

Peak, boundary and cohesion characteristics of prosodic grouping

Daniel Hirst

Institut de Phonétique, CNRS - Parole et Langage, Université de Provence

29 avenue Schuman, 13621 AIX (France)

e-mail : Hirst @ fraix11.bitnet

INTRODUCTION

Prosodic grouping can be defined as the way in which phonological units are assumed to join together to form larger units. Such grouping has, at various times and by various authors, been accounted for by appealing to one or more of three distinct but inter-related characteristics which I refer to here as *peaks* (= prominence, headedness), *boundaries* (= disjuncture, discontinuity) and *cohesion*. (= juncture, continuity).

There is a certain redundancy in these three characteristics. Thus boundaries and cohesion, for example, are mutually exclusive, together defining prosodic groups (=units, constituents or domains). The relationship between domains and groups is more subtle and also more theory dependent. Halle & Vergnaud (1987) argue that the pure representation of heads (= peaks) and the pure representation of domains (= groups) constitute "conjugate representations" which together define the notion of government : "*the property of being a head is the same as that of being a governing element in a constituent, and the property of being in some domain is the same as that of being governed by some head.*" (p 16)

Despite this redundancy, I argue that these three characteristics are worth distinguishing when we look at the way prosodic constituents are defined. Before looking at the way they apply to spoken language, as an exercise it may be interesting to see how these concepts can be applied to written language where the notion of constituents seems more easily defined.

Grouping in written language.

Note that we are interested here in looking at the formal graphic constituents of written language, not the linguistic constituents. At the lowest level we may distinguish the *letter*. Letters combine to form *words*. Words combine to form *punctuation groups* (i.e. groups delimited by punctuation marks). Punctuation groups combine to form *paragraphs*. Above the level of the paragraph, constituents vary depending on the type of document. A book may be further organised into *chapters* and *volumes* : other documents may have a much more highly organised and specific structure.

Assuming that we are concerned with printed texts using the roman alphabet, letters, as constituents are defined by the characteristic of cohesion : a letter is a graphic symbol presenting a horizontal continuity (some letters such as i, j and accented letters are discontinuous but the discontinuity is always vertical). To take into account cursive handwriting (or other writing systems such as Arabic or Devanagari where individual letters are not necessarily disjoint) we should need to appeal to a notion of graphic peak to define the way in which we separate one letter from another. The graphic word can also be thought of as being defined by its cohesive properties : a word consisting of a sequence of contiguous letters. The punctuation-group is a clearcut case of a boundary-defined constituent marked. Paragraphs, like words, are units which can be defined in terms of cohesion. Chapters are defined by boundaries (chapter headings) and usually by cohesion (new page for new chapter). Volumes too are defined both by cohesion (binding) and boundaries (covers).

Besides these formal constituents, graphic texts can also be divided into lines and pages which are not constituents although as I have mentioned they are used as part of the cohesive properties defining paragraphs and chapters respectively. If we set up a hierarchy among punctuation boundary symbols defining minor symbols (, ; :) and major symbols (. ? !) then we can define a further hierarchical level : the sentence, intermediate between the punctuation group and the paragraph. Unlike the other graphic constituents we have defined, this distinction is not based on formal properties of the boundary

symbols themselves. Sentences are also marked by a boundary symbol in the form of an initial capital letter but initial capitals are not sufficient defining characteristics (this is even more true in German).

It is controversial whether reading appeals to intermediate constituents between the letter and the word. Against the classic idea that reading can make use of both direct (orthographic) and indirect (phonological) lexical access, recent research (Treiman & Chafetz 1987) suggests that intermediate units between the letter and the word may play a greater role in reading than has hitherto been thought. Interestingly, if we look at characterisations of the *sonority hierarchy* of phonemes such as the following (adapted from Goldsmith 1990, p112) :

$$\{ a > e; o > i; u > r > l > m; n > s > v; z; D; f; Q > b; d; g > p; t; k \}$$

it is a rather striking fact, which to my knowledge has not been pointed out before, that the shape of the letters partially encodes this hierarchy. All of the symbols on the lefthand (strong) side of the hierarchy (except for /l/) are small letters whereas all of those on the right are ascending or descending letters. It seems difficult to imagine that such a fact is a mere coincidence. Small letters seem, then, to act as potential syllable peaks, which explains both why lower-case letters are generally considered more legible than upper case (Tinker 1963), and how the shape of a word can function as a cue to its lexical identity (Walker 1987).

