, waiting for her call to be put through, and that unpreparedness may account for why she misses the moderator's cue. Yet since presumably all callers to the show are put on hold, this fails to explain why the large majority of them have no trouble at all following the moderator's cue. In most calls a regular rhythm is established across speaking turns from the very beginning.

REFERENCES

Abercrombie, D. (1965). *Conversation and spoken prose* . In D. Abercrombie (ed.), *Studies in Phonetics and Linguistics* (pp. 1 9). Oxford: Oxford University Press .

Armstrong, L. E., and Ward, I. C. (1926). *A Handbook of English Intonation* . (Second edition, 1931). Cambridge: W. Heffer and Sons .

Auer, P. (1992). *Introduction: J. Gumperz' approach to contextualization* . In P. Auer and A. d. Luzio (eds), *The Contextualization of Language* (pp. 1 38). Amsterdam: Benjamins .

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

Auer, P., Couper-Kuhlen, E., and Mueller, F. (1999). *Language in Time: The Rhythm and Tempo of Spoken Interaction* . New York: Oxford University Press .

Bateson, G. (1956). *The message "This is play"* . In B. Schaffner (ed.), *Group Processes* (pp. 145 242). New York: Josiah Macy Jr. Foundation .

Bateson, G. (1972). *A theory of play and fantasy* . In G. Bateson (ed.), *Steps to an Ecology of Mind* (pp. 177 93). New York: Ballatine .

Bateson, G., Birdwhistell, R., McKeown, N., Mead, M., Brasen, H. W., and Hockett, G. F. (n.d.). *The Natural History of an Interview* . Chicago: University of Chicago, unpublished.

Bolinger, D. (1964). *Around the edge of language* . *Harvard Educational Review* , (34) , 282 96.

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

Chafe, W. (1979). *The flow of thought and the flow of language* . In T. Givon (ed.), *Syntax and Semantics* (Vol. 12: Discourse and Syntax, pp. 159 81). New York: Academic Press .

Chafe, W. (1980). *The deployment of consciousness in the production of a narrative* . In W. Chafe (ed.), *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production* (pp. 9 50). Norwood, NJ: Ablex .

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

Chafe, W. (1993). *Prosodic and functional units of language* . In J. A. Edwards and M. D. Lampert (eds), *Talking Data: Transcription and Coding Methods for Language Research* . Hillsdale, NJ: Lawrence Erlbaum .

Cook-Gumperz, J., and Gumperz, J. J. (1976). *Context in children's speech* . Papers on Language and Context, Working Paper No. 46, Language Behavior Research Laboratory , Berkeley.

Couper-Kuhlen, E. (1986). *An Introduction to English Prosody* . London and Tuebingen: Edward Arnold and Niemeyer .

Couper-Kuhlen, E. (1993). *English Speech Rhythm: Form and Function in Everyday Verbal Interaction* . Amsterdam: Benjamins .

Couper-Kuhlen, E. (1996). *The prosody of repetition: on quoting and mimicry* . In E. Couper-Kuhlen and M. Selting (eds), *Prosody in Conversation: Interactional Studies* (pp. 366 405). Cambridge: Cambridge University Press .

Couper-Kuhlen, E., and Selting, M. (1996). *Towards an interactional perspective on prosody and a prosodic perspective on interaction* . In Couper-Kuhlen and M. Selting (eds), *Prosody in Conversation* (pp. 11 56). Cambridge: Cambridge University Press .

Cruttenden, A. (1986). *Intonation* . Cambridge: Cambridge University Press .

Crystal, D. (1969). *Prosodic Systems and Intonation in English* . London: Cambridge University Press .

Du Bois, J. W., Schuetze-Coburn, S., Paolino, D., and Gumming, S. (1992). *Discourse transcription* . *Santa Barbara Papers in Linguistics* , (4) , 1 225.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., and Paolino, D. (1993). *Outline of discourse transcription* . In J. A. Edwards and M. D. Lampert (eds), *Talking Data: Transcription and Coding in Discourse Research* (pp. 45 89). Hillsdale, NJ: Lawrence Erlbaum .

