

## SYMPOSIUM 1: Semantics, Syntax and Prosody

Chairman: *Ilse Lehiste, Columbus, U.S.A.*

Panel members: *E. Gårding, P. Martin, A. Cutler, V. Fromkin*

Discussant: *H. Fujisaki*

The symposium started with a brief introduction by the chairman, who observed that the relationship between prosody and syntax has been explored quite intensively in the past few years, but that the relationship between prosody and semantics has not been in the focus of research to a comparable degree. Contributors to the symposium have tried to address themselves to both questions. Since the topic - the relationship between semantics, syntax, and prosody - is very broad, some limitations have to be imposed on its treatment. We will not attempt to define precisely the relationship between semantics and syntax. Neither will we try to draw internal distinctions within the broad field of semantics. Prosody will be discussed at sentence level. Obviously, prosodic features are used in language also to convey both lexical and grammatical meaning at the word level: a tonal pattern, a quantity pattern, or difference in stress can distinguish between two lexical items or indicate, for example, a difference between two case forms of the same word. These functions of prosodic features will be taken for granted. Context more extensive than a single sentence will enter into the discussion by implication in connection with pragmatic considerations, i.e. when prosody is used to relate the sentence to a particular situation.

Sentence-level prosody comprises aspects of timing, accent, and intonation. Two of the four contributions (those by Gårding and Martin) address themselves primarily to intonation, and two (those by Cutler and Fromkin) focus on accent. Since timing was not discussed to any great extent by any member of the panel, the chairman started the symposium by saying a few words about the sentence-level function of the time dimension of spoken language.

In a series of papers summarized in Lehiste's contribution to the 13th International Congress of Linguists (held in Tokyo in the autumn of 1982) she has explored the relationship between timing and syntax, and has come to the conclusion that timing is the primary cue to the syntactic structure of a sentence. Boundaries of syntactic constituents can be signalled by pauses; they are also signalled by preboundary lengthening. At least in English pauses are not obligatory. It has been Lehiste's claim that lengthening does not achieve this effect (of signalling the presence of a boundary) by itself, but indirectly through its effect on rhythm: both preboundary lengthening and insertion of a pause disrupt the expected rhythmic structure of a sentence. In

a more recent paper (about to be published in *Folia Linguistica*) Lehiste has shown that the presence of fundamental frequency is not necessary for signalling syntactic boundaries - speakers and listeners are equally successful in disambiguating whispered sentences as they are in producing and perceiving normally phonated speech.

The existence of a link between syntax and prosody seems to be generally accepted by now. Two issues are currently being debated. The first is the question whether prosodic structure is entirely predictable on the basis of syntactic structure, or whether there exists an independent prosodic structure that may or may not correspond to the syntactic structure of a sentence. The second issue is the question whether the syntactic structure of a sentence is reflected primarily in timing or whether it is primarily reflected in some other prosodic feature, such as accent or intonation. Lehiste has come out both in favor of an independent prosodic structure, namely sentence-level rhythm described in terms of metric feet, and in favor of timing as the primary means for signalling syntactic structure - by controlled disruption of the rhythmic structure of the spoken utterance.

What does not seem to have been given any serious consideration is the relationship between speech timing and semantics. It appears that duration has only an indirect role here: if it can be shown that *accent* is directly connected to semantics, then duration enters in as one aspect of the phonetic manifestation of accent. If it can be shown that *intonation* is directly connected with semantics, duration could conceivably play a role in helping establish the domains over which particular intonational patterns are manifested: intonation units must have a duration. At the present time, there seems to be no immediate connection between semantics and sentence timing.

*Gårding* in her contribution developed a theory of intonation. Using examples from Swedish, French and Chinese, she showed how intonation can serve to express modality and syntactic and information structure.

An intonation curve has local maxima and minima, turning points. These turning points are parts of a larger pattern, the grid, which is obtained by connecting the main maxima by a topline and the main minima by a baseline. A grid may be rising, falling, or level. Such grids may in turn be part of a global pattern which may be rising, falling or level. A grid may have a normal, expanded or compressed width even to the extent of being best represented by one line only. The part of the grid where the direction or width are changed, or where the grid takes a jump, is called a pivot. An intonation unit is a piece of an utterance with an unbroken grid.