We can conclude from this discussion that even in the case of printed language where a document is entirely composed of a sequence of discrete symbols, there are a number of formal constituents (line, page) whose role is only marginal while other constituents (syllable, sentence) are only indirectly related to or derivable from the printed form.

GROUPING IN SPEECH

As is to be expected, the relation between the observable data and the phonological units which are assumed to structure this data is even more indirect in the case of speech than in the case of written language. In speech, unlike in writing, there are no unequivocal units on which an analysis can be based. The only direct formal constituents of a speech-signal are portions delimited by silences and/or by breathing. Both of these have in the past been taken as boundary signals for phonological constituents, ie 'breath-groups' (Sweet 1890; Jones 1949; Lieberman 1980) or 'pause-defined units' (Brown, Currie & Kenworthy 1980). Lieberman's categorical affirmation : "*the breath-group is the primary element that people use to segment the flow of speech into sentence-like units*" (p 240) is worth contrasting with Sweet's far more cautious : "*These breath-groups correspond partially to the logical division into sentences*". The parallel with line ends and page breaks in reading is instructive since these, as we saw above, just like breaths and silences, are neither necessary nor sufficient criteria for isolating linguistic units even though higher order linguistic units will often make use of such direct physical characteristics to reinforce the internal cohesion of the units.

Since the observation of directly delimited prosodic units does not prove very rewarding, the next step is to look at those indirect units which have been proposed in various analyses. Between the phoneme and the utterance, the following hierarchy is a fairly traditional one in prosodic analysis :

syllable < accent group < intonation unit

None of these prosodic units are uncontroversial. I propose to review each in turn together with alternative proposals that have been made for units of approximately the same level. There have been a number of proposals concerning prosodic units above the intonation unit - paratones - the prosodic equivalent of paragraphs. Research in this area, in particular into the prosodic structure of discourse, is a world unto itself and I shall not attempt to deal with it at all in this paper.

AROUND THE SYLLABLE

It could be claimed that the syllable is the prosodic unit for which there is the greatest consensus among linguists and phoneticians. Speakers' intuitions concerning the number of syllables for example are far more consistent than those concerning phonemes. Thus de Cornulier (1982) has shown that French speakers can easily spot an odd-man-out in

sequences of words consisting of three syllables (*quantité, apéro, Nicolas,...*) whereas they are totally incapable of performing the same sort of operation on words made up of three phonemes (*nuit, avis, tard, train...*) (p 59.). The syllable, it has been claimed is a good contender for the role of basic unit of speech perception (Mehler et al. 1981). The demonstration of a syllabic effect in French (i.e. shorter reaction time for a detection task when targets were syllables than when they were smaller or larger than syllables) was not however replicated for English (Cutler, Mehler, Norris & Segui 1986). This was taken to imply the existence of a language-specific comprehension strategy and seemed to support the claim that certain consonants in English are ambisyllabic, that is belong to both adjacent syllables. (Anderson & Jones 1974, Kahn 1976). Other phonologists (Kiparsky 1979, Selkirk 1982) have claimed that ambisyllabicity is not necessary within a framework making use of the notion of 'foot'.

Most of the arguments in favour of the syllable as a unit are really arguments in favour of syllabicity rather than syllabic constituency : syllables, in other words (in languages like English) are identified by their peak characteristics rather than by their boundaries or their internal coherence. Recent work on temporal patterning has brought up an alternative candidate for modelling segmental durations. This is the V-to-V unit used by Fant & Kruckenberg 1988, Fant, Kruckenberg & Nord 1991 consisting of a sequence beginning with the vowel onset and continuing up to but not including the following vowel onset. There is considerable evidence that the vowel onset is a particularly crucial part of the speech signal (Dogil 1986). Barbosa & Bailly 1992 suggest that normalised duration of V-to-V units gives a better image of prosodic boundaries than does syllable duration. My own analysis of a reading of a one-minute continuous text in French reveals that V-to-V units have smaller variance than the corresponding syllables in the same text. This is to be put into relationship with the finding of Wightman et al (1992) that final lengthening only applies to the segments in the final rime, not the final syllable. The picture which emerges from these findings is that the syllable is not perhaps so strongly entrenched as a phonological unit as has often been believed. Instead of syllables, it could turn out to be their immediate constituents (onsets and rimes) which are the basic building blocks for speech.

