

ToBIG and PROLAB

Two prosodic transcription systems for German compared

K. J. Kohler
IPDS Kiel, Germany

Report at Workshop on Prosodic Labelling
ICPhS Stockholm, 13 August 1995

1. Preliminaries

There are at present two symbolic systems that have been developed for the computer compatible transcription of German prosody:

- ToBIG [2], a modified version of ToBI [1a,b], which is a consensus system among different German labelling groups using ToBI, and
- PROLAB [3], an independently devised notation framework.

The latter is the older of the two [4]. It is based on KIM, the Kiel Intonation Model for German [5a,b,c], which in turn has been implemented in the Infovox TTS system for German as a commercial product as well as a research environment for testing the model and its symbolic representation [6a,b,c]. PROLAB has also been applied to the prosodic labelling task taken on by IPDS Kiel within the VERBMOBIL project [7].

A ToBI inventory was first presented for German in 1994, within VERBMOBIL [8], to be used by labellers with an engineering background at Braunschweig University. In parallel other ToBI modifications were proposed in linguistic environments at the Universities of Saarbrücken and Stuttgart [3,9]. As a result of discussions held by the three ToBI groups (Braunschweig, Saarbrücken, Stuttgart) at Stuttgart in February 1995, a unified single ToBIG system emerged [3], which is the basis for this comparison with PROLAB.

During a one-day workshop held at IPDS Kiel in April 1995 representatives of the ToBIG system from Braunschweig and Saarbrücken, and of the PROLAB system from Kiel compared ToBI and PROLAB transcriptions of the same spontaneous MAP TASK and VERBMOBIL dialogue speech data and discussed correspondences and divergencies as well as the underlying theoretical issues. As a follow-up from this meeting two dialogue turns from the Kiel VERBMOBIL and 23 sentence-type utterances from the Saarbrücken MAP TASK corpus were selected for transcription by both Kiel and Saarbrücken in their respective systems in preparation of a comparative report to be given at the Stockholm labelling workshop by K. J. Kohler. At the same time the speech data were circulated among the prospective German participants at this symposium for individual labelling in fulfilment of the organizers' stipulation in their circular letter of 18 May 1995.

Since the Stuttgart and Kiel meetings a few more people have joined the group of German participants at the Stockholm workshop, and Anton Batliner, Andreas Kiessling and Matthias Reyelt have solicited a questionnaire from each of them on their respective labelling systems:

- Esther Grabe: follows ToBI Saarbrücken, i.e. ToBIG, for the moment but works on an adaptation, which she has not specified; so she can be grouped together with ToBIG.
- Carsten Guenther: likewise, since he applies the inventory of [8].
- Andreas Kiessling (and Anton Batliner, ERBA-Korpus): this labelling is ""grammatical" (syntactic) automatic labelling of boundaries and accents" with "no formal labelling of intonation" but drawing on a repertoire of 4 phrase boundary labels B1-3 plus B0 for simple word boundaries, and 3 related phrase accent labels A1-3 (plus A0 for unaccented syllables); these label categories are different from the corresponding ones used in [8]; the essential ingredient of prosodic labelling, i.e. the specification of intonation, is lacking, furthermore the very limited inventory has only been applied to read speech, and since special reference is made to a mapping onto boundary and functional (accent) tiers in [8], the system need not be considered separately in this report.
- Bernd Möbius: uses "internal conventions" laid down in an "internal documentation", which he classifies as "not available", and covers "word boundaries, syntactic phrase boundaries (based on punctuation), intonational phrase boundaries (1 level), perceived syllable stress (1 level), hesitations and laryngealizations, sentence mode"; since his labelling system is not specified it cannot be compared with ToBIG and PROLAB, quite apart from the fact that again it lacks the essential ingredient of a prosodic labelling frame; so it will have to be left aside in this report.
- Thomas Portele, Barbara Heuft: their labelling system is classified as "own development" but not specified; it is moreover largely "signal labelling" with signal measures (F0, duration) as prosodic parameters, supplemented by some "perceptual labelling", referring to "place of boundaries" and "syllable prominence" (to the extraordinary delicacy of 32 levels!); it thus falls outside the scope of the symbolic labelling systems we have to deal with in this report.

