

MULTTEXT - LRE Project 62-050

Prosody Encoding Survey

WP 1 Specifications and Standards

T1.5 Markup Specifications

Deliverable 1.5.3.

Final version

15 September 1994

Joaquim Llisterri,

Universitat Autònoma de Barcelona

Contributors: Daniel J Hirst,

Laboratoire Parole et Langage, CNRS

1.- Introduction

The process of prosodic encoding can be defined as the symbolization of the linguistically relevant variations that occur in the domains of time, frequency and intensity in the sound wave corresponding to a speaker's utterance. The process of encoding implies deciding which variations in the physical parameters of the speech wave carry out linguistic information and finding a way to describe them by means of a symbolic system. Since physical parameters such as frequency and intensity are continuously varying over time, a symbolic coding implies also converting continuous information to a set of discrete units. Thus, symbolic coding of prosody involves at least two different levels of abstraction: a linguistic interpretation of changes in physical properties of the speech wave, and a classification of these changes into discrete categories. Finally, a notational system has to be designed in order to represent these categories; the discussion and evaluation of the systems that have been developed is the aim of the present report.

Traditional prosodic categories are length, stress and intonation (Grønnum Thorsen, 1987); pauses can also be added to prosodic phenomena, Variations in length or duration occur over the time domain, and they carry out contrastive segmental phonological distinctions - e.g. in languages with short vs. long vowels or consonants - together with information about stress in words and in sentences. Information about stress is carried out not only by changes in duration, but also in frequency and intensity that contribute to the perceived relative prominence of one syllable over others. Intonation can be understood as a variation in fundamental frequency (F_0) articulatorily related to the rate of vocal folds vibration and perceived as changes in pitch or tonal level. The situation is then rather complex, because a given prosodic category can have several acoustic or perceptual correlates. The opposite is also true, since variations in one acoustic parameter can be interpreted as a perceptual cue to different prosodic phenomena; variations in fundamental frequency, for example, are correlated with stress, with intonation, or can be used to signal boundaries between prosodic units.

The definition of prosodic categories is not the aim of our work, and for each coding or annotation system described we will use the categories defined by the authors of the system, providing an explanation if necessary. The readers interested in a general basic description of prosody are referred to chapter 10 of Ladefoged (1975) or chapter 5 of Pickett (1980); a more in depth discussion can be found in Cutler and Ladd (Eds.) (1983), Lehiste (1970), Cruttenden (1986), Rossi et al. (1981) or 't Hart, Collier and Terken (1990); for a wide perspective of current work see House and Touati (Eds.) (1993). A glossary with definitions of the technical terms can be found in part 4 of the report.

Prosodic coding or annotation can be performed manually by a human transcriber or in an (semi)automatic way. The operation can be also carried out for a variety of purposes and with different levels of detail which depend on the aim and applications. This report will review some of the traditional fields in which prosodic annotation has been used - discourse analysis, conversation analysis and text linguistics - but will

concentrate on systems developed for the annotation of speech databases aimed at speech technology applications or linguistic descriptions.

Moreover, prosodic coding can be included within the segmental transcription of an utterance or can be represented at a different level. As we will see, the first approach is usually taken in discourse and conversation analysis and in text linguistic studies, in order to provide a representation of a spoken text. The second is more traditionally linked to speech databases having speech technology applications in mind.

MULTEXT will provide an automatic prosodic coding derived from the speech wave. The orthographic text, a phonemic transcription and the prosodic coding will be represented in different tiers or levels, that will be temporally synchronized - or aligned - with the corresponding speech signal (Hirst, 1994). The materials that will be used in the project are 40 short passages of an average length of 5 thematically connected sentences, taken from EUROM.1, a multilingual speech corpus collected within the ESPRIT project 2589 SAM (Multilingual Speech Input/Output Assessment, Methodology and Standardization) (SAM, 1992).

The present deliverable has two main objectives and, therefore, is divided in two different parts:

- (1) a survey of different types of prosodic labeling systems
- (b) a comparison and evaluation of those systems in terms of different needs for speech and for natural language processing

2.- Prosodic labeling survey

The first part of the report aims at presenting some of the systems currently used in prosodic annotation, with special attention to those that have been developed having labelling of databases in mind. There are presently a number of systems developed within different traditions in separate fields and with different aims; it would be beyond the scope of this report to describe all of them in detail. For more details the reader is referred to Grønnum Thorsen (1987) who provides a review of the elements that should be coded in a prosodic transcription and of some of the classical notational conventions; Léon and Martin (1970) also contains a chapter devoted to classical approaches to prosodic transcription; Wells *et al.* (1992) summarizes the prosodic annotated systems discussed within the SAM Prosody Group, that will be presented here in more detail.

It is worth mentioning that during the ESCA Prosody Workshop (Lund, Sweden, September 1993) a task force composed by Gösta Bruce, Nick Campbell, Dafydd Gibbon, Daniel Hirst and Jacques Terken was set up to produce a survey and to prepare a report on Machine-Readable Labelling for Prosodic Notation. Contributions to this report from groups involved in prosodic labeling are expected by the end of September, and a draft version of the report will be circulated at the end of October '94. Unfortunately, it is not possible for timing reasons to include the results of this work in the present report.

References: general surveys

GIBBON, D. (1989) Survey of Prosodic Labelling for EC Languages. SAM-UBI-1/90, 12 February 1989; Report e.6, in ESPRIT 2589 (SAM) *Interim Report, Year 1*. Ref. SAM-UCL G002. University College London, February 1990.

GRØNNUM THORSEN, N. (1987) "Suprasegmental transcription", *ARIPUC, Annual Report of the Institute of Phonetics, University of Copenhagen* 21: 1-28

2.1.- Prosodic coding in discourse and conversation analysis

Discourse transcription is usually understood as "the process of creating a written representation of a speech event so as to make it accessible to discourse research" (Du Bois *et al.*, 1993:45). A great diversity of proposals exist in this field, and it is not practical to review each of them; interested readers may find examples of notational conventions in Atkinson and Heritage (1984), Du Bois (1991) or in Gumperz and Berenz (1993). All those systems share the fact that the transcription is based in conventional spelling, enriched with some conventions to represent information that is present in the spoken discourse but can not be conveyed by means of normal spelling conventions. The transcription of prosodic aspects requires the design of a set of symbols to represent intonation unit boundaries, terminal pitch direction, accent, and accent unit boundaries pitch movements and pauses.

As an illustration of this tradition, we reproduce the sub-set of symbols used in prosodic transcription and the definitions provided by the authors in the system proposed by Du Bois *et al.*, (1993)

Units

{ carriage return } Intonation unit : a stretch of speech uttered under a single coherent intonation contour

-- Truncated intonation unit

Transitional continuity

. Final: a class of intonation contours whose transitional continuity is regularly understood as final in a given language

, Continuing: a class of intonation contours whose transitional continuity is regularly understood as continuing in a given language

? Appeal: a class of intonation contours whose transitional continuity is

	regularly understood as an appeal, in a given language
Terminal pitch direction	
\	Fall
/	Rise
–	Level
Accent and lengthening	
^	Primary accent
˘	Secondary accent
!	High booster: a higher than expected pitch in a word
;	Low booster: a lower than expected pitch in a word
=	Lengthening
Tone	
\	Fall
/	Rise
\/	Fall-rise
/\	Rise-fall
–	Level
Pause	
...(N)	Long pause: .7 seconds or longer
...	Medium pause: between 0.3 and 0.6 seconds inclusive
..	Short pause: about 0.2 seconds or less
(0)	Latching: lack of pause between speakers' turns
Specialized notations	
(N)	Duration

&	Intonational unit continued
	Accent unit boundary
< >	Embedded intonation unit

Table 1: Prosodic transcription in Du Bois et al. (1993)

References: prosodic coding in discourse and conversation analysis

ATKINSON, J.M. - HERITAGE, J. (Eds.) (1984) *Structures of social action. Studies in conversation analysis*. Cambridge / Paris: Cambridge University Press / Editions de la Maison des Sciences de l'Homme

DU BOIS, J.W. (1991) " Transcription design principles for spoken discourse research", *Pragmatics* 1: 71-106

DU BOIS, J.W.- SCHUETZE-COBURN, S.- CUMMING, S.- PAOLINO, D. (1993) "Outline of discourse transcription" in EDWARDS, J.A.- LAMPERT, M.D. (Eds.) *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NY: Lawrence Erlbaum. pp. 45-89.

GUMPERZ, J.J.- BERENZ, N. (1993) " Transcribing Conversational Exchanges", in EDWARDS, J.A.- LAMPERT, M.D. (Eds.) *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, N.J.: Lawrence Erlbaum Associates. pp. 91-122

2.2.- Prosodic coding in corpus linguistics

As far as we know, it is not possible to document a tradition of prosodic coding in corpus linguistics as we have identified in the field of discourse analysis. Leech (1991) reports that notable exceptions to the lack of prosodic coding in spoken corpora are the London-Lund Corpus (LLC) - described in Svartvick et al. (1982) and in Svartvick (Ed.) (1990) - and the Lancaster/IBM Spoken English Corpus (SEC) - described in Taylor and Knowles (1998) and in Knowles and Lawrence (1987) - It is worth noting that both of them contain only English data.

An example of the kind of work carried out in prosodic coding in corpus linguistics is found in the papers published by Knowles (1991) and by Wichmann (1991) using the SEC. The corpus contains, according to Knowles (1991), an outline prosodic transcription. Transcription was made in three steps: division of the text into tone groups, identification of accented and stressed syllables in the tone groups and assignation of a pitch contour to the accented syllables.