AROUND THE ACCENT GROUP

Above the syllable, the accent group, or foot, is probably the most respectable phonological unit in common use. This unit, like the syllable, is obviously essentially a peak phenomenon. It is not evident that foot boundaries are themselves in any way well defined. The term 'accent group' or 'foot' is often used indiscriminately to refer to units which should perhaps be more carefully distinguished. Thus the foot can be taken to be a strictly phonological unit (Selkirk 1978) a purely rhythmic unit (Abercrombie 1964), a pitch accent or tonal unit, or even a mixed syntactic/phonological unit. The relationship of the accent group to the word has always been controversial. Thus Pike (1945) and Jassem (1951) both distinguished rhythmic units from tonal units. For both authors rhythm units were essentially groups of words. Unlike Pike's 'contour', however Jassem's 'Tonal Unit' does not take into account word boundaries but groups unstressed syllables with the preceding stressed syllable, irrespective of word boundaries. This is precisely the same unit which Abercrombie (1964), followed by Halliday (1968) later promoted as the Foot. Most of the evidence on final lengthening seems to suggest that any boundary effects are associated with the word (Beckman & Edwards 1990) or with higher order constituents (Wightman et al. 1992) rather than with the foot itself. The controversy as to the relevance of the Foot to prosodic analysis is, however, far from over as can be seen from the discussion between Nooteboom 1991, Fant 1991 and Kohler 1991 for example. It also remains to be seen whether the Foot can be generalised to account for language-specific characteristics. Hirst & Di Cristo (in press) suggest that Germanic languages in general are characterised by left-headed accent groups (initial prominence) whereas similar structures in Romance languages are right-headed (final prominence) (cf the analysis of French in Wenk & Wioland 1982).

A second question relates to the position of the boundary of the foot. Although most authors have defined the foot as a group of syllables, others, following Lea (1980), have preferred to work with the 'interstress interval' defined as the sequence running from the

onset of one stressed vowel to the next (Fant Kruckenberg & Nord 1991, Wightman et al. 1992). Systematic comparisons of the appropriateness of different prosodic constituents in different languages remain to be made.

AROUND THE INTONATION UNIT

Feet are generally held to be grouped into higher order constituents for which there are a great number of different proposals in the literature. Most work within the British tradition of intonation analysis has assumed, with a few exceptions, a single hierarchical level variously referred to as the "sense-group", "tone-group" or "intonation unit" (see Hirst & Di Cristo in press for a review). Recent work on the TOBI transcription system (Silverman et al 1992, Hirschberg & Beckman 1993) has followed Beckman & Pierrehumbert (1986) in assuming that at least in English and Japanese there are two levels of intonation units : called 'intermediate phrase' and 'intonational phrase' respectively. Although it is claimed that these units are distinguished by tonal boundary signals ('phrase accent' and 'boundary tone' respectively, this distinction is as yet extremely theory internal, since there is no simple one-to-one correspondance between acoustic data and such boundaries. A valley in the F0 curve for example may be interpreted as a L% boundary tone, a L phrase accent, the -L or *L of a pitch accent or simply an unspecified sagging transition between two high pitch accents.

Recent corpus-based studies of durational cues for prosodic constituency have suggested that more than two levels of intonation units can be systematically distinguished. Thus both Campbell & Ladd (1991) and Wightman & al. (1992) found that at least four degrees could be established on the basis of final lengthening. This has been interpreted by Ladd & Campbell as evidence in support of the claim (Ladd 1986) that there is no principled limit to the depth of prosodic structure. A similar claim has been made by Martin (1975) and is implicit in the 'performance structures' of Gee & Grosjean (1983), Monnin & Grosjean (in press). It still remains to be shown whether there are categorical differences between different types of intonation units and how far such distinctions hold across different languages.

One of the most promising lines of research into cues for prosodic grouping is that of tonal cohesion. Downstepping or downdrift sequences have long been known to provide important cohesion cues in many languages. Thus in Bambara (Mali) a downdrifted sequence : /sísé ná ^hkúma/ will be interpreted as a single clause : "Sissé is going to talk!" whereas the same sequence without downdrift /sísé ná kúma/ will be interpreted as two clauses : "Sissé! Come and talk!" (Hirst 1979). An upward resetting of pitch has a distinct boundary effect even in the absence of a nuclear pitch movement in the preceding intonation unit (Kingdon 1958 p62). Research into prosodic parsing on the basis of predictive models of tonal cohesion only just begun (Schuetze-Coburn, Shapley & Weber 1991: Bruce, Granström, Gustafson & House in press) but promises to be one of the most rewarding fields in prosody in years to come.

