

Chapter 12

Prosody

What is most comprehensible in speech is not the word itself, but rather the tone, intensity, modulation, tempo, with which a sequence of words is spoken – in short, the music behind the words, the passion behind this music, the person behind this passion: in other words, everything that cannot be written (Nietzsche, 1980, p. 89, Fragment Nr. 296; cited in Blank, 1991, p. 9; our translation).

Chapter Prospectus

Chapter 12, *Prosody*, emphasizes on-time in spoken discourse for the most part. Variations in the on-time variables of loudness, articulation rate, and intonation, and in the off-time variable of pauses are precisely the elements absent in written texts. Such deletion in the written relative to the spoken makes the written barren in itself and lacking in potential for expressiveness; only the prosodic skill of an expressive reader can bring a written text to life by reading it aloud. The continuous variability of prosody defies transcription into discrete units, even though its variability in spoken discourse constitutes an important determinant of meaning. Researchers are gradually coming to terms with the necessity of considering all these factors of prosodic variability as simultaneously operative, if indeed they are to adequately assess the contribution of prosody to meaning in spontaneous spoken discourse.

The Concept

Merriam-Webster's collegiate dictionary (11th ed., 2003, p. 998) has offered the third – and for our purposes, relevant – meaning of *prosody* as “the rhythmic and intonational aspect of language.” The *APA dictionary of psychology* (American Psychological Association, 2007, p. 742) is more detailed: “A phonological feature of speech, such as stress, intonation, intensity, or duration, that pertains to a sequence of PHONEMES rather than to an individual SEGMENT” and refers the reader further to “PARALANGUAGE” and

“SUPRASEGMENTAL.” Note that the APA definition comes down on the side of phonology (over phonetics) and lists stress independently of intensity. Couper-Kuhlen and Selting (1996b, p. 11) have conceptualized prosody as comprising “the ‘musical’ aspects of speech”: “Auditory effects such as melody, dynamics, rhythm, tempo and pause.” A phonetic definition of prosody has been provided by Kohler (1995, p. 13 f.; our translation), including “variations in pitch, intensity, levels of emphasis, tempo, register, general voice quality.”

In the previous chapter, we have already analyzed the role of pauses in spoken discourse. Here it needs only to be further emphasized that the patterning of on-time and off-time by the frequency and duration of pauses constitutes an important contribution to prosody. In this respect, the generalization – normative rather than empirical – on the part of Duranti (1991, p. 137) and others that silence is out of place in conversation is an unwarranted oversimplification:

In certain kinds of verbal exchanges – what conversation analysts call “conversations” – silence is to be avoided and gaps between turns are to stay as short as possible (cf. Sacks, Schegloff & Jefferson, 1974; Schegloff & Sacks, 1973).

In the context of prosody, the term tempo is also popular. Couper-Kuhlen and Selting (1996b, p. 11; see previous paragraph) have clearly associated tempo with on-time and have distinguished it from pauses. Hence, tempo is here synonymous with articulation rate as traditionally measured in terms of syllables per second (syl/s) of on-time, and it presumes assessment (perceptual or instrumental) of off-time. Just as with the off-time or pause components of prosody, the on-time elements of prosody also work in consort with one another and with pauses to construct the rhythmic and intonation patterns of spoken discourse.

Paradoxically, definitional problems arise far more with respect to adjunct concepts than with respect to the basic physical variables of loudness, pitch, and time. Thus, *stress*, *emphasis*, *prominence*, and *salience* are all in need of clarification. They are used sometimes interchangeably and sometimes with specific meanings exclusive of one another. Perhaps the most opaque of all is *stress*. *Merriam-Webster’s collegiate dictionary* (11th ed., 2003, p. 1235) offers the following subordinate meanings of *stress* relevant to prosody:

4: intensity of utterance given to a speech sound, syllable, or word producing relative loudness **5 a**: relative force or prominence of sound in verse **b**: a syllable having relative force or prominence.