Knowles' study deals with the marking of tone group boundaries in the process of transcription. Tone group boundaries are defined as "a discontinuity in the prosodic pattern" (Knowles, 1991:151); discontinuities are classified into three categories: temporal discontinuities - pauses greater than 25 cs. - , pitch discontinuities, and segmental discontinuities - evidenced by the absence of connected speech processes or by the presence of segmental patterns showing the existence of a separation between segments -. The paper concludes that in order to study prosodic breaks, marking of segmental discontinuities, pitch discontinuities and pauses is essential.

Wichmann (1991) deals with upward pitch movements in the same corpus. As far as transcription is concerned, 5 pitch levels are distinguished:

1	much higher
2	higher
3	the same
4	lower
5	much lower

Table 2: Notational system for pitch levels in Wichmann (1991)

Accented syllables are assigned superscript tonetic stress marks indicating the direction of pitch movement in the syllable: high level, rise, fall and fall-rise

An up arrow is assigned to any syllable when a step up in pitch is perceived as greater than normal. When up arrows occur together with tonetic stress marks it means that "the syllable is prominent for pitch, in addition to loudness and duration, and the step up in pitch is greater than that which is implied by the superscript tonetic stress mark alone" (Wichmann, 1991:167); when up arrows do not co-occur with tonetic marks "they indicate a syllable which is prominent only in terms of pitch and not for reasons of vowel quality, duration or loudness" (Wichmann, 1991:167). Wichmann's study links the occurrence of up arrows to boundaries between sections of texts - equivalent to paragraphs -, parenthesis, direct speech, rhetorical questions, lexical emphasis or pre-closing signals.

References: prosodic coding in corpus linguistics

KNOWLES, G. (1991) "Prosodic labelling: the problem of tone group boundaries", in JOHANSSON, S.- STENSTRÖM, A. (Eds.) *English Computer Corpora. Selected Papers and Research Guide*. Berlin: Mouton de Gruyter. pp. 149-163

KNOWLES, G.- LAWRENCE, L. (1987) "Automatic intonation assignment" in GARSIDE, R.- LEECH, G.- SAMPSON, G. (Eds.) *The Computational Analysis of English: A Corpus-based Approach*. London: Longman. pp. 139-148

LEECH, G. (1991) "The State of the Art in Corpus Linguistics" in AIJMER, K.- ALTENBERG, B. (Eds.) *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. London: Longman. pp. 8-29.

SVARTVIK, J. (Ed) (1990) *The London-Lund Corpus of Spoken English: Description and Research*. Lund: Lund University Press.

SVARTVIK, J.- EEG-OLOFSSON, M.- FORSHEDEN, O.- ORENSTRÖM, B.- THAVENIUS, C. (1982) *Survey of Spoken English. Report on Research 1975-1981*. Lund: Lund University Press.

TAYLOR, L.- KNOWLES, G. (1988) *Manual of Information to Accompany the SEC Corpus*. UCREL, University of Lancaster

WICHMANN, A. (1991) "A study of up-arrows in the Lancaster/IBM Spoken English Corpus", in JOHANSSON, S.- STENSTRÖM, A. (Eds.) *English Computer Corpora. Selected Papers and Research Guide*. Berlin: Mouton de Gruyter. pp. 165-178

2.3.- TEI (Text Encoding Initiative)

Chapter 11 of the Text Encoding Initiative Guidelines (Sperberg-McQueen and Burnard, (Eds.) 1994) discusses the transcription of spoken language. Since the main aim of this standardization effort concerns written texts, the guidelines presented in this chapter are oriented towards the transcription of speech as a text enriched with a set of conventions for phenomena that can not be adequately described with standard spelling. TEI Guidelines on spoken texts were mainly the result of work carried out within a subgroup composed of Stig Johansson - chair- , Jane Edwards and Andrew Rosta.

The following prosodic phenomena are discussed:

(1) Pauses

Pauses are signaled by the element <pause>, "to indicate a perceived pause either between or within utterances" (Sperberg-McQueen and Burnard, (Eds.) 1994 § 11.2.2). Two attributes are accepted:

- who, identifying the speaker

- type, used to specify the category of the pause - e.g. long, short - if needed. Duration is indicated as dur = . The example given in the Guidelines is the following (<u> and </u> indicate the beginning and end of an utterance):

" <u> Okay <pause dur=200>U-m<pause dur=75>the s the scene opens up <pause dur=50> with <pause dur=20> um <pause dur=145> you see a tree okay? </u> " (Sperberg-McQueen and Burnard, (Eds.) 1994 § 11.2.2)

Mechanisms for synchronizing pauses with other transcribed phenomena are also provided in the Guidelines.

(2) Tone units or intonational phrases

Since it is possible with TEI mechanisms to divide an utterance into smaller units or segments delimited by the <seg> element, it is possible to use this feature to signal prosodic units. An example is provided:

" <u who=F1> <seg> although its an old ide&stress;a</seg> <seg>it hasnt been on the mar&stress;ket very long</seg> " (Sperberg-McQueen and Burnard, (Eds.) 1994 § 11.3.1)

The beginning of the segment or unit is marked as <seg> and the end as </seg>; &tress appears immediately following the syllable bearing the primary accent.

(3) Shifts

The element <shift> "marks the point at which some paralinguistic feature of a series of utterances by any one speaker changes" (Sperberg-McQueen and Burnard, (Eds.) 1994 § 11.2.6). A series of features - tempo, pitch range, tension, rhythm, and voice quality - with a set of values for each is proposed, as reproduced here:

Tempo

a	allegro (fast)
aa	very fast
acc	accelerando (getting faster)
l	lento (slow)
ll	very slow
rall	rallentando (getting slower)

loud - Loudness

f	forte (loud)
ff	very loud
cresc	crescendo (getting louder)
p	piano (soft)
pp	very soft
dimin	diminuendo (getting softer)

pitch - Pitch range

high	high pitch range
low	low pitch range
wide	wide pitch range
narrow	narrow pitch range
asc	ascending
desc	descending
monot	monotonous
scand	scandent, each succeeding syllable higher than the last, generally ending in a falling tone

Tension

sl	slurred
----	---------

lax	lax, a little slurred
ten	tense
pr	very precise
st	staccato, every stressed syllable being doubly stressed
leg	legato, every syllable receiving more or less equal stress

Rhythm

rh	beatable rhythm
arrh	arrhythmic, particularly halting
spr	spiky rising, with markedly higher unstressed syllables
spf	spiky falling, with markedly lower unstressed syllables
glr	glissando rising, like spiky rising but the unstressed syllables, usually several, also rise in pitch relative to each other
glf	glissando falling, like spiky falling but with the unstressed syllables also falling in pitch relative to each other

voice - Voice Quality

whisp	whisper
breath	breathy
husk	husky
creak	creaky
fals	falsetto
reson	resonant
giggle	unvoiced laugh or giggle
laugh	voiced laugh

trem	tremulous
sob	sobbing
yawn	yawning
sigh	sighing

Table 3: Coding of changes in prosodic and paralinguistic features in TEI
(Sperberg-McQueen and Burnard, (Eds.) 1994)

Use of these features is exemplified in the TEI Guidelines:

" <u who=LB><shift feature=loud new=f> Elizabeth </u>
<u who=EB> Yes </u>
<u who=LB><shift> Come and try this <pause> <shift feature=loud
new=ff> come on </u> " (Sperberg-McQueen and Burnard, (Eds.) 1994 §
11.2.6)

For each <shift> element the feature that changes is identified and the new value is defined.

(4) Representation of stress and pitch patterns

The examples are the following:

.	low fall intonation
,	fall rise intonation
?	low rise intonation
!	rise fall intonation
-	truncated syllable
:	lengthened syllable

Table 4: Example of representation of stress and pitch patterns
(Sperberg-McQueen and Burnard, (Eds.) 1994)

When a detailed representation of stress and pitch patterns is needed, TEI recommends to maintain the prosodic transcription in parallel with the transcription of the text and make use of the alignment mechanisms which are described in the Guidelines.

References: TEI (Text Encoding Initiative)

JOHANSSON, S. (forthcoming) "Encoding a Corpus in Machine-Readable Form", in ATKINS, B.T.S. et al. (Eds.) *Computational Approaches to the Lexicon: An Overview*. Oxford: Oxford University Press.

JOHANSSON, S.- BURNARD, L.- EDWARDS, J.- ROSTA, A. (1991) *Working paper on spoken texts*. Document TEI AI2 W1.

SPERBERG-McQUEEN, C.M.- BURNARD, L. (Eds.) (1994) *Guidelines for Electronic Text Encoding and Interchange. TEI P3*. Chapter 11: Transcriptions of Speech. Association for Computational Linguistics / Association for Computers and the Humanities / Association for Literary and Linguistic Computing: Chicago and Oxford.

2.4.- NERC (Network of European Reference Corpora)

One of the aims of NERC (Network of European Reference Corpora) was to provide recommendations for the transcription of spoken language or "conventions for putting representations of the spoken language into machine readable form" (NERC, 1994:88) Thus, the final report of NERC devotes chapter 3B to the transcription of spoken language and chapter 5.2. to phonetic, phonemic and prosodic annotation. As in the TEI, the main emphasis was on the written representation of speech, but an important recommendation is that a corpus should contain a digitised representation of the sound wave temporally aligned with the transcription (NERC, 1994:94).

NERC uses a four level transcription system developed by French (1992) for the COBUILD project. Prosodic information is contained in Level Three and Level Four as follows:

Level Three contains identification of tone boundaries and tonic syllables:

/	tone unit boundary
CAPITALS	tonic syllables

Table 5: Conventions for the transcription of tone boundaries and tonic syllables in NERC (French , 1992)

Level Four contains the identification of head syllables and tone, using five basic tones for English. There is also an orthographic and a phonemic transcription aligned with a spectrogram and an Fo contour.