Some of my remarks will apply to ToBI, the parent frame, in general, some to ToBIG, the off-spring, in particular. In pursuing my goal of critically comparing ToBIG and PROLAB, as in the evaluation of any labelling system, three aspects have to be taken into account:

- the theoretical bases of the systems
- the categories of prosodic models for German (descriptive adequacy)
- the symbolic inventories, their syntax and application.

Before the question of labelling reliability between different labellers and the same labellers at different times can be successfully tackled, the validity of labels as descriptively adequate representations of prosodic speech phenomena has to be ascertained. As I shall show a great deal of research still has to be done to achieve this validity criterion so that any reliability statistics carried out on labelled files are premature.

2. The theoretical bases of ToBIG and PROLAB

2.1. The phonetics/phonology issue

Both ToBI and PROLAB stress the point that they represent prosodic phonologies and therefore only symbolize contrastive differences, leaving contextual phonetic effects, e.g. microprosody, unmarked. But PROLAB applies this theoretical stand more consistently. Thus it symbolizes all perceptually determined (low-ending) medial peaks by "&2 &2.", irrespective of differences of F0 maximum alignment with the stressed vowel nucleus due to syllable structure, number of syllables, voicing state of preceding and/or following consonants. ToBI, however, introduced "<" and ">" for shifts of H* or L+H* to a later or earlier time position in the stressed syllable, and ToBIG has taken this over.

Furthermore, in PROLAB all categories are perceptually defined, usually involving bundles of acoustic properties, e.g. F0, duration, intensity for sentence stresses, or segmental lengthening, F0 patterning before and after prosodic phrase boundaries. And these categories are allocated by labellers to speech files in a procedure of paradigmatic contrasting. It establishes, e.g., whether an F0 peak contour on "Toby" or "Anna" is a medial peak, in spite of a continuously falling F0, on a par with a rising-falling F0, for example on "Marianna"; it categorically contrasts this with an early peak, which would manifest itself as a continuously falling F0 starting at a lower point in "Toby" and "Anna" and as a rise-fall synchronized on "i" of "Marianna". In ToBI, on the other hand, the acoustic property of the F0 pattern is the principal guideline for the definition of accents and break indices, and in the above cases a medial peak on "Toby", "Anna" is marked as H*, but as L+H* on "Marianna", although they represent the same phonological category, with phonetic variation due to the position of the accented syllable in the utterance. Contrariwise, "but/aber Anna", "but/aber Toby" with medial peaks are treated like "Marianna" and marked as L+H*, again for contextually determined phonetic reasons, where PROLAB uses the same symbolization for the same phonological categories, disregarding phonetic variation.

ToBIG practises the same inconsistency in the application of the phonologizing principle. It is, e.g., not clear how the notational series of H+!H*, H*>, H*, H*<, L+H*>, L+H*, L+H*<, L*+H (in each case followed by L-L%) are distributed across the prosodic phonological categories of early, medial, and late peaks. But one thing is clear: they cannot all be different phonological categories, but must in part constitute low-level phonetic variation, which should not be represented symbolically, since it is more efficiently retrieved at the signal level.

A comparison with PROLAB notations is thus severely hampered, all the more so as ToBIG uses the terms 'late/early peak' (e.g. with reference to L+H*< or (H+!)H*) without, however, meaning the same prosodic categories as PROLAB. In [8] H* is even equated with PROLAB's 'early peak', but is then used in a way that corresponds to its 'medial peak', for which ToBIG, like ToBI, also uses L+H* as a phonetic variant. So in spite of the explicit reference to KIM and its notational system in [8], the authors lack a complete understanding of the use of peak categories in the Kiel framework.

ToBIG has also added a further (complex) pitch accent to the above-mentioned series: H+L*. No reasons are given why ToBI was modified and the Pierrehumbert notation [10] reintroduced, nor why the exclusion of the new symbolization, expressly stated in [8], was reversed. Was it on language-independent grounds or because the labelling of German made it mandatory, and if so, was it in order to symbolize phonological distinctions or to mark phonetic exponency? The definition that is supplied speaks in favour of the latter interpretation: "'step-down to low' a preaccented high or mid target followed by a target on the

accented syllable which is clearly at or very near the bottom of the speaker's range" [3, p.5]. On scientific principles one would expect that no changes are introduced into a standard labelling system without detailed justification for the modifications.