Note that definition number **4** limits *stress* to loudness produced by intensity, whereas definition number **5 b** refers generically to force of a syllable or prominence. Crystal (1997, p. 174) has very clearly limited stress to loudness in at least one context: “In English words, each syllable is pronounced with a certain level of loudness, or stress.” Maynard, Houtkoop-Steenstra, Schaeffer, and van der Zouwen (2002, p. 494) have introduced the term *emphasis* and have made it broader than just loudness: “Emphasis is done with some combination of a changing pitch, rise in volume, stretch of a sound, or stress on a vowel.” In

this case, stress is clearly distinguished from “rise in volume.” We are faced with a confusing array of terminology. Sometimes, the definitions and their examples even get in one another’s way, as in *Collins cobuild advanced learner’s English dictionary* (2003, p. 1433), where *stress* has been defined in terms of a syllable that “sounds slightly louder.” But the corresponding example thereof is precisely an example of equal stress: “‘*Sit down,*’ *she replied, stressing each word,*” so that one must ask, “Louder than what?”

Accent in turn is defined in *Merriam-Webster’s collegiate dictionary* (11th ed., 2003, p. 7) as “**1** : an articulative effort giving prominence to one syllable over adjacent syllables; *also* : the prominence thus given a syllable.” It would seem reasonable to assume that the “prominence” intended here for both *stress* and *accent* can be contributed by a number of prosodic variations, including relative loudness, syllabic prolongation, surrounding pauses, and – paradoxically – diminution of loudness. Even an extra-linguistic variable, e.g., an accompanying gesture, bodily movement, or deixis, can contribute such prominence to a speech sound, syllable, or word.

Intonation too must be defined: “The rise and fall in pitch of the voice in speech” (*Merriam-Webster’s collegiate dictionary*, 11th ed., 2003, p. 656). The richness of this concept is reflected in Crystal’s (1997, p. 173) listing of the many functions of intonation: emotional, grammatical, informational, textual, psychological, and indexical.

A danger in the midst of all this variability is the possibility that the researcher may analyze the cycles per second (cps) of pitch, the decibels (db) of loudness, and/or the milliseconds (ms) of time as isolated physical phenomena, without acknowledgement that they are used by speakers together and in conjunction with extra-linguistic factors for an overall effect on meaning. In this respect, Quirk, Greenbaum, Leech, and Svartvik (1985, p. 1589) have warned us “against simple equations such as regarding stress as identical with loudness. . . other factors are or can be involved – notably duration and pitch.” Unfortunately, this very complexity has provided an occasion for some qualitative analysts to pooh–pooh quantitative analyses as purely physicalistic and irrelevant to the interactional situation. It is hardly our intent to join in this chorus, but rather simply to warn against the ever present danger of reductionism and oversimplification.

Prosody and Meaning

It is a basic fact that the *way* something is *said* may alter what is *meant*. What we will later (see our Chapter 23) refer to as a veritable *somaticization* of the syntax of spontaneous spoken discourse that sometimes transcends, modifies, supplements, or supplants the sentential syntax of well-formed grammatical units is saliently subject to the uses of these prosodic means. For example, in context, a very emphatically uttered masculine third-personal pronoun in the

assertion *he did it* can constitute a firm statement of the innocence of a female suspect for whom the assertion *she did it*, with a correspondingly emphatically uttered feminine third-personal pronoun, would have been appropriate. Her innocence is not so much a logical conclusion that must be inferred from the emphatic *he*. Instead, it is a shift in the very meaning of *he* in this context from simply *he* to *he, not she*. There are innumerable cases of irony, sarcasm, play on words, and other usages in everyday speech that display such *somaticized* syntax. Even if the traditional structural linguists may not be quite ready to acknowledge this determination of meaning in the interaction of spontaneous spoken discourse, it is nonetheless very important psychologically for the understanding on the part of the listener of the meaning intended by the speaker.

The Transcription of Prosody

Couper-Kuhlen and Selting (1996b, p. 11), in their edited volume *Prosody in conversation: Interactional studies* (1996a), which has recently been reprinted in paperback (2006), have deplored the vacuum of research on prosody: “It is surely no exaggeration to state that a large part of this field has been left untilled by modern structural linguistics.” They have ascribed this neglect to the failure to allow “speech prosody and language-in-use” to cross-fertilize one another and have concluded that “it is doubtless the overwhelming influence of literacy on thinking about language which has been responsible for the neglect of prosody.” Couper-Kuhlen and Selting have further alleged three sources for the neglect of prosodic phenomena: (1) They are not “segment-based, referential units”; (2) they are continuous rather than discrete units; and (3) they are not systematically codified in writing.