UNDERLINING	head syllables
‘	falling tone
'	rising tone
G	fall-rise
H	rise-fall
.	level tone

Table 6: Conventions for the transcription of head syllables and tone in NERC (French , 1992)

Some examples are provided:

" /The thing _G IS/ we're not going to let them get aWAY with it this time./
/I can't recall their having tried it on be_HFORE/. " (NERC, 1994:97)

As far as recommendations for prosodic annotations are concerned, the NERC report suggests the use of SAMPA (SAM Phonetic Alphabet) and SAMPROSA (SAM Prosodic Alphabet) (Teubert, 1993; NERC, 1994).

In a detailed report by Payne (1992) the conventions for transcribing spoken language developed within the TEI have been compared to those used by French (1992) and adopted by NERC. Payne discusses the problems with the use of the <s> tag for indicating tone units and with the use of <shift> to indicate changes in what TEI describes as "paralinguistic features". The need to provide a way to mark tonic syllables and tone levels in the orthographic transcription is stressed, this point being considered a major shortcoming for the coding of prosody within the TEI scheme.

References: NERC (Network of European Reference Corpora)

FRENCH, J.P. (1992) "Transcription proposals: multilevel system", Working paper, University of Birmingham, October 1992. NERC-WP4-50

NERC-1 (1994) *Network of European Reference Corpora. Final Report.* ILC-CNR-Pisa.

PAYNE, J. (1992) "Report on the compatibility of J P French's spoken corpus transcription conventions with the TEI guidelines for transcription of spoken texts", Working Paper, COBUILD Birmingham and IDS Mannheim, December 1992, NERC - WP8/WP4 -122

TEUBERT, W. (1993) "Phonetic/Phonemic and Prosodic Annotation" Final Report, IDS Mannheim, February 1993, NERC-WP8-171

2.5.- IPA (International Phonetic Alphabet)

The IPA (International Phonetic Alphabet) has a set of symbols for the representation of suprasegmental elements. On the occasion of the Kiel convention in 1989 a working group on Suprasegmental Categories coordinated by Gösta Bruce was set up (Bruce, 1988,1989). It was concluded that additions were needed to represent suprasegmentals within the IPA framework. As far as intonation was concerned, it was noted that there are no specific symbols for the notation of intonation - except for tones - in the IPA. Bruce's conclusions are that "there exists an apparent need for a direct way of symbolizing intonation in a phonetic transcription. However, the opinions diverge regarding the exact way of transcribing intonation. For a phonological transcription of intonation the symbolization is very much dependent on the language and the analysis" (Bruce, 1989: 36-37); the author concludes that " perhaps the best thing is to let the IPA Principles just state the problem and exemplify the techniques" (Bruce, 1989:37).

In the revised IPA chart (IPA, 1993) published in 1993 the following symbols appear under the heading "suprasegmentals":

'	Primary stress
---	----------------

"	Secondary stress
:	Long
;	Half-long
,	Extra-short
.	Syllable-break
	Minor (foot) group
	Major (intonation) group
˘	Linking (absence of a break)
Tones and word accents	
Level	
ˆ ˆˆˆˆ or ˘	Extra high
ˆˆ ˆˆ or ˘	High
- or ˘	Mid
ˆ or ˘	Low
˘˘ or ˘	Extra low
Y	Downstep
U	Upstep
Contour	
È or /	Rising
Ï or \	Falling
˘˘ or ˘	High rising
˘˘ or ˘	Low rising
˘ˆ or ˆ	Rising-falling
U	Global rise

\ Global fall

Table 7: Representation of suprasegmental elements in the IPA (IPA, 1993)

These symbols are not different from those published after the 1989 revision (IPA, 1989). It is worth noting that in the Kiel convention two systems were approved for the notation of pitch: the diacritical tone marks system -on the left-hand side of table 7- , in which pitch marks are placed above the segmented material and the "tone letters" system -on the right-hand side of table 7- in which marks are placed before or after segmented materials (IPA, 1989).

References: IPA (International Phonetic Association)

BRUCE, G. (1988) "2.3. Suprasegmental categories and 2.4. The symbolization of temporal events", *Journal of the International Phonetic Association* 18,2: 75-76

BRUCE, G. (1989) "Report from the IPA working group on suprasegmental categories", *Lund University Department of Linguistics and Phonetics, Working Papers* 35: 25-40

IPA (1989) "Report on the 1989 Kiel Convention", *Journal of the International Phonetic Association* 19,2: 67-80.

IPA (1993) "IPA chart, revised to 1993", *Journal of the International Phonetic Association* 23,1: center pages, unnumbered.

2.6.- TOBI (Tone and Break Index)

TOBI (Tone and Break Index Tear) was developed to fulfill the need of a prosodic notation system providing a common core to which different researchers can add additional detail within the format of the system; it focuses on the structure of American English, but transcribes word grouping and prominences, two aspects which are considered to be rather universal (Price, 1992). TOBI was has resulted from collaboration between academics and industrials; a workshop a MIT hosted by Victor Zue in 1991 and a second one at NYNEX hosted by Kim Silverman in 1992 have helped to come to an agreement.

According to Silverman *et al.* (1992) the system shows the following features: (1) it captures categories of prosodic phenomena; (2) it allows transcribers to represent some uncertainties in the transcription; (3) it can be adapted to different transcription requirements by using subsets or supersets of the notation system; (4) it has demonstrated high inter-transcriber agreement; (5) it defines ASCII formats for machine-readable representations of the transcription; and (6) it is equipped with software to support transcription using Waves™ and UNIX programmes.

A TOBI transcription for an utterance consists of symbolic labels for events on four parallel tiers: (1) orthographic tier, (2) break-index tier, (3) tone tier and (4) miscellaneous tier. Each tier consists of symbols representing prosodic events, associated to the time in which they occur in the utterance. The conventions for annotation according to TOBI are defined for text-based transcriptions and for computer-based labeling systems such as Waves™; a detailed description can be found in Hirschberg & Beckman.

(1) The Orthographic Tier

It is used for the transcription of orthographic words using ordinary spelling conventions

(2) The Break Index Tier

The break index tier specifies strength of coherence or the degree of disjuncture between adjacent words in the orthographic transcription. The break index is a rating for the degree of juncture between every pair of words and after the final word in the utterance. Values for the break index are chosen from the following set:

0	clear phonetic marks of clitic groups
1	most phrase-medial word boundaries

2	a strong disjuncture marked by a pause or virtual pause, but with no tonal marks; or a disjuncture that is weaker than expected at what is tonally a clear intermediate or full intonation phrase boundary
3	intermediate intonation phrase boundary
4	full intonation phrase boundary
-	affixed directly to the right of the break index indicates a transcriber's uncertainty about break-index strength

Table 8: Notation of break index values in TOBI (Hirschberg and Beckman)

Disfluencies

The perception of an audible hesitation is marked by the diacritic ‘p’ immediately to the right of the break index in the following manner:

1p	abrupt cutoff
2p	prolongation
3p	hesitation after the onset of the tonal marks for an intermediate phrase

Table 9: Notation of disfluencies in the break index tier in TOBI (Hirschberg and Beckman)

3.- The Tonal Tier

The tone tier specifies the tonal properties of the Fo contours of the utterance. These contours are represented as a sequences of pitch events.

Two types of tones are marked in the tone tear: phrasal tones and pitch accents.

Phrasal tones are pitch events associated with intonational boundaries; they are assigned at every intermediate or intonation phrase. The basic tone levels are defined in terms of the local pitch range and marked High or Low:

H	High tone level
L	Low tone level

Table 10: Notation of phrasal tone levels in TOBI (Hirschberg and Beckman)

Phrase Accents appear at an intermediate phrase boundary (level 3 and above) and are signaled as follows:

H-	High phrase accent
L-	Low phrase accent
!H-	Downstepped Phrase Accent

Table 11: Notation of phrase accent levels in TOBI (Hirschberg and Beckman)

Final Boundary Tones occur at every full intonation phrase boundary (level 4) and are transcribed in the following manner:

H%	High final boundary tone
L%	Low final boundary tone

Table 12: Notation of final boundary tone levels in TOBI (Hirschberg and Beckman)

The High initial Boundary Tone marks a phrase that begins relatively high in the speaker's pitch range when a high pitch at the beginning of an utterance can not be attributed to a high accent on the first or second syllable of the utterance and when the utterance contrasts with possible rendition with a lower-pitched onset.

%H	High initial boundary tone
----	----------------------------

Table 13: Notation initial boundary tone in TOBI (Hirschberg and Beckman)

Since intonation phases are composed of one or more intermediate phrases plus a boundary tone, full intonation phrase boundaries will have two final tones:

L- L%	the standard declarative contour of American English
L- H%	'continuation rise'
H- H%	as in a 'yes-no question' in American English
H- L%	a final level 'plateau'

Table 14: Notation of tone at intonation phrase boundaries in TOBI (Hirschberg and Beckman)

Disfluencies

%r is used to mark the left edge of an intonation phrase which begins after a hesitation or disfluency; indicates a contour restart after a disruption when the disfluency has caused a clear contour discontinuity.