'Downstep' is another instance of marking phonetic constraints in ToBI(G). PROLAB treats it as what it is: the downdrift within a prosodic phrase (not on a time scale, but on a structural basis from sentence accent to sentence accent), as the phonetic default, which needs no symbolization. Lack of 'downstep' and 'upstep', however, have to be marked as phonological deviations from default. The latter category is not marked at all in ToBI(G). 'Upstep' here only means phonetic realization of H-L% and H-H%. A sequence of upsteps in, for instance, "What on earth are you doing that for?" (as against the equally possible downstepped sequence) has no symbolization in ToBI(G), but it should because these are clearly phonological differences that perform functions on the semantic and pragmatic levels. PROLAB has recognized and filled this need by the introduction of "|" after the digit accent marker. Cf. also the German utterance "rote gelbe blaue grüne weiße schwarze" as a sequence of upstepped falls or rises or a sequence of downstepped falls.

2.2. The atomistic linear vs the molecular superpositional approach

PROLAB contrasts with ToBI(G) in that it does not represent pitch patterns as linear sequences of elementary tones – H and L – entering into serial orders of pitch accent and boundary tone units. It replaces this atomistic approach by a more global one that recognizes whole pitch peak and valley contours as prosodic molecules for concatenation, without the need for a further decomposition into successive elements at the levels of prosodic phonology and hence prosodic labelling. The different peaks (8) and valleys (4) are defined as holistic entities broadly synchronized with the onset of sentence-stressed vowels (early, medial, late peak/valley), the finer details of F0 pattern and alignment with the sound segments being left unspecified at this level of phonological contrasts. This more abstract prosodic representation has a number of advantages over the marking of concrete pitch levels:

- A great variety of sometimes widely differing phonetic shapes (in terms of H and L), due to segmental set-up, syllable structures, position in utterance, speech rate can all be subsumed under the same symbolization if they perceptually represent the same category in the prosodic phonology of the particular language; it is thus no longer odd if peak maxima, symbolized as H* have lower F0 than the low portion of a late rise-fall, marked as L*+H (as in the Saarbrücken sentence "Hasch e blaues Wohnmobil?").
- It is no longer necessary to explicate the non-arbitrary selection of certain H/L-sequences for the formation of new composite building elements beside H and L, and the exclusion of all other possible combinations - a justification which is, however, never given, neither in ToBI nor in ToBIG.
- Peak and valley contours are internally highly constrained, compared with the sequencings of ToBI(G) tones, which would need elaborate phonotactic restriction rules, but usually do not have them provided; of course PROLAB also requires concatenation rules for the ordering of peaks and valleys, but being located at a higher structural level they are less constrained.
- This also implies that pitch information is grouped together into unitary elements instead of being distributed across different types of tones (pitch accents and boundary tones).
- The linearity of PROLAB pitch patterns stops at the sequencing of holistic peaks and valleys between phrasing markers, everything else is handled by superposition:

- sandhi between peak and valley units, distributed over broad domains, e.g. hat patterns vs. dipped contours
- phrasing markers influence several prosodic features (timing, pitch, segmental exponency) on either side
- downstep characterizes whole phrasing units unless otherwise indicated
- speech rate and global register regulate timing and and F0 level as long-term overlays.

The Stuttgart system [9], based on [11], comes very close to PROLAB by treating sequences of H and L as unitary patterns, such as 'fall', 'rise', 'fall-rise', 'early peak' (taken over from KIM), rather than sequences of pitch/phrase accents and boundary tones. But distinct from PROLAB it still preserves the tonal composition, more for historical reasons than for a system-internal need.

2.3. The marking of sentence stress, intonation and phrasing

PROLAB also differs from ToBI(G) in that it marks sentence stress separately from intonation. So it becomes possible

- to recognise different sentence-stress levels to which the same types of peak or valley patterns can be attached (e.g. partially deaccented prenuclear medial peaks before a reinforced late peak in the Saarbrücken sentence quoted in 2.),
- to take other acoustic properties than F0 into account in the determination of sentence stress.

[8] also handles sentence accent separately from intonation and break indices by using the labels PA (primary accent of the intonational phrase), NA (secondary accents in the intonational phrase), EK (emphasis/contrastive accent). But they are located on the functional tier as perception of prominence, marking sentence modality, instead of being formal phonological categories. They still cannot differentiate non-nuclear (=secondary) accent from partial and complete deaccentuation in "er hat einen Brief geschrieben." So this part of ToBIG would benefit from a revision.