Related to this third source, there seems to be one additional reason for the neglect of research on prosody: conflicting and/or inefficient transcription notations. The *GAT* system (Selting et al., 1998), used by Couper-Kuhlen and Selting (1996b), reflects some of the problems of transcription notation systems. In examples of German utterances transcribed according to *GAT* in Kowal and O’Connell (2003a, p. 100), one finds emphasis notated as “!PIK!”; rising intonation as “gewesen?”; falling intonation as “nich.”; syllabic prolongation as “:.”; loudness as “<<f>wir>”; and quiet speech as “<< p>wir>”. None of these notations can reflect either the *suprasegmental* nature of the variables or their *continuous* rather than discrete nature (see Selting, 2001, p. 1065 ff., on problems of prosodic transcription).

Herrmann and Grabowski (1994, pp. 32 f.; our translation) have insisted that there is frequent agreement in the notation conventions of various research groups with regard to “the verbal, the nonverbal, and the utterance-accompanying components” of spoken discourse. However, Kowal and O’Connell (2003a, p. 102) have found only 30% of a total of all the notation

symbols across five German and three English transcription systems in complete agreement with one another. For example, *GAT* (Selting et al., 1998, p. 114) uses uppercase lettering preceded and followed by an exclamation mark as one option for notating emphasis (in the original German, “extra starker Akzent”), whereas all the other systems use a different codification of symbols to notate emphasis. Another example of the problems arising from the diversity of transcription notation systems can be seen in the very first chapter of Couper-Kuhlen and Selting (1996b). They have presented a multitude of cited examples transcribed in accord with various notation systems. These examples make it abundantly clear that there is no unified system for presenting prosodic data. In addition, the diversity makes it very difficult for a reader to understand the transcripts, not to mention the difficulties of reproducing already published transcripts in further publications, as we have discussed in Chapter 10. Quite in accord with these observations, Crystal (1997, p. 172) has referred to the notation systems for transcribing intonation as “competing descriptive frameworks” that vary greatly precisely because they reflect different theoretical views (e.g., phonetic vs. phonological; auditory vs. acoustic).

Research

Most of the prosodic research that has been undertaken in recent years has been concerned with intonation. Many years ago, Abercrombie (1965, p. 6; as cited in Couper-Kuhlen & Selting, 1996b, p. 12) set the stage for such investigation:

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.

While crediting conversation-analytic research with an interest in intonation, Couper-Kuhlen and Selting (1996b, p. 13) have contended that the interest of this research stops largely at the level of the transcript and rarely figures “in the analyses which conversation analysts have so far offered.” Gumperz (1982, p. 100; 1992) they have considered to be the exception, with his process of contextualization through the use of prosodic features.

Couper-Kuhlen and Selting (1996b) have found a number of problems with current intonation research. Criteria for the identification of intonation units have become controversial because they pit phonetic and phonological persuasions against one another. Couper-Kuhlen and Selting themselves have proposed to bypass the problem by going beyond traditional grammatical units and “taking a discourse perspective” (p. 16), in accordance with which

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-constructive unit, a ‘phonetic chunk’ which speakers use to constitute and articulate turns-at-talk.

Thus, they (p. 21) have linked intonation to interactional functions and goals and have referred to this as “*pragmatic* ‘meaning,’” “situated, inference-based interpretation” rather than “semantic meanings of decontextualized linguistic forms.” Once again, we seem to be dealing with a syntax for spontaneous spoken discourse that transcends and supplements traditional grammatical categories, and in the case of intonation analyses goes beyond a long historical tradition of such grammatical categories. For our own part, we would find Couper-Kuhlen and Selting’s “*pragmatic* ‘meaning’” quite compatible with our own understanding of the basic semantic meaning of an utterance from a psychological perspective.

It should be noted, however, that their “situated inference-based interpretation” may well go beyond the evidence. There may be some confusion between the immediate understanding of meaning on the part of interlocutors and the inferential processes of research analysts, in the sense that the researchers are indeed referring to their own (quite legitimate) research inferences. However, the interlocutors do not necessarily proceed by “uncoupling intonation from lexico-syntax” (p. 22); rarely does the interactional use of intonation to determine meaning require a throw-away of lexico-syntax. Quite the contrary, there is most commonly a co-determination on the part of intonation (and many other prosodic and contextual factors) and lexico-syntax. In fact, we would argue that Brazil, Coulthard, and Johns (1980, p. 18; cited in Couper-Kuhlen & Selting, 1996b, p. 22) have made an artificial dichotomy between “linguistic features of the message” and “the speaker’s assessment” in the following:

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.