%r	left edge of an intonation phrase which begins after a hesitation or disfluency
----	---

Table 15: Notation of disfluencies in the tone tier in TOBI (Hirschberg and Beckman)

Pitch accents are pitch events associated with accented syllables, or "pitch movements or configurations that lend prominence to their associated word" (Silverman et al., 1992); lack of pitch accent assignment for a syllable means that the syllable is not accented. Pitch Accents are transcribed as follows:

H* ‘	'peak accent': an apparent tone target on the accented syllable which is in the upper part of the speaker's pitch range for the phrase; includes tones in the middle of the pitch range but precludes very low Fo targets
L* ‘	'low accent': an apparent tone target on the accented syllable which is in the lowest part of the speaker's pitch range
L*+H	'scooped accent' : a low tone target on the accented syllable which is immediately followed by relatively sharp rise to a peak in the upper part of the speaker's pitch range
L+H*	'rising peak accent': a high peak target on the accented syllable which is immediately preceded by relatively sharp rise from a valley in the lowest part of the speaker's pitch range
H+!H*	a clear step down onto the accented syllable from a high pitch which itself cannot be accounted for by a H phrasal tone ending the preceding phrase or by a preceding H pitch accent in the same phrase

Downstepped Pitch Accents can also be transcribed as : !H*, L*, L*+!H, L*+!H*, !H+!H*

Table 16: Notation of pitch accents in TOBI (Hirschberg and Beckman)

There are means in TOBI to deal with under specification and uncertainty:

*	tonally unspecified pitch accent
-	tonally unspecified phrase accent
%	tonally unspecified boundary tone
*?	uncertain pitch accent
-?	uncertain phrase accent
%?	uncertain boundary tone

Table 17: Notation of transcriber's uncertainties in TOBI (Hirschberg and Beckman)

The following diacritics might be used in the transcription:

!	downstepped (high) tones; the diacritic precedes the downstepped pitch accent peak or downstepped H phrase accent
HiF0	marks the local pitch range for each intermediate phrase (interval between level 3 boundaries)

Table 18: Transcription diacritics in TOBI (Hirschberg and Beckman)

4.- Miscellaneous Tier

This tier is used for comments or markings desired by particular transcription groups. Events should be labeled at their temporal beginnings and endings with labels of the form 'event <...event>'

When the WAVES™ format is used, a speech waves and separate label files for the orthographic tier, the break index tier, the tonal tier and the miscellaneous tier are produced. In non-WAVES™ format, each line contains five fields, corresponding to the following items:

Field 1: the orthographic transcription; the syllable containing a pitch accent is marked by ' * ' before the vowel

Field 2: contains the tonal transcription, including pitch accents, phrase accents and boundary tone; phrasal accents are associated with the last word in the phrase

Field 3: contains the break index value, giving the strength of the break between the word on the current line and the word on the next line

Field 4: contains the time markers associated with the break indices

Field 5: contains miscellaneous information comments

Documentation on TOBI, tools for labeling with WAVES+™ and utterances transcribed can be accessed via anonymous ftp to kiwi.nmt.edu (129.138.1.82) (login:anonymous, password: your e-mail address)

Hirst (1994) remarks the difficulties in using TOBI as a multi-language prosodic annotation system, since it presupposes knowledge of the set of relevant pitch patterns for a given language.

References: TOBI (Tone and Break Index)

HIRSCHBERG, J.- BECKMAN, M. (1992) *Report on proposed transcription system and some recommendations*. unpublished ms.

HIRSCHBERG, J.- BECKMAN, M.E. *The ToBI Annotation Conventions*. unpublished ms

PRICE, P. (1992) Summary of the Second Prosodic Transcription Workshop: the TOBI (TOnes and Break Indices) Labeling System. Nynex Science and Technology, Inc. 5-6 April, 1992. *Linguist List* vol. 3-761, 9 October 1992.

SILVERMAN, K.- BECKMAN, M.- PITRELLI, J.- OSTENDORF, M.- WIGHTMAN, C.- PRICE, P.- PIERREHUMBERT, J.- HIRSCHBERG, J. (1992) "TOBI: A standard for labeling English prosody", *Proceedings of the Second International Conference on Spoken Language Processing, ICSLP-92*. Banff, October 1992. pp. 867-870

TERKEN, J.- OSTENDORF, M. Conventions for Non-WAVES™ format. unpublished ms.

2.7.- SAMPA (SAM Phonetic Alphabet)

SAMPA (SAM Phonetic Alphabet) is a multi-lingual computer-readable transcription system developed within the ESPRIT project 2589 SAM (Multilingual Speech Input/Output Assessment, Methodology and Standardization). Its aim is to provide ASCII encodings for the IPA symbols required for European languages. SAMPA includes a number of symbols for prosodic transcription, attempting to avoid any model-dependency. It is mainly intended to support signal-oriented labelling and provide a basis for cross-language comparisons. The SAMPA final standard system is presented in Wells *et al.* (1992)

The prosodic features considered in SAMPA are the following (each symbols is followed by its ASCII number):

:	58	length mark
"	34	primary stress and accent I words in Norwegian and Swedish

" "	34,34	accent II words in Norwegian and Swedish
%	37	secondary stress
-	45	level tone, if followed by a tone group boundary
'	39	rising tone
‘	96	falling tone
‘ ’	96,39	fall-rise
”“	39,96	rise-fall
\$	36	syllable boundary
+	43	morpheme boundary
#	35	word boundary
...	46,46,46	silent pause
	124	tone group/intonation phrase boundary
§	21	phonological phrase/rhythm group boundary
##	35,35	sentence boundary
-	45	separator (e.g. to distinguish hiatuses from diphthongs)

Table 19: SAMPA (SAM Phonetic Alphabet (Wells *et al.*, 1992))

Gibbon (1990) criticizes the theory-oriented character of the system and its inseparability from the tonetic theory of stress marking. However, more work on prosodic transcription has been carried out within the SAM project as will be shown below.

References: SAMPA (SAM Phonetic Alphabet)

SAM (1992) "Speech acquisition and Annotation Protocols and Index of Mnemonics (SAM-UCL-018)-Section IV: SAMPA" in *SAM User Guide to ETR Tools*. ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. Ref., SAM-UCL-G007.

WELLS, J.- BARRY, W.- GRICE, M.- FOURCIN, A.- GIBBON, D. (1992) *Standard Computer-Compatible Transcription*. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992. In SAM (1992) ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. *Final Report. Year Three: 1.III.91-28.II.1992*. Ref. SAM-UCL-G004. London: University College London.

WELLS, J.C. (1989) " Computer-coded phonemic notation of individual languages of the European Community ", *Journal of the International Phonetic Association* 19,1: 31-54

2.8.- PROSPA

PROSPA was developed by Margaret Selting and Dafydd Gibbon (Selting, 1987, 1988) specially to meet the needs of discourse and conversation analysis but has also been discussed within the Prosody Group in the ESPRIT 2589 SAM (Multilingual Speech Input/Output Assessment, Methodology and Standardization) project.

PROSPA is aimed at the high-level broad transcription which is needed for discourse analysis,. Intonation is defined as "the contour or melody of speech in terms of the temporal organization of perceived pitch of utterances" (Selting, 1987:779).The categories are based on auditive criteria. Transcriptions consist of :

- (1) an overall inclination or declination specified over the domain of an intonation unit
- (2) peaks and troughs internal to the unit
- (3) a final dynamic tone

Local categories are defined as accent and accent types or "short range pitch movements usually realized on lengthened vowels" (Selting, 1987:780); they include

+	upward pitch movement
-	downward pitch movement
=	level pitch accent

Table 20: Notation of pitch accents in PROSPA (Selting, 1987)

Since accents can be realized together with pitch changes, the following symbols are introduced:

U+	upward local pitch jump co-occurring with an upward accent
Y+	downward local pitch jump co-occurring with an upward accent

Table 21: Notation of pitch movements in PROSPA (Selting, 1987)

Global categories are defined according to rhythmical or pitch contour properties in a cohesive series of accents. Length of a global contour and the direction of pitch or tone level are indicated as follows:

()	extent of a sequence of cohesive accents
F	globally falling intonation

R	globally rising intonation
H	level intonation on high tone level
M	level intonation on middle tone level
L	level intonation on low tone level
H/F	falling intonation on a globally high tone level
...	sequence of weakly accented or unaccented syllables

Table 22: Notation of global categories in PROSPA (Selting, 1987)

The intonation after the last accent of a global unit - or "tails" - is noted after the parentheses in the following manner:

`	falling tails
/	rising tails
-	level tails
/`	combinations of tails (rising-falling here)

Table 23: Notation of "tails" in PROSPA (Selting, 1987)

Accent strength can be noted in this system by using repetitions of the symbol +. It is also possible to make use of the transcription line containing the orthographic texts by marking there primary and extra-strong accents.

Wells *et al.* (1992) report that the final tone inventory will have to be augmented in order to transcribe languages other than German, for which it was designed.

References: PROSPA

SELTING, M. (1987) "Descriptive categories for the auditive analysis of intonation in conversation", *Journal of Pragmatics* 11: 777-791

SELTING, M. (1988) "The role of intonation in the organization of repair and problem handling sequences in conversation", *Journal of Pragmatics* 12: 293-322.

WELLS, J.- BARRY, W.- GRICE, M.- FOURCIN, A.- GIBBON, D. (1992) *Standard Computer-Compatible Transcription*. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992. In SAM (1992) ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. *Final Report. Year Three: 1.III.91-28.II.1992*. Ref. SAM-UCL-G004. London: University College London.

2.9.- SAMSINT (SAM System for Intonation Transcription)

SAMSINT (SAM System for Intonation Transcription) has been defined by the SAM Prosody Working Group within the ESPRIT project 2589 SAM (Multilingual Speech Input/Output Assessment, Methodology and Standardization). It is intended to be a computer-readable system for the transcription of intonation contours within defined intonation units.