The investigation of the relationship between Intonation Units and syntactic, semantic and pragmatic representations has a long and varied history. Prosodic structure theory (Selkirk 1978, Nespor & Vogel 1986) presents a rich articulated theory of the relation between syntax and phonology. This is further developed in Selkirk (1986, 1990) where it is claimed that each lower-order constituent in a given language (phonological word, phonological phrase) is characterised by two parameters : a designated syntactic category (maximal projection/ lexical word) and a designated edge (left/right). Given these two parameters the mapping from syntax to phonology is defined by ensuring that the designated edge of the designated syntactic constituents aligns with the corresponding edges of the prosodic constituents. Selkirk has claimed several times that Intonational phrases seem to be governed by semantic rather than by syntactic constraints. I have argued (Hirst 1993) that a non-deterministic version of Selkirk's rule will in fact account for the data which she found troublesome for a syntactic account. One way of formulating this Mapping Rule is as follows :

- Map a syntactic structure exhaustively onto a linear sequence of intonational phrases such that :*
- a.) *the left end of each intonational phrase corresponds to the left end of a major syntactic constituent*

b.) *the intonational phrase is no longer than the corresponding syntactic constituent.*

The Mapping Rule also gives an interesting explanation for the fact that certain Intonation Units boundaries are optional while others appear to be compulsory. I claim that the difference stems from the non-linear nature of certain syntactic structures (parentheticals, vocatives, non-restrictive relatives etc) and that linear order is only imposed after the (optional) choice of Intonation Unit boundaries.

CONCLUSION

There is clearly a need for considerably more empirical data concerning the appropriateness of prosodic constituents at different levels for modelling durational and tonal characteristics of utterances. In particular very little is known concerning the way in which such characteristics vary from one language to another and this is an obvious priority task for the years to come.

ACKNOWLEDGEMENTS

This research has been partially supported by the French CNRS Cogniscience programme : *Les unités de traitement dans la perception de la parole et la lecture.* (Responsable : J. Pynte).

REFERENCES

- Abercrombie, D. (1964) "Syllable quantity and enclitics." in Abercrombie, Fry, MacCarthy, Scott & Trim (eds) *In Honour of Daniel Jones* (Longman), 216-222.
- Anderson, J. & Jones, C. 1974, 'Three theses concerning phonological representation.' *Journal of Linguistics* 10, 1-26.
- Barbosa, P. & Bailly, G. (1992) "Prédiction de la durée segmentale : le paradigme des groupes inter-P-Centers". in *Actes du Séminaire Prosodie* (La Baume-lès-Aix, October 1992), 73-79.
- Beckman, M. & Edwards, J. (1990) "Lengthenings and shortenings and the nature of prosodic constituency." in Kingston & Beckman (eds) *Papers in Laboratory Phonology*
- Beckman, M. & Pierrehumbert, J. (1986) "Intonational structure in English and Japanese" *Phonology Yearbook* 3, 255-309
- Brown, G.; Currie, K.L. & Kenworthy, J. (1980) *Questions of Intonation.* (Croom Helm; London).
- Cornulier, Benoît de (1982) *Théorie du vers : Rimbaud, Verlaine, Mallarmé.* (Seuil, Paris).
- Cutler, A; Mehler, J; Norris, J; Segui, J; (1986) "The syllable's differing role in the segmentation of French and English." *Journal of Memory and Language* 25, 385-400
- Dogil, Grzegorz (1986) *The pivot theory of speech parsing.* (University of Vienna Press, Vienna)
- Fant, G. (1991) "Units of temporal organization. Stress groups versus syllables and words." *Proc. I.C.Ph.S. XII* 1, 247-250.
- Fant, G.; Kruckenberg, A. & Nord, L. (1991) "Durational correlates of stress in Swedish, French and English." *Journal of Phonetics* 19, 351-363.
- Gee, J.-P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15, 411-458.
- Goldsmith, J. (1990) *Autosegmental and Metrical Phonology* (Blackwell; Oxford)
- Halle, M. & J.-R. Vergnaud *An Essay on Stress* (MIT Press; Cambridge, Mass.)
- Halliday, M.A.K. (1968) *Intonation and Grammar in British English.* (Mouton; the Hague)
- Hirschberg, J. & Beckman, M. (1992) "Report on proposed transcription system and some recommendations." unpublished ms.
- Hirst, D.J. (1979) 'Pitch features for tone and intonation.' *Travaux de l'Institut de Phonétique d'Aix* 6, 177-191.
- Hirst, D.J. (1993) "Detaching intonational phrases from syntactic structure." *Linguistic Inquiry* 24 (4).