Erickson, F., and Shultz, J. (1982). *The Counselor as Gate Keeper: Social Interaction in Interviews* . New York: Academic Press .

Goffman, E. (1981). *Forms of Talk* . Oxford: Blackwell .

Goldberg, J. A. (1978). *Amplitude shift: a mechanism for the affiliation of utterances in conversational interaction* . In J. Schenkein (ed.), *Studies in the Organization of Conversational Interaction* (pp. 199 218). New York: Academic Press .

Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers* . New York: Academic Press .

Gumperz, J. (1982). *Discourse Strategies* . Cambridge: University Press .

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

Halliday, M. A. K. (1970). *A Course in Spoken English: Intonation* . London: Oxford University Press .

Hirschberg, J., and Ward, G. (1992). *The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English* . *Journal of Phonetics* , (20) , 241 51.

Kendon, A. (1988). *Goffman's approach to face-to-face interaction* . In P. Drew and W. Antony (eds), *Erving Goffman: Exploring the Interaction Order* (pp. 14 40). Oxford: Polity .

Klewitz, G., and Couper-Kuhlen, E. (1999). *Quote-unquote: the role of prosody in the contextualization of reported speech sequences* . *Pragmatics* , (9.4) , 459 85.

Laver, J. (1980). *The Phonetic Description of Voice Quality* . Cambridge: Cambridge University Press .

Laver, J. (1994). *Principles of Phonetics* . Cambridge: Cambridge University Press .

Martinet, A. (1962). *A Functional View of Language, being the Waynflete Lectures* . Oxford: Clarendon .

Menn, L., and Boyce, S. (1982). *Fundamental frequency and discourse structure* . *Language and Speech* , (25.4) , 341 83.

O'Connor, J. D., and Arnold, G. F. (1961). *Intonation of Colloquial English* . (Second edition, 1973). London: Longman .

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation* . PhD thesis, MIT.

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

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

Pittenger, R. E., Hockett, C. F., and Danehy, J. J. (1960). *The First Five Minutes: A Sample of Microscopic Interview Analysis* . Ithaca, NY: Paul Martineau .

Schuetze-Coburn, S., Shapley, M., and Weber, E. G. (1991). *Units of intonation in discourse: a comparison of acoustic and auditory analyses* . *Language and Speech* , (34.3) , 207 34.

Steedman, M. (1991). *Structure and intonation* . *Language* , (67) , 260 96.

Swerts, M., and Geluykens, R. (1994). *Prosody as a marker of information flow in spoken discourse* . *Language and Speech* , (37.1) , 21 43.

Tannen, D. (ed.), (1993). *Framing in Discourse* . New York and Oxford: Oxford University Press .

't Hart, J., Collier, R., and Collier, A. (1990). *A Perceptual Study of Intonation* . Cambridge: Cambridge

University Press .

Trager, G. L. (1958). *Paralanguage: a first approximation* . *Studies in Linguistics* , (13.1) , 1 12.

Trager, G. L., and Smith, H. L. J. (1957). *An outline of English structure* . *Studies in Linguistics: Occasional Papers* , (3) , 40 52.

Ward, G., and Hirschberg, J. (1985). *Implicating uncertainty: the pragmatics of fall-rise intonation* . *Language* , (61) , 747 76.

Cite this article

COUPER-KUHLEN, ELIZABETH. "Intonation and Discourse: Current Views from Within." *The Handbook of Discourse Analysis*. Schiffrin, Deborah, Deborah Tannen and Heidi E. Hamilton (eds). Blackwell Publishing, 2003. Blackwell Reference Online. 13 December 2007 <http://www.blackwellreference.com/subscriber/tocnode?id=g9780631205968_chunk_g97806312059682>

Bibliographic Details

The Handbook of Discourse Analysis

Edited by: Deborah Schiffrin, Deborah Tannen And Heidi E. Hamilton
eISBN: 9780631205968
Print publication date: 2003