SAMSINT is based on INTSINT (see below), incorporating additional facilities and simplifications such as the specification of a global pitch direction over the domain of a tone group or an intonation unit as in PROSPA (see above) and simplifying boundary transcriptions which are transcribed independently, contrary to INTSINT (Wells *et al.*, 1992)

The movements of the intonation contours that can be symbolized using SAMSINT are the following (the ASCII number appears after the symbol):

T	84	top
B	66	bottom
+	43	higher
-	45	lower
^	94	upstep
!	33	downstep
>	62	same
[91	initial intonation unit boundary
]	93	final intonation unit boundary
/	47	global rising in an intonation unit
\	92	global falling in an intonation unit

Table 24: SAMSINT (Wells *et al.*, 1992)

The following points should be noted:

- Top and Bottom points are defined relative to a given pitch range.
- Higher, Lower, Upstep, Downstep and Same are defined relative to the point immediately previous within the intonation unit.

- The initial pitch of an intonation unit can be either Top or Bottom, or be unspecified.

References: SAMSINT (SAM System for Intonation Transcription)

WELLS, J.- BARRY, W.- GRICE, M.- FOURCIN, A.- GIBBON, D. (1992) *Standard Computer-Compatible Transcription*. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992. In SAM (1992) ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. *Final Report. Year Three: 1.III.91-28.II.1992*. Ref. SAM-UCL-G004. London: University College London.

2.10.- SAMPROSA

SAMPROSA has been proposed by Dafydd Gibbon, incorporating results from discussions within the SAM Prosody Working Group - ESPRIT project 2589 SAM (Multilingual Speech Input/Output Assessment, Methodology and Standardization)-. Sources for SAMPROSA according to Gibbon and Bleiching (1993), are traditional phonetic transcription marks taken from SAMPA, transcriptions used in discourse analysis - Selting (1987, 1988)- , autosegmental phonology - Goldsmith (1990) - and pitch-oriented phonetic elements such as those used in INTSINT (International Transcription System for Intonation, see below).

SAMPROSA is conceived as a descriptive representational system based on explicit theoretical assumptions. If SAMSINT was conceived by the SAM Prosody Working Group as " a system for transcribing intonation contours within defined intonation units, but making no further theoretical demands on the transcriber" (Wells *et al.* 1992:11), SAMPROSA was developed with a more phonologically-oriented approach.

The system consists in a vocabulary of symbols and a syntax for its combination. The vocabulary is an inventory in which each symbol is given an operational definition in terms of phonetic reproducibility and recognisability; the inventory is universal and allows definition of language-specific subsets. The symbols can be combined according to principles which are partly universal and partly language specific (Gibbon and Bleiching, 1993).

The symbols used in SAMPROSA are the following (ASCII number appear after the symbol):

Local tone		
H	72	High pitch
L	76	Low pitch
T	84	Top pitch (extreme H)
B	66	Bottom pitch (extreme L)
M	77	Mid pitch
+	43	Higher pitch

++	43,43	Much higher pitch
+ -	43,45	Peak (upward-downward)
-	45	Lower pitch
--	45,45	Much lower pitch
^	94	Upstep
^^	94,94	Wide upstep
!	33	Downstep
!!	33,33	Wide downstep
= or > or S	61,62 or 83	Level or same tone

Global tone: from Local and Nuclear tone repertoire

Terminal tone: from Local and Nuclear tone repertoire

Nuclear tone

-	45	Level tone (if followed by a tone group boundary)
' or / or R	39, 47 or 82	Rising tone
' or \ or F	96, 92 or 70	Falling tone
' ' (etc.)	96,39 (etc.)	Fall-rise
' ' (etc.)	39,96 (etc.)	Rise-fall

Length

:	58	Segment length mark
---	----	---------------------

Stress

"	34	Primary stress
%	37	Secondary stress

Pause

...	46,46,46	Silence
-----	----------	---------

Boundary

\$	36	Syllable boundary
----	----	-------------------

#	35	Word boundary
	124	Tone group boundary (non-directional)
[91	Tone group boundary (left)
]	93	Tone group boundary (right)
Metasymbols		
-	45	Separator (may be replaced by _)
*	42	Conjunctor

Table 25: SAMPROSA (Wells et al., 1992)

SAMPROSA can be applied in multi-tier transcription systems in which an signal is represented by means of an independent parallel symbolic representation. The signal and its representation can be related either by association - by defining phonological rules that relate prosodic and segmental units - or by synchronization - assigning symbols to the signal as tags or annotations -. The system is intended to be used in prosodic transcription for linguistic purposes and for prosodic labelling in speech technology and experimental phonetics research (Gibbon and Bleiching, 1993).

References: SAMPROSA

GIBBON, D. (1989) *Survey of Prosodic Labelling for EC Languages*. SAM-UBI-1/90, 12 February 1989; Report e.6, in ESPRIT 2589 (SAM) *Interim Report, Year 1*. Ref. SAM-UCL G002. London: University College London, February 1990.

GIBBON, D.- BLEICHING, D. (1993) *EAGLES Working Group 5: Spoken Language. Interim Report*. September 1993.

GOLDSMITH, J. (1990) *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.

WELLS, J.- BARRY, W.- GRICE, M.- FOURCIN, A.- GIBBON, D. (1992) *Standard Computer-Compatible Transcription*. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992. In SAM (1992) ESPRIT PROJECT 2589 (SAM) *Multilingual Speech Input/Output Assessment, Methodology and Standardisation. Final Report. Year Three: 1.III.91-28.II.1992*. Ref. SAM-UCL-G004. London: University College London.

2.11.- INTSINT (International Transcription System for Intonation)

INTSINT (International Transcription System for Intonation) aims at providing a system for cross-linguistic comparison of prosodic systems It has been developed by Daniel Hirst, based on a stylization procedure of the Fo (Fundamental frequency) contour build up from interpolation between points.

According to the description presented in Wells *et al.* (1992) transcription in INTSINT is based on prosodic target points aligned with an orthographic or phonetic

transcription. It can be used at different levels of detail, allowing a narrow as well as a broad phonetic transcription. Although it is conceived as a system for cross-language comparisons, language-specific subsets of elements can be recommended.

INTSINT is based on the postulate that "the surface phonological representations of a pitch curve can be assumed to consist of phonetically interpretable symbols which can in turn be derived from a more abstract phonological representation" (Hirst, 1991:307). The pitch contour - or pitch curve - can be represented as a sequence of pitch target points that can be interpolated by a function. In favour of this approach to the representation of pitch curves, Hirst (1991) quotes evidence from acoustic modelling studies showing that pitch targets account better for the data than pitch changes and from perceptual studies claiming that pitch patterns are predominantly interpreted in terms of pitch levels. INTSINT aims therefore at the symbolization of pitch levels or prosodic target points, each characterising a point in the fundamental frequency curve.

The symbolization of prosodic target points is made by means of arrow symbols corresponding to different pitch levels.

Higher, Upstepped, Lower, Downstepped or Same are tonal symbols describing relative pitch levels defined in relation to a previous pitch target or to the beginning of an intonation unit:

<i>U</i>	Higher
\	Lower
^o	Upstep
_o	Downstep
o	Same

Table 26: Notation of relative pitch levels in INTSINT (Wells et al., 1992)

Top or Bottom are tonal symbols describing absolute pitch levels described in relation to the operative range of the intonation unit:

U	Top
Y	Bottom

Table 27: Notation of absolute pitch levels in INTSINT (Wells et al., 1992)

Mid is assumed to occur only at the beginning of an intonation unit, and is then considered unmarked.

Hirst, Nicolas & Espesser (1991) have shown that, at least for French, the prosodic targets can be defined with respect to the speaker's Fo (Fundamental frequency) mean -

Mid-, to one point fixed at a half-octave interval above the mean - Top - and to one point fixed at a half-octave interval below the mean - Bottom -.

The Fo modelling is carried out automatically by a program called MOMEL (Hirst & Espesser, 1991) that, after Fo detection, provides the best fit for a sequence of parabolas, dividing the F0 curve into a microprosodic and a macroprosodic profile. The microprosodic component is caused by the individual segmental elements of the utterance, and the macroprosodic component reflects the intonation patterns produced by the speaker (Hirst & Espesser, 1991). The output of the programme is a sequence of target points with a time value in ms. and a frequency value in Hz. Target points can be then automatically coded into INTSINT symbols, once the position of the intonation unit boundaries has been manually introduced.

An experiment comparing listener's evaluation of a synthesized text using original target points and INTSINT-coded target points has shown that the INTSINT version attained more than 80% of the score attributed to the version synthesized with the original target points (Hirst, Nicolas & Espesser, 1991).

Within the MULTEXT project a tool will be developed for the automatic symbolic coding of Fo target points using INTSINT. A preliminary description of such an algorithm is given in Hirst (1994) (see also Hirst *et al.*, 1994) which attempts to provide an optimal INTSINT coding of a given curve by seeking to minimise the mean squares error of the predicted values from the observed values. Absolute pitch values Top, Mid and Bottom are modelled by their mean values and Relative pitch levels are modelled by a linear regression on the preceding target point.

References: INTSINT (International Transcription System for Intonation)

HIRST, D.J. (1991) "Intonation models: Towards a third generation" in *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol. 1 pp. 305-310

HIRST, D.J. (1994) "The symbolic coding of fundamental frequency curves: from acoustics to phonology", in FUJISAKI, H. (Ed) *Proceedings of International Symposium on Prosody*, Satellite Workshop of ICSLP 94, Yokohama, September 1994.