These concepts - the local turning points, pivots, direction and width of the grid - are associated with the syntax and semantics of the utterance in the following way. The local turning points signal words and morphemes, that is semantic and syntactic entities. The pivots signal semantic constituents, like the theme and the rheme, or syntactic constituents like the subject and the predicate. The general direction of the grid over the utterance, often in combination with the direction of the last intonation unit, determine the

speech act type, which may be declarative, interrogative and imperative, to use classical terms. In all cases, this is also a syntactic sentence type but not necessarily of the corresponding class. An interrogative speech act type may very well correspond to a declarative sentence type.

The width and position of the grid of an intonation unit signal its information weight relative to other intonation units. In other words, these intonation units may be in focus and out of focus. In this way the width and position may express coordination and subordination of semantic constituents and of syntactic ones.

Again, the classes are not isomorphic. We may very well focus a subordinate clause by prosodic means, and we may focus any clause or part of a clause by syntactic means, e.g. word order, without assistance from prosody.

*Gårding* summed up her presentation by asserting her belief that the grid, pivot and intonation unit are useful tools in intonation analysis. They are strongly connected with syntax and semantics in a similar way in different languages. Whenever there is a conflict between syntactic and semantic categories, it is the semantic function that has the greater impact on intonation.

*Martin* defined intonation as the sequence of prosodic contours located on the stressed vowels of an utterance. Intonation is perceived to be linked in certain ways to both the syntactic structure of a sentence and its semantic content. Using examples from French, *Martin* demonstrated that intonation can be constrained by syntax and semantics; nevertheless all three are considered to be independent systems, functioning according to their own rules. *Martin* developed a theory of intonation, according to which utterances have both a syntactic structure and a prosodic structure. Both structures are hierarchical, and they are not necessarily isomorphic, although some connection must exist between the two. Sentences are divided into *rheme* and one or more *themes*; prosodic division into rheme and theme is indicated by a specific prosodic contour located on the last stressed syllable of the rheme. The prosodic division into rheme and theme can be either prevented or forced by a semantic or syntactic constraint. According to *Martin*, pauses are direct manifestation of syntax and should not be considered as part of an independent prosodic system.

The papers by *Cutler* and *Fromkin* were primarily concerned with the relationship between sentence accent and syntax and semantics: whether the position of accent is determined by syntax, or whether the speaker can assign it more or less freely to express a particular communicative intent. Both papers used slips of the tongue as relevant research data; it was interesting to note that the same kinds of data served to support rather different conclusions.

*Cutler* finds that in producing accent patterns, language users behave as if sentence accent placement were concerned with the semantic and pragmatic structure of utterances rather than with their syntax. Speakers place accents to reflect the information structure of the message they wish to convey;

listeners use accentual structure to locate points of information focus. Cutler adduced evidence from children's acquisition of the production and comprehension of accent to strengthen her claim that accentual structure is associated with the meaning of a message rather than its form.

Fromkin argued that the semantic function of accents does not exclude a dependence on syntax, and that no new evidence exists to counter the claim that phrasal stress (which can coincide with accent) is determined by syntactic structure. According to Fromkin, normal, non-contrastive intonation, too, is determined syntactically. Primary stress or accent must be assigned after the syntax is determined; accent placement is independent of the particular lexical items on which it falls even if the semantic focus is thereby confused or changed. Nevertheless, Fromkin concluded that assignment of prosodic structure depends both on the syntactic structure and information structure of utterances.

Commenting on Gårding's paper, Fujisaki agreed with her concerning the need for a generative model in order to be able to interpret correctly a given  $F_0$  contour. This is so because an  $F_0$  contour is generated as a consequence of control of the vocal fold vibration by a set of commands that are directly related to the linguistic structure of an utterance. As Fujisaki has shown elsewhere, the underlying linguistic structure becomes more explicit if we *deconvolve* the mechanical/physiological characteristics of the laryngeal control mechanism from an observed  $F_0$  contour. In this sense, the use of a piecewise-linear tonal grid is a rather crude and unrealistic approximation of the observed characteristics of  $F_0$  contours. In Fujisaki's opinion, an objective and quantitative way to extract the underlying linguistic information is possible not by stylization but by deconvolution, using analytically expressible transfer functions (describing the quantitative relationships between the linguistic information as input and an  $F_0$  contour as output) with physically and physiologically meaningful formulations.