Similarly, PROLAB allocates phrasing markers with reference to all prosodic features on perceptual grounds and thus avoids their definition with exclusive or predominant reference to F0 contours. So it also labels intonation and phrasing independently of each other, whereas in ToBI(G) the allocation of break indices and tones is circular. AS PROLAB is usually a supplement to segmental labelling, ordinary word boundaries are always indicated, and cliticization and various other reductions in the joining of words are also marked, so that the addition of two phrasing markers PG1,2 covers all the boundary categories necessary in prosodic labelling. If prosodic labels are inserted into orthographic text, word boundaries are again provided, but sandhi adjustments are usually not, unless the transliterations symbolize them specially. In case they do not it might therefore be helpful to introduce a special conjunction marker.

ToBIG in [3] does not deal with break indices, but [8] does. Here a consensus will still have to be reached.

3. Categories of prosodic models for German

Before prosodic labels can be successfully applied to speech files a prosodic model has to be developed for the particular language. An appropriate symbol inventory is then a mere tool for the unique representation of the categories of this model, i.e. different, contrastive categories of the model must always be symbolized differently, and the same categories must receive the same notation.

As far as ToBI [1] is concerned the underlying model is the one developed for (American) English by Pierrehumbert [10]. In the case of ToBIG things are far less clear. The authors of [8] did not base their proposed prosodic symbolizations on a comprehensive model of German prosody - either worked out by themselves or taken over from others who had - but simply adopted the notational device provided by ToBI and applied it to German, without asking whether this framework is a necessary and sufficient representation for the model categories to be recognized in German.

The Stuttgart system [9] is different: it follows Fery [11], which constitutes at least a partial model of German intonation, albeit limited in scope and functionally oriented towards focus and grammatical phrasing (in read speech), rather than constituting a formal phonological model, whose categories are related to grammatical, semantic and pragmatic functions a posteriori.

The Saarbrücken system [3] is not explicit in this respect because it does not cite [11], but refers to [9], and of course Martine Grice dealt with questions of intonation theory in her PhD thesis [12], however, it should be remembered, in relation to Italian. So the question of the Saarbrücken system being built on a model of German prosody remains an open issue.

PROLAB presupposes KIM, the Kiel Intonation Model for German, and is very explicit about the categories that require symbolization in this formal phonological model. It must always be remembered that KIM is a phonological interface between the levels of meaning (syntax, semantics, pragmatics) and the level of speech signal manifestation. Contrary to [11], the prosodic categories of KIM are not defined as morphemes with independent meanings of their own.

This does, of course, not mean that the form - meaning relationships are not analysed. On the contrary, it is a subsequent step after having established the formal categories of the prosodic phonology in their own right: they are formal elements distinguishing meanings in the first instance, not the meanings themselves. Such terms as "question intonation" should therefore be avoided because they lead to a great deal of confusion.

Nevertheless it cannot be ruled out that in certain restricted cases a direct link between pitch pattern and meaning may be found, e.g. in the case of the early peak in languages like German and English. This is then on a similar level to sound symbolism, as when high vowels symbolize small, low vowels large things. This also raises the question of prosodic universals, but before they can be adequately tackled we need independent formal models for the individual languages.

There is, on the other hand, no denying the fact that English and German are very close to each other, also in terms of their prosodic phonologies, but they are not identical because, after all, they sound quite different. In order to achieve descriptive adequacy for the prosodic system of a language it should therefore always be treated in its own right first, correspondences can then be established by comparison afterwards, and prosodic labelling

only comes in right at the end. This argument, of course, raises the question as to whether the ultimate goal we are to aim for at this workshop - "cross-language generalisations of transcription" - is not hopelessly premature. There is no doubt that a number of prosodic labelling advocates actually put the labelling cart before the prosody horse.

How important it is to have a clear conception of the prosodic categories before labels are put to them can be seen

- by the ToBI examples in 2.1
- and by the following ToBIG labels for g071a004:
 - "Dienstag" has a late peak, marked as H* \lt
 - similarly "nachmittags" has a late peak, but it is now labelled as L+H*
 - "Donnerstag" has a medial peak but is also marked as L+H*
 - the first occurrence of "schwierig" has an early peak, the second one a medial peak, but both are marked as H*.

These examples show that the principle of unique representation of categories in the prosodic phonology of German is violated in both directions: the same categories are labelled differently, and different categories are also labelled the same. PROLAB, at least at these points, provides consistent labellings.