HIRST, D.J. (1994) *WP2 Corpus Annotation Tools. Task 2.6 Prosody tools. Task 2.7. Post-editing tools. Speech Tools. Milestone A2 Deliverables*. MULTEXT Report, 16 June 1994.

HIRST, D.J. - DI CRISTO, A. (Eds.) (forthcoming) *Intonation Systems. A Survey of 20 Languages*. Cambridge: Cambridge University Press.

HIRST, D.J. - DI CRISTO, A. (forthcoming) "A survey of intonation systems" in HIRST, D. - DI CRISTO, A. (Eds.) *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

HIRST, D.J. - DI CRISTO, A.- LE BESNERAIS, M.- NAJIM, Z.- NICOLAS, P.- ROMÉAS, P. (1993) "Multilingual modelling of intonation patterns", in HOUSE, D.- TOUATI, P. (Eds.) *Proceedings of an ESCA Workshop on Prosody*. September 27-29, 1993, Lund, Sweden. *Lund University Department of Linguistics and Phonetics, Working Papers* 41. pp. 204-207

HIRST, D.J. - ESPESSER, R. (1993) "Automatic modelling of fundamental frequency using a quadratic spline function", *Travaux de l'Institut de Phonétique d'Aix* 15: 71-85

HIRST, D.J. - IDE, N.- VÉRONIS, J. (1994) "Coding fundamental frequency patterns for multi-lingual synthesis with INTSINT in the MULTEXT project", *Proceedings of the ESCA/IEEE Workshop on Speech Synthesis*, New York, September 1994.

HIRST, D.J.- NICOLAS, P.- ESPESSER, R. (1991) "Coding the Fo of a continuous text in French: An experimental approach" in *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol. 5 pp. 234-237

WELLS, J.- BARRY, W.- GRICE, M.- FOURCIN, A.- GIBBON, D. (1992) *Standard Computer-Compatible Transcription*. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992. In SAM (1992) ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. Final Report. Year Three: 1.III.91-28.II.1992. Ref. SAM-UCL-G004. London: University College London.

3.- Comparison and evaluation of the systems

The second part of this report will compare the systems reviewed and will attempt to evaluate the needs of speech research and the needs of natural language processing as far as prosodic coding is concerned.

3.1.- Comparison of the systems

In a classical book on English intonation, Crystal (1969) discusses the principles that should guide a prosodic transcription system. According to him these are the following:

- (1) accuracy
- (2) consistency
- (3) be as automatically applicable as possible
- (4) use the minimum of symbols
- (5) establish degrees of complexity of symbols to reflect the different significance attached to the data
- (6) be broad, covering only those aspects which are linguistically significant.

On the overall, most of the systems examined in the first part of the report fulfill the conditions proposed by Crystal. However, it is clear that the systems described have been designed for different purposes and within different traditions. Nevertheless, it is possible to find some parameters that may help to compare the external features of each transcription system. The following dichotomies are suggested:

Multi-tiered vs. one-tiered systems

One-tiered systems include the symbols for the symbolization of prosodic events within the segmental - orthographic or phonetic / phonemic - transcription, while in multi-tiered systems it is possible to distinguish different layers or levels, separating the segmental transcription from the suprasegmental coding. Examples of one-tiered systems can be found in the domain of discourse analysis or text linguistics or in the conventions adopted by the TEI and NERC projects; IPA, SAMPA and its derivations

can be also classified within this category. TOBI and INTSINT are very clear examples of multi-tiered systems allowing the separation of different types of events from the segmental transcription. As far as the labelling of speech databases is concerned, the later systems seem to offer clear advantages.

Machine readable symbols vs. non-machine readable symbols

It can be seen in the tables presented in part one that in some of the transcription systems proposed the authors provide ASCII numbers for each symbol (e.g. SAMPA, SAMSINT or SAMPROSA). Other systems such as those used in discourse analysis and in corpus linguistics make use of characters which are usually available in computer keyboards; TOBI is another example of this category. However, some of the symbols used in the IPA, PROSPA or INTSINT are sometimes difficult to be reproduced using the conventional character set that is found on a personal computer. It seems that a prosodic coding system aimed at facilitation exchange of labelled databases should ideally make use of machine-readable symbols.

Systems that can be applied automatically vs. systems that rely on the transcriber's judgment

The great majority of the systems described in part one of this report depend on the transcriber's judgment, in the sense that the transcriber himself decides, after an auditory or acoustic analysis of the utterance, which is the symbol that more adequately reflects a given prosodic phenomenon. Only INTSINT can be automatically applied, taking the speech wave as a starting point and producing an abstract representation in a completely automatic way. Of course, this is an advantage when labelling of large speech databases has to be undertaken, since it ensures at least homogeneity of criteria.

It is also possible to compare the prosodic coding systems described so far by means of more "internal" parameters. The following are suggested:

Multilingual vs. non-multilingual systems

Systems such as TOBI or PROSPA have been developed having one language in mind. Others such as SAMPA or SAMSINT address European languages, and IPA and INTSINT have been designed to cover a wider range of languages - actually both of them contain the term "international" in their denomination -. For the purposes of a multilingual project like MULTEXT, it is essential that the coding system should be able to convey prosodic contrasts in a number of languages, and it seems logical to use a system conceived with that purpose.

Theory-oriented systems vs. "neutral" systems

Some authors explicitly claim that their system is not model-dependent; this is the case of SAMPA. On the contrary, other authors provide the theoretical background in which their coding system is based; examples are SAMPROSA, TOBI or INTSINT. In both cases the assumptions behind the system are of phonological nature, or are based on the author's conception of the phonetics-phonology interface.

On the other hand, the theory behind systems used in discourse and conversation analysis is provided by the needs, the practices and the models used in the field. The events which are coded are those which are known to be relevant in order to explain the discursive or the interactional behaviour of the speaker.

Types of prosodic events encoded in each system

The comparison of the different prosodic events that can be encoded by means of each of the systems described is not a trivial task due to the differences in terminology among the authors. The following elements seem to be common to many systems:

- Prosodic boundaries and prosodic units
- Tone or pitch level, terminal and non-terminal
- Pitch movements, pitch direction or pitch contour, both local and global
- Accent, at word or phrase level
- Lengthening
- Pauses

The range and the nature of phenomena which are covered by each system depends basically on its aim. While systems used in discourse analysis or in text linguistics tend to cover global phenomena and include sometimes information about the so-called "paralinguistic features" (TEI, for example), systems designed for an accurate representation of prosody such as TOBI, SAMSINT or SAMPROSA offer a detailed coding of boundaries, tone levels and pitch accents. INTSINT is essentially aimed at the symbolization of F_0 movements, and therefore the notational system concentrates on pitch levels.

3.2.- Prosody encoding and Natural Language Processing

To carry out a proper evaluation of the systems proposed for prosodic coding it would be necessary to take into account published work on prosodic parsing and related areas, in order to:

- (a) determine which information about prosody is necessary to improve higher level analysis such as syntactic parsing
- (b) determine which morphological and syntactic information is related to prosodic events and can be used to improve prosodic parsing.

However, the bibliography in this field has experienced an important growth in the recent years. The need to derive natural prosodic information from a written text in text-to-speech synthesis has probably been one of the reasons for some important breakthroughs in this area. Only some general remarks can be presented here, and the

reader interested in more details can refer to the papers presented at the sessions on "Speech synthesis and prosody", "Generation of prosody" during Eurospeech 91 (Eurospeech 91), on "Prosody IV: Phrasing", "Synthesis: Systems, Syntax, Prosody" during Eurospeech 93 (Eurospeech 93) and on "Grouping" and "Grouping and Discourse" during the ESCA Workshop on Prosody (House and Touati (Eds.) 1993).

It seems to be commonly agreed that prosody is related both to syntactic structure and to parts of speech; in a general introductory paper on corpus linguistics Leech and Fligelstone (1992:123) remark that "the syntactic structure of a sentence helps to determine, for example, its prosodic features such as pauses and the tone group boundaries. Similarly, grammatical analysis is needed to determine (say) whether a word pronounced /led/ should be written lead, or vice versa". These observations, and many others that can be found in the literature, show that both information from parsing and from tagging can be useful in assigning prosody.

3.2.1.- Prosody and parts of speech tagging

Knowles and Lawrence (1987) exemplify the relationship between grammatical tagging and intonation assignment. Working with a corpus of 100.000 words consisting on the recordings, an unpunctuated written transcription, an orthographic transcription, a prosodic transcription and a grammatically tagged text, they tried to automatically derive prosody from the written text and the tagging.

In their experiment, a prosodic transcription of the corpus was made by two phoneticians using a modified version of O'Connor and Arnold's (1971) system, and tagging was carried out using the LOB tagging programmes; it was concluded from the tagging process that "with minor amendments to the tagset, the programs originally designed for written texts can be adapted to deal with spoken texts" (Knowles and Lawrence, 1987:145).

In the automatic assignation of prosody, tone-grouping was carried out following a bottom-up approach that took into account grammatical information such as clitic attachment to lexical words, grammatical collocations such - for example adjective + noun or verb + adverb - and deaccentuation phenomena in grammatical words

Although the paper does not offer qualitative results, it shows that it is possible to devise rules to automatically assign intonation to a written text using information from a tagger.

3.2.2.- Prosody and syntactic parsing

The relationship between syntactic parsing and prosody has been recently studied in the development of text-to-speech synthesis systems, although an extensive literature on the interaction between prosody and syntax deals with more theoretically-oriented problems such as disambiguation (see, for example, Lehiste (1973), Price *et al.* 1991), prosodic marking of syntactic boundaries (see Cooper and Sorensen, 1971, 1981, Cooper and Paccia-Cooper, 1980, Strangert, 1992 or the papers from the session on "Phrasing" at Eurospeech'93) or the syntax-phonology interface (see Ladd, 1986 or Nespor and Vogel, 1986 as classical contributions).