- Hirst, D.J. & Di Cristo, A. (in press) "A survey of intonation systems" in Hirst & Di Cristo (eds) (in press).
- Hirst, D.J. & Di Cristo, A. (eds) (in press) *Intonation Systems : a Survey of Twenty Languages* (Cambridge University Press; Cambridge)
- Jassem, W. (1952) *Intonation of colloquial English*. (Panstowe Wydawnictwo Naukowe; Warszawa).
- Jones, D. (1949) *An Outline of English Phonetics* (Heffer; Oxford)
- Kahn, D. (1976) *Syllable based generalisations in phonology*. PhD Thesis, MIT.
- Kingdon, R. (1958) *The Groundwork of English Intonation*. (Longmans, London)
- Kiparsky, P. (1979) "Metrical structure assignment is cyclic." *Linguistic Inquiry* 10 (3), 521-541
- Kohler, K.J. (1991) "Isochrony, units of rhythmic organisation and speech rate." *Proc. I.C.Ph.S. XII* 1, 257-261.
- Ladd, D.R. (1986) "Intonational phrasing : the case for recursive prosodic structure." *Phonology Yearbook* 3, 311-340.
- Ladd, D.R. & Campbell, W.N. (1991) "Theories of prosodic structure : evidence from syllable duration." *Proceedings I.C.Ph.S. XII* 2, 290-293.
- Lieberman, P. (1980) "The acquisition of intonation by infants : physiology and normal control." in C. Johns-Lewis (ed) *Intonation in Discourse*. (Croom Helm; London), 239-257.
- Martin, P. (1975) "Eléments pour une théorie de l'intonation." *Rapport d'Activités de l'Institut de Phonétique de Bruxelles* 9(1), 97-126
- Mehler, J; Segui, J; Frauenfelder, U. (1981) "The role of the syllable in language acquisition and perception." in Myers, Laver & Anderson (eds) *The Cognitive Representation of Speech*. (Amsterdam, North Holland).
- Monnin, P., & Grosjean, F. (in press). "Les structures de performance en français: caractérisation et prédiction." *L'Année Psychologique*.
- Nespor, M. & Vogel, I. (1986) *Prosodic Phonology* (Foris, Dordrecht)
- Nooteboom, S.G. (1991) "Some observations on the temporal organisation and rhythm of speech." *Proc. I.C.Ph.S. XII* 1, 228-237.
- Pierrehumbert, J. & Beckman, M. (1986) "Intonational structure in Japanese and English." *Phonology Yearbook* 3, 255-310.
- Schuetze-Coburn, S.; Shapley, M. & Weber, E.G. (1991) "Units of intonation in discourse: a comparison of acoustic and auditory analyses." *Language and Speech* 34(3), 207-234.
- Selkirk, E.O. (1978) "On prosodic structure and its relation to syntactic structure." *Indiana University Linguistics Club*. (in *Nordic Prosody II* 1981)
- Selkirk, E.O. (1982) "The Syllable." in Van der Hulst & Smith (eds) *The Structure of Phonological Representations*. vol II (Foris, Dordrecht),
- Selkirk, E.O. (1986) "On derived domains in sentence prosody." *Phonology Yearbook* 3, 371-405
- Selkirk, E.O. (1990) "On the nature of prosodic constituency : comments on Beckman and Edward's paper." in Kingston & Beckman (eds) *Papers in Laboratory Phonology*, 179-200
- Silverman K. et al., (1992) "TOBI: A Standard for Labeling English Prosody," *Proc. of the Intern. Conf. on Spoken Language Processing*, 867-870.
- Sweet, H. (1890) *A Primer of Phonetics* (Clarendon; Oxford)
- Tinker, M.A. (1963) *The Legibility of Print* (Iowa State University Press)
- Treiman, R. & Chafetz, J. 1987 "Are there onset- and rime-like units in printed words." in M. Colheart (ed.) *Attention and Performance XII. The Psychology of Reading*. (Erlbaum; Hillsdale, NJ).
- Walker, P. (1987) "Word shape as a cue to the identity of a word : an analysis of the Kucera and Francis (1967) word list." *Quart.Jour. of Exper. Psy.* 39A, 675-700.
- Wenk, B. & Wioland, F. (1982) "Is French really syllable-timed?" *Journal of Phonetics* 10, 193-216.
- Wightman, C.; Shattuck-Hufnagel, S.; Ostendorf, M. & Price, P. (1992), "Segmental Durations in the Vicinity of Prosodic Phrase Boundaries." *J.A.S.A.*, 1707-1717