Fujisaki then presented his own model for the mechanism of generating an  $F_0$  contour from a set of linguistically meaningful commands. According to this model (based originally on an idea by Öhman, but elaborated by Fujisaki), an  $F_0$  contour – expressed in terms of  $\log F_0$  as a function of time – can be decomposed into two types of components: 1) phrase components corresponding to prosodic phrases, and 2) accent components corresponding to prosodic words. Each of these components can be considered as a consequence of control of the laryngeal mechanism by a separate linguistic command: 1) the phrase command, which is a sharp pulse (an impulsive force) applied to the laryngeal structure and which generates a phrase component, and 2) the accent command, which is a stepwise force applied to another part of the laryngeal structure, which generates an accent component. Responses of the laryngeal mechanism to these two kinds of commands are different, but each of them can be approximated very closely by the transfer function of a second-order linear system. By expressing these transfer functions in a quantitative way, one can decompose a given  $F_0$  contour

into its constituent components (i.e. phrase components and accent components) and deconvolve the inputs and the respective transfer characteristics.

Phrase components are the causes of declination and constitute what some people call the baseline – roughly corresponding to the lowest line of Gårding's tonal grid. Both a complete resetting (at a respiratory pause) and a mere superposition of a new phrase upon the old one are possible, and this corresponds roughly to Gårding's pivot. Fujisaki's research has shown that his model can approximate quite closely  $F_0$  contours of spoken words and sentences of Japanese, English, and Estonian. The analysis of Japanese sentences has shown that speakers of Japanese generally use three levels of phrase components and two levels of accent components; these numbers, however, may differ from language to language.

Commenting on Martin's paper, Fujisaki expressed his readiness to agree with Martin that the prosodic structure and the syntactic structure are two different systems. He also liked the concept of 'prosodic word', which is appropriate for describing the phenomenon of 'accent sandhi' in Japanese, where two or more lexical items, each of them with its individual word accent types, are often *concatenated* (not *compounded*) and behave as if they were one long word with its own word accent type.

In the case of Japanese, however, accent sandhi is not necessarily all or none. Fujisaki has encountered cases where the coupling between two words is somewhat intermediate and where it is difficult to decide whether they should be regarded as one prosodic word or two prosodic words. He suspects that this may also be the case in spoken French. He asked Martin to define a prosodic word, to describe its acoustic-phonetic manifestation, and to say to what an extent prosodic words are stable or variable.

Commenting on the papers of Cutler and Fromkin, Fujisaki expressed his concern about the discrepancy of conclusions drawn by Cutler and Fromkin from the same kind of speech error data. Although he has no doubts about the validity of the research technique – of collecting and interpreting speech error data, – it nevertheless seems that the patterns of speech errors are of such a great variety that it is possible to draw two entirely different conclusions. He would interpret the apparent discrepancy between Cutler's conclusions and Fromkin's conclusion in the following way: the pattern of errors in stress placement could sometimes depend mainly on the syntactic structure, but sometimes mainly on the semantic and pragmatic structure of the utterance.

Generalizing from all four papers, Fujisaki observed that there exist at least two types of languages which differ in the ways of expressing information concerning the focus. Judging from Gårding's paper, focal information in Swedish seems to be signalled mainly by pivots (or by the presence of a new phrase command in Fujisaki's model), while in English it seems to be signalled mainly by accent placement (or by an increase of the accent command on a particular word) according to Cutler and Fromkin. Fujisaki added that in the case of Japanese, phrase commands are mainly determined

by syntax and there is very little chance for focal information to be signalled by intonation.

While agreeing with the idea that the prosody of a sentence has its own structure, Fujisaki pointed out the differences in the constraints imposed on these structures. Syntactic and semantic structures of a sentence are constrained by the limited capacity of the short-term memory and the limited depth of, e.g., syntactic embedding; the prosodic structure; however, is more severely constrained by the limited characteristics of peripheral mechanisms, both in production and in perception. For instance, one cannot utter a very long sentence without pause, even though the syntax may allow it. One cannot have more than three or four different sizes of phrase commands to signal the presence of syntactic or semantic units of different size or different complexity, and so on. In other words, the prosodic structure is more constrained as compared with the syntactic structure or the semantic structure. Therefore one often finds a degenerate (incomplete) representation of the syntactic or the semantic structure. Whether one tries to preserve (a part of) the syntactic information or to preserve (a part of) the semantic information in the prosodic structure will depend on the individual, on the situation, and on the language.