One of the major topics in this area seems to be the congruence between prosodic and syntactic boundaries. Strangert and Strangert (1993: 1210) conclude that as far as perceived pauses at syntactic boundaries -sentence, clause and phrase boundaries - are concerned "it is possible to differentiate between syntactic boundaries on the basis of prosodic cues alone". Perception of word boundaries seems to be determined by pauses, melodic discontinuities and declination resets (de Pijper and Sanderman, 1993).

A more detailed analysis of the interaction between syntactic and prosodic boundaries is found in Beaugendre and Lacheret-Dujour (1993). They suggest that prosodic groups correspond in most of the cases to syntactic constituents, and define different types of prosodic boundaries linked to syntactic categories. Major boundaries, for example, close a subject composed by a nominal group or precede a subordinate clause; minor boundaries close a verbal group or a lexical word complement of a verb or precede a word which begins by a preposition or a determiner. The conclusion of their work is a set of "intonosyntactic rules" that involve morphological and syntactic information.

The importance of boundary strength is stressed in many other studies dealing with automatic assignation of prosody in text-to-speech synthesis. Use of break indices such as those proposed in TOBI has been made in order to provide prosodic disambiguation of syntactically ambiguous sentences in text-to-speech synthesis experiments (Campbell and Whightman, 1992).

It is possible to ask from a different perspective which is the prosodic information that could be useful to improve syntactic parsing. Although the literature on parsing does not seem to provide an immediate answer to this question, it does seem possible to think that the marking of prosodic boundaries and pauses would be helpful, at least as far as disambiguation is concerned. Also pitch movements at certain points in the utterance could provide important information for the determination of syntactic structure. However, this is a major research topic in itself and can not be the object of an in-depth treatment here.

4.- Glossary

Accent: "Accent may refer to prominence given to a syllable, usually by the use of pitch [...] In this sense, accent is distinguished from stress, which is more often used to refer to all sorts of prominence (including prominence resulting from increased loudness, length or sound (quality), or to refer to the effort made by the speaker in producing a stressed syllable" (Roach, 1992: 1)

Contour: "A movement of the pitch of the voice in speech" (Roach, 1992: 27)

Declination: "The tendency for the pitch to fall throughout a syntactic unit such as a sentence" (Ladefoged, 1993:293)

Duration: "The amount of time that a sound lasts. In the study of speech it is usual to use the term length for the listener's impression of how long a sound lasts for, and duration for the physical, objectively measurable time" (Roach, 1992: 33)

Frequency: "The rate of variation in air pressure in a sound" (Ladefoged, 1993:293)

Fundamental frequency (Fo) : "The frequency of vibration of the vocal folds" (Roach, 1992: 45)

Head: "In the standard British treatment of intonation, one of the components of the tone unit" (Roach, 1992: 51)

Intensity: "The amount of acoustic energy in a sound" (Ladefoged, 1993:294)

Intonation: "The pattern of pitch changes that occur during a phrase, which may be a complete sentence" (Ladefoged, 1993:294)

Level tone: "A tone produced with an unchanging pitch level" (Roach, 1992: 66)

Loudness: "The auditory property of a sound that enables a listener to place it on a scale going from soft to loud without considering the acoustic properties, such as the intensity of the sound" (Ladefoged, 1993:294)

Nucleus: "The most prominent syllable of the tone unit" (Roach, 1992: 74)

Pause: "A break in the flow of continuous speech" (Roach, 1992: 78)

Pitch: "The auditory property of a sound that enables a listener to place it on a scale going from low to high, without considering the acoustic properties, such as the frequency of the sound" (Ladefoged, 1993:296)

Prominence: "The extent to which a sound stands out from others because of its sonority, length, stress and pitch" (Ladefoged, 1993:294)

Prosody: "Features of speech (such as pitch) that can be added to the basic segments, usually to a sequence of more than one sound" (Roach, 1992: 87)

Sentence stress: "The placing of the strongest stress (or accent) on a syllable, or word of a particular sentence" (Roach, 1992: 98)

Stress: "A property of syllables which make them stand out as more noticeable than others" (Roach, 1992: 102)

Suprasegmental: "Phonetic features such as stress, length, tone and intonation, which are not properties of single consonants or vowels" (Ladefoged, 1993:296)

Tone group: "The part of a sentence over which a particular intonation pattern extends" (Ladefoged, 1993:297)

Tone unit: "A unit of speech consisting of one or more syllables" (Roach, 1992: 113)

Tone: "A pitch that conveys part of the meaning of a word" (Ladefoged, 1993:297)

Tonic syllable: "The syllable within a tone group that stands out because it carries the major pitch change: (Ladefoged, 1993:297)

Word stress: "The stress pattern of a word" (Roach, 1992: 124)

5.- References

ATKINSON, J.M. - HERITAGE, J. (Eds.) (1984) Structures of social action. Studies in conversation analysis. Cambridge / Paris: Cambridge University Press / Editions de la Maison des Sciences de l'Homme

BEAUGENDRE, F.- LACHERET-DUJOUR, A. (1993) "Automatic generation of French intonation based on a perceptual study and morphosyntactic information", in *Eurospeech'93. 3rd European Conference on Speech Communication and Technology*. Berlin, Germany, 21-23 September 1993. Vol. 2 pp. 1219-1222

BRUCE, G. (1988) "2.3. Suprasegmental categories and 2.4. The symbolization of temporal events", *Journal of the International Phonetic Association* 18,2: 75-76

BRUCE, G. (1989) "Report from the IPA working group on suprasegmental categories", *Lund University Department of Linguistics and Phonetics, Working Papers* 35: 25-40

CAMPBELL, W.N.- WIGHTMAN, C.W. (1992) "Prosodic encoding of syntactic structure for speech synthesis", in *ICSLP 94 Proceedings of the International Conference on Spoken Language Processing*, October 12-26 1992, Banffs, Alberta, Canada. Vol. 2 pp. 1167-1170

COOPER, W.E. - SORENSEN, J.M. (1981) *Fundamental Frequency and Sentence Production*. New York: Springer Verlag.

COOPER, W.E.- PACCIA-COOPER, J. (1980) *Syntax and Speech*. Cambridge, MA: Harvard University Press.

COOPER, W.E.- SORENSEN, J.M. (1971) "Fundamental Frequency Contours at Syntactic Boundaries", *Journal of the Acoustical Society of America* 62,3: 683-692

CRUTTENDEN, A. (1986) *Intonation*. Cambridge: Cambridge University Press (Cambridge Textbooks in Linguistics).

CRYSTAL, D. (1969) *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press (Cambridge Studies in Linguistics, 1)

CUTLER, A.- LADD, R.D. (Eds.) (1983) *Prosody. Models and Measurements*. Heidelberg: Springer Verlag.

DE PIJPER, J.R.- SANDERMAN, A. (1993) "Prosodic cues to the perception of constituent boundaries" in *Eurospeech'93. 3rd European Conference on Speech Communication and Technology*. Berlin, Germany, 21-23 September 1993. Vol. 2 pp. 1211-1214

DU BOIS, J.W. (1991) "Transcription design principles for spoken discourse research", *Pragmatics* 1: 71-106

DU BOIS, J.W.- SCHUETZE-COBURN, S.- CUMMING, S.- PAOLINO, D. (1993) "Outline of discourse transcription" in EDWARDS, J.A.- LAMPERT, M.D. (Eds.) *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NY: Lawrence Erlbaum. pp. 45-89.

Eurospeech 91. 2nd European Conference on Speech Communication and Technology. Genova, Italy, 24-26 September 1991. 3 vols.

Eurospeech'93. 3rd European Conference on Speech Communication and Technology. Berlin, Germany, 21-23 September 1993. 3 vols.

FRENCH, J.P. (1992) "Transcription proposals: multilevel system", Working paper, University of Birmingham, October 1992. NERC-WP4-50

GIBBON, D. (1989) Survey of Prosodic Labelling for EC Languages. SAM-UBI-1/90, 12 February 1989; Report e.6, in ESPRIT 2589 (SAM) *Interim Report, Year 1*. Ref. SAM-UCL G002. University College London, February 1990.

GIBBON, D.- BLEICHING, D. (1993) *EAGLES Working Group 5: Spoken Language. Interim Report*. September 1993.

GOLDSMITH, J. (1990) *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.

GRØNNUM THORSEN, N. (1987) "Suprasegmental transcription", ARIPUC, Annual Report of the Institute of Phonetics, University of Copenhagen 21: 1-28

GUMPERZ, J.J.- BERENZ, N. (1993) "Transcribing Conversational Exchanges", in EDWARDS, J.A.- LAMPERT, M.D. (Eds.) *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, N.J.: Lawrence Erlbaum Associates. pp. 91-122

HART, J. T- COLLIER, R.- COHEN, A. (1990) *A Perceptual Study of Intonation. An Experimental - Phonetic Approach to Intonation*. Cambridge: Cambridge University Press. (Cambridge Studies in Speech Science and Communication)

HIRSCHBERG, J.- BECKMAN, M. (1992) *Report on proposed transcription system and some recommendations*. unpublished ms.

HIRSCHBERG, J.- BECKMAN, M.E. *The ToBI Annotation Conventions*. unpublished ms

HIRST, D. (1991) "Intonation models: Towards a third generation" in *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol. 1 pp. 305-310

HIRST, D. (1994) *WP2 Corpus Annotation Tools. Task 2.6 Prosody tools. Task 2.7. Post-editing tools. Speech Tools. Milestone A2 Deliverables*. MULTTEXT Report, 16 June 1994.