After all, it is the speaker who deliberately chooses the linguistic features – along with tone – precisely to aid and abet in the communication of the message to the audience. Intonation is decidedly *not* “primarily a symptom of how we feel about what we say” (Bolinger, 1989, p. 1; as cited in Couper-Kuhlen & Selting, 1996b, p. 23), it is a constituent determinant of what we say, part of what we say. Nonetheless, the distinction between syntactic and prosodic units remains extremely important (see Kern & Selting, 2006, p. 244).

The reader will notice that mainstream psycholinguistics has not played a prominent part in the intonation research detailed above. By and large, the psychologists have not been ready to go beyond the lexico-syntax as Couper-Kuhlen and Selting (1996b, p. 21) have done with their “*pragmatic* ‘meaning.’” For example, the prosodic research of Grosjean and his colleagues (e.g., Grosjean, Grosjean, & Lane, 1979) was based on “isolated passages without

any communicative intent” (O’Connell, 1988, p. 162). More recent psycholinguistic research displays the same neglect. We have sampled a number of English- and German-language psycholinguistic texts with the following outcome: Clark (1996, p. 182) has limited himself to several paragraphs in which he seems to make “intonation or prosody” synonymous; Harley (2001, p. 106) has mentioned prosody only in the context of the language development of infants, but has not included intonation in his index; in Dietrich’s index (2002), neither intonation nor prosody is to be found; and Rickheit, Sichelschmidt, and Strohner (2002, p. 52) have included only one short paragraph regarding a generic definition of prosody. Hence, it is gratifying to find in Carroll’s (2007, p. 70 f.) recent textbook an extended treatment of both prosody and intonation.

Our Research on Articulation Rate

We wish to present here a set of our own research projects on one particular topic in the domain of prosody – articulation rate. Goldman-Eisler (1968, p. 25) had considered articulation rate to be “a personality constant of remarkable invariance” and had accordingly neglected to observe its variation across settings and genres. However, in a number of projects, spanning 1986 to 2004 – all with a cut-off point of 0.12 or 0.13 s for pauses – one of our clearest findings has been that there was no overlap whatsoever between mean articulation rates for rhetorical readings of poems (3.69, 4.20, and 4.72 syl/s; Sahar, Brenninkmeyer, & O’Connell, 1997, p. 453) and inaugural addresses (4.37 syl/s; Kowal et al., 1997, p. 14) on the one hand and TV interviewers and interviewees (means ranging from 5.04 to 6.14 syl/s; Kowal & O’Connell, 1997, p. 313; Kowal & O’Connell, 2004b, p. 91; O’Connell & Kowal, 1998, p. 549) on the other. Articulation rate seems to be far more complex than Goldman-Eisler’s personality constant makes room for.

Futuristics

At the beginning of the twenty-first century, Couper-Kuhlen (2001, p. 16) set out to review research on the use of intonation in discourse. Essentially, her approach has been to record as historical the competitions of the past, to note the divisions of researchers into several schools in the present, and to emphasize the importance for all of dealing with the complexity of prosody in the future:

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 the contextualization-cue approach there has been a subtle shift away from the study of “intonation” to the study of *prosody* and discourse.

Her mention (p. 25) of “a second type of new territory in the field of interactional prosody” is even more futuristic. She has emphasized the universality of

the temporal dimension of spoken discourse and has predicted that “the focus here will be on *timing*.” She has acknowledged the availability of objective measures of timing, but has striven to go beyond that level in search of “the metric which is behind participants’ subjective judgment of time.” Perhaps mainly as a consequence of our original training as experimental psychologists, our own research has always been couched in terms of objective – instrumentally measured – time. We have indeed confirmed in a series of studies (see our Chapter 11) that both experimental subjects’ and trained transcribers’ judgments of time are sometimes quite different from objectively measured time. Moreover, we are aware that the study of time as prosody is not about “participants’ subjective *judgment* of time,” but about participants’ *use* of time. It is the research analyst who *judges* time; participants *use* time.