Responding to Fujisaki's comments, Gårding explained the notion of grid in a more detailed way. She pointed out that her model and that of Fujisaki are similar; both are based on Öhman's model. Grid is a concept that is useful for describing focus over part of the utterance. Responding to a question by Klaus Kohler about the degree of abstractness of her model, Gårding stated that the model is quite concrete: it is used both for analysis and for the generation of a concrete pitch contour.

Martin answered Fujisaki's question about prosodic words by reference to the relative amplitude of melodic variation. There is one lexical stress per prosodic word; in French, the prosodic word comprises the accented syllable and preceding unaccented syllables.

Cutler and Fromkin responded to Fujisaki by emphasizing the validity of the technique involving collection and interpretation of speech errors.

The discussion now became more general. The chairman had requested that those contributors to the discussion who wanted their comments included in the report provide a written version before the end of the congress; not all speakers complied with this request, and their comments can therefore be included only in a general way.

Tore Janson commented about a statement made by Fujisaki, stressing that even if prosody is implemented at a low level in articulation, it reflects complicated syntactic and semantic/pragmatic facts. The basic problem in speech communication is that the speaker has to convey information to the listeners about complex hierarchical structures over a channel that permits only linear order. The role of prosody is largely to hint the nature of the hierarchical structure. This can be done only partially and imperfectly, due to the relatively poor expressive capacity of the prosodic signals. Therefore,

the listeners will have to do much guesswork when evaluating this information. For that reason, it seems certain that prosodic processing is going on even at the highest levels at the listeners' end.

C.W. Temu criticized the members of the panel for the narrow choice of examples from European languages. He asked also about differences in the way prosody works in languages in which there are changes possible in word order, versus those languages in which there is no change in word order. Fromkin responded to this comment with examples from Twi; Gårding's Chinese examples were likewise relevant in that context.

Mario Rossi, in his extensive comment, concentrated on the functions of intonation: expressive, demarcative and hierarchical. The units of expressive function are specific contours; the units of demarcative function are continuative and terminal intonemes; the units of hierarchical function may be units of demarcative function and/or semantic accent. The demarcative function is constrained by syntax. The hierarchical function is not always congruent with syntax, because it is mainly constrained by semantic organization of focus/presupposition and rheme/theme. To the extent that the theme/rheme structure corresponds to subject/predicate structure, intonation is congruent with syntax; when rheme indicates information weight (i.e. focus), intonation is used to scale the semantic content (as demonstrated many years ago by Mathesius and Karčevsky), and intonation is not necessarily congruent with syntax.

Several speakers suggested that if there is a conflict between semantics and prosody, semantics has the upper hand. Lehiste responded to that by arguing that prosody has to be independent of semantics at least to a certain degree, since it can be used to turn the meaning of a sentence into its opposite in the case of conscious expression of irony. It was suggested that prosody, syntax and semantics should be considered three 'channels of communication', all three of which are independent in a certain way.

It remains to assess the state of the art, as it is reflected in the four contributions to the symposium, the comments of the discussant, and the contributions made from the floor. It appears to be generally accepted that a relationship exists between prosody, syntax and semantics. The disagreements pertain to the relative independence of these three aspects of language. In their written contributions, Gårding had viewed intonation as basically independent of syntax, and Martin as independent of both syntax and semantics. Gårding, however, associated pivots (i.e. changes in the direction of fundamental frequency movements) with syntactic boundaries as well as with the information structure of the utterance. Cutler saw accents as determined primarily by the semantic import of an utterance, while Fromkin argued for a more important role for syntax in determining the location of the accent. Lehiste claimed that the rhythmic structure of a sentence is independent of its syntactic structure, but can be modified to reflect differences in the placement of syntactic boundaries. Fujisaki viewed prosody as a rather severely constrained way to represent syntactic and semantic struct-

ure. As a result of the discussion, panel members seemed willing to concede that alternative interpretations are possible; it appears rather remarkable that no strong claims for universality were made, and the opinion seemed to prevail that languages may differ in the ways in which they relate prosody, syntax and semantics.

## References

- Lehiste, Ilse (1982). The role of prosody in the internal structuring of a sentence. *Preprints of the Plenary Session Papers, The XIIIth International Congress of Linguists, Tokyo, 1982*. Pp. 189-198.
- Lehiste, Ilse (1983). Signalling of syntactic structure in whispered speech. To appear in: *Folia Linguistica*, Vol. XVII.