HIRST, D. - DI CRISTO, A. (Eds.) (forthcoming) *Intonation Systems. A Survey of 20 Languages*. Cambridge: Cambridge University Press.

HIRST, D. - ESPESSER, R. (1993) "Automatic modelling of fundamental frequency using a quadratic spline function", *Travaux de l'Institut de Phonétique d'Aix* 15: 71-85

HIRST, D.- DI CRISTO, A. (forthcoming) "A survey of intonation systems" in HIRST, D. - DI CRISTO, A. (Eds.) *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

HIRST, D.- DI CRISTO, A.- LE BESNERAIS, M.- NAJIM, Z.- NICOLAS, P.- ROMÉAS, P. (1993) "Multilingual modelling of intonation patterns", in HOUSE, D.- TOUATI, P. (Eds.) *Proceedings of an ESCA Workshop on Prosody*. September 27-29, 1993, Lund, Sweden. *Lund University Department of Linguistics and Phonetics, Working Papers* 41. pp. 204-207

HIRST, D.- NICOLAS, P.- ESPESSER, R. (1991) "Coding the Fo of a continuous text in French: An experimental approach" in *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol. 5 pp. 234-237

HIRST, D.J. (1994) "The symbolic coding of fundamental frequency curves: from acoustics to phonology", in FUJISAKI, H. (Ed) *Proceedings of International Symposium on Prosody*, Satellite Workshop of ICSLP 94, Yokohama, September 1994.

HIRST, D.J. - IDE, N.- VÉRONIS, J. (1994) "Coding fundamental frequency patterns for multi-lingual synthesis with INTSINT in the MULTTEXT project", *Proceedings of the ESCA/IEEE Workshop on Speech Synthesis*, New York, September 1994.

HOUSE, D.- TOUATI, P. (Eds.) (1993) Proceedings of an ESCA Workshop on Prosody. September 27-29, 1993, Lund, Sweden. *Lund University Department of Linguistics and Phonetics, Working Papers* 41.

IPA (1989) "Report on the 1989 Kiel Convention", *Journal of the International Phonetic Association* 19,2: 67-80.

IPA (1993) "IPA chart, revised to 1993", *Journal of the International Phonetic Association* 23,1: center pages, unnumbered.

JOHANSSON, S. (forthcoming) "Encoding a Corpus in Machine-Readable Form", in ATKINS, B.T.S. et al. (Eds.) *Computational Approaches to the Lexicon: An Overview*. Oxford: Oxford University Press.

JOHANSSON, S.- BURNARD, L.- EDWARDS, J.- ROSTA, A. (1991) *Working paper on spoken texts*. Document TEI A12 W1.

KNOWLES, G. (1991) "Prosodic labelling: the problem of tone group boundaries", in JOHANSSON, S.- STENSTRÖM, A. (Eds.) *English Computer Corpora. Selected Papers and Research Guide*. Berlin: Mouton de Gruyter. pp. 149-163

KNOWLES, G.- LAWRENCE, L. (1987) "Automatic intonation assignment" in GARSIDE, R.- LEECH, G.- SAMPSON, G. (Eds.) *The Computational Analysis of English: A Corpus-based Approach*. London: Longman. pp. 139-148

LADD, D.R. (1986) "Intonational phrasing: the case for recursive prosodic structure", *Phonology Yearbook* 3: 311-340

LADEFOGED, P. (1975) *A Course in Phonetics*. New York: Harcourt, Brace, Jovanovich, 1982 2nd ed., 1993 3rd ed.

LEECH, G. (1991) "The State of the Art in Corpus Linguistics" in AIJMER, K.- ALTENBERG, B. (Eds.) *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. London: Longman. pp. 8-29.

LEECH, G.- FLIGELSTONE, S. (1992) "Computers and corpus analysis" in BUTLER, C.S. (Ed) (1992) *Computers and Written Texts*. Oxford: Basil Blackwell. pp. 115-140

LEHISTE, I. (1970) *Suprasegmentals*. Cambridge, Mass.: The MIT Press.

LEHISTE, I. (1973) "Phonetic Disambiguation of Syntactic Ambiguity", *Glossa* 7,2 : 107-121

LÉON, P.- MARTIN, P. (1970) *Prolegomènes à l'étude des structures intonatives*. Montréal: Didier (Studia Phonetica 2).

NERC-1 (1994) *Network of European Reference Corpora. Final Report*. ILC-CNR-Pisa.

NESPOR, M.- VOGEL, I. (1986) *Prosodic Phonology*. Dordrecht: Foris (Studies in Generative Grammar, 28)

O'CONNOR, J.D.- ARNOLD, G.F. (1961) *Intonation of Colloquial English*. London: Longman, 1973 2nd. edition.

PAYNE, J. (1992) "Report on the compatibility of J P French's spoken corpus transcription conventions with the TEI guidelines for transcription of spoken texts", Working Paper, COBUILD Birmingham and IDS Mannheim, December 1992, NERC - WP8/WP4 -122

PICKETT, J.M. (1980) *The Sounds of Speech Communication. A Primer of Acoustic Phonetics and Speech Perception*. Baltimore: University Park Press. Austin: Pro-Ed.

PRICE, P. (1992) Summary of the Second Prosodic Transcription Workshop: the TOBI (TOnes and Break Indices) Labeling System. Nynex Science and Technology, Inc. 5-6 April, 1992. *Linguist List* vol. 3-761, 9 October 1992.

PRICE, P.J.- OSTENDORF, M.- SHATTUCK-HUFNAGEL, S.- FONG, C. (1991) "The use of prosody in syntactic disambiguation", *Journal of the Acoustical Society of America* 90, 6:2956-2970

ROACH, P. (1992) *Introducing Phonetics*. London: Penguin (Penguin English Linguistics)

ROSSI, M.- DI CRISTO, A.- HIRST, D.J. - MARTIN, P.- NISHINUMA, Y. (1981) *L'intonation. De l'acoustique à la sémantique*. Paris: Klincksieck.

SAM (1992) *Guide to EUROM.1 Speech Database*. Doc no. SAM-NPL-102, Final version 21 April 1992.

SAM (1992) "Speech acquisition and Annotation Protocols and Index of Mnemonics (SAM-UCL-018)-Section IV: SAMPA" in *SAM User Guide to ETR Tools*. ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. Ref., SAM-UCL-G007.

SELTING, M. (1987) "Descriptive categories for the auditive analysis of intonation in conversation", *Journal of Pragmatics* 11: 777-791

SELTING, M. (1988) "The role of intonation in the organisation of repair and problem handling sequences in conversation", *Journal of Pragmatics* 12: 293-322.

SILVERMAN, K.- BECKMAN, M.- PITRELLI, J.- OSTENDORF, M.- WIGHTMAN, C.- PRICE, P.- PIERREHUMBERT, J.- HIRSCHBERG, J. (1992) "TOBI: A standard for labeling English prosody", *Proceedings of the Second International Conference on Spoken Language Processing, ICSLP-92*. Banff, October 1992. pp. 867-870

SPERBERG-McQUEEN, C.M.- BURNARD, L. (Eds.) (1994) *Guidelines for Electronic Text Encoding and Interchange. TEI P3*. Chapter 11: Transcriptions of Speech. Association for Computational Linguistics / Association for Computers and the Humanities / Association for Literary and Linguistic Computing: Chicago and Oxford.

STRANGERT, E. (1992) "Prosodic cues to the perception of syntactic boundaries" in *Proceedings of the International Conference on Spoken Language Processing*, Banff, Alberta, Canada, 1992. Vol. 2, pp. 1283-1285

STRANGERT, E.- STRANGERT, B. (1993) "Prosody in the perception of syntactic boundaries", in *Eurospeech'93. 3rd European Conference on Speech Communication and Technology*. Berlin, Germany, 21-23 September 1993. Vol. 2 pp. 1209-1210

SVARTVIK, J. (Ed) (1990) *The London-Lund Corpus of Spoken English: Description and Research*. Lund: Lund University Press.

SVARTVIK, J.- EEG-OLOFSSON, M.- FORSHEDEN, O.- ORENSTRÖM, B.- THAVENIUS, C. (1982) *Survey of Spoken English. Report on Research 1975-1981*. Lund: Lund University Press.

TAYLOR, L.- KNOWLES, G. (1988) *Manual of Information to Accompany the SEC Corpus*. UCREL, University of Lancaster

TERKEN, J.- OSTENDORF, M. Conventions for Non-WAVES™ format. unpublished ms.

TEUBERT, W. (1993) "Phonetic/Phonemic and Prosodic Annotation" Final Report, IDS Mannheim, February 1993, NERC-WP8-171

WELLS, J.- BARRY, W.- GRICE, M.- FOURCIN, A.- GIBBON, D. (1992) *Standard Computer-Compatible Transcription*. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992. In SAM (1992) ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. Final Report. Year Three: 1.III.91-28.II.1992. Ref. SAM-UCL-G004. London: University College London.

WELLS, J.C. (1989) " Computer-coded phonemic notation of individual languages of the European Community ", *Journal of the International Phonetic Association* 19,1: 31-54

WICHMANN, A. (1991) "A study of up-arrows in the Lancaster/IBM Spoken English Corpus", in JOHANSSON, S.- STENSTRÖM, A. (Eds.) *English Computer Corpora. Selected Papers and Research Guide*. Berlin: Mouton de Gruyter. pp. 165-178