4. Prosodic notation and data bases

4.1. Notational conventions

All German ToBI variants, including ToBIG, do prosodic labelling on different tiers in orthographic text. Usually the traditional tiers are used: orthographic, break index, tone, miscellaneous, but [8] distinguishes "phrase boundaries", "intonation" and "functional tier", which includes a strange mixture of perceptual features (prominence => sentence accent) and functional aspects (sentence modality and 4 different types of question). [9] includes voice quality labels on the miscellaneous tier, which will be of great importance for the information processing of spontaneous discourse. In all versions, the alignment of labels with the speech wave follows the English ToBI conventions; in particular, the orthographic word is put at the end of the corresponding signal portion, pitch accents are placed somewhere within (the vowel of) the accented syllable.

PROLAB is structured in such a way that prosodic labels can be integrated

- either into a complete segmental label file
- or into an orthographic file.

There are thus no separate labelling tiers and files for different speech phenomena, and consequently the system of prosodic symbols is constrained by the demands of 7 bit ASCII notation in segmental labelling and orthographic transliteration. To achieve a simple division between prosodic and segmental labels, the former are prefixed by "&". The integration of segmental and prosodic labelling aims at the simultaneous availability of both information sources on the same level of representation in order to be able to study the interplay of segmental and prosodic elements more efficiently.

In segmental files word-initial segments are prefixed by "##", marking word boundaries at the beginning of the signal portion for the particular word; similarly in orthographic files the orthographic word is put at its signal onset. All prosodic labels are allocated to points in time that are either segment – usually word – beginnings, or the ends of segments in file-final position; they are not given durations. If several coincide on the same time point, they occur in the serial order

- (1) intonation
- (2) phrasing
- (3) speech rate change
- (4) sentence stress
- (5) synchronization of pitch contour with stressed syllable nucleus (early, medial, late).

Categories (1) and (2) refer principally to left-hand domains, the others to right-hand ones. Sentence-stress labels mark whole words as stressed; where in the word the stress properties manifest themselves is deducible from the position of the separately marked lexical stress vowel. Further categories included in the labelling are reset, upstep, high prehead, disfluencies. Categories still to be added are global register shift up or down from default and global voice quality changes.

4.2. Labelling platforms and labelling procedures

All German ToBI groups, with the exception of [8], use ESPS/waves+; [8] have developed their own portable environment, with tools for conversion from/to ESPS format. They can all select tiers individually for processing, but in [8] the intonation tier presupposes the marking of accents, and the recommended progression is phrasing, accents, intonation. For the allocation of pitch and phrase accents as well as boundary tones ToBIG labellers rely principally on the representation of the physical parameter of F0 and thus have to perform the

extremely difficult task of mapping individual phonetic signal manifestations onto given notational phonology-related categories. Since H and L furthermore evoke an association with high(er) and low(er) pitch values, respectively, it is not surprising that in the actual everyday labelling work the phonetic deviations from the phonological goal and the inconsistencies discussed in 2. should occur.

PROLAB has developed its own portable platform under X11. The system accepts segmental label files and F0 analysis data as input, allows the display of the speech wave form, the F0 contour and the segmental labels under visual as well as auditory control, the latter being the decisive factor for the allocation of labels. The manual labelling proceeds in cycles dealing with one prosodic domain after another, in the progression from phrasing to sentence stress, to intonation patterns (peaks, valleys), to their alignment, to speech rate changes and to the other variables mentioned in 4.1. In this progression the labelling can stop at any degree of delicacy, defined by the purpose the resulting label files are to be put to. PROLAB is a comprehensive labelling frame for all prosodic variables, but it is at the same time very flexible to allow a wide range of detail and complexity for different applications.

4.3. Support tools

[8] is developing an acoustic verification procedure of prosodic labelling using PSOLA-resynthesis for systematic modifications of prosodically labelled natural speech productions.

PROLAB has developed two verification procedures:

- LPC resynthesis controlled by intonation category labels to test the functional equivalence of the natural speech signal and the resynthesized variations
- a complete TTS synthesis-by-rule system for German [6a,b,c], controlling all the prosodic categories that are labelled in PROLAB and generating speech output from symbolic PROLAB input (orthographic strings supplemented by PROLAB labels); it has also been adapted to the synthesis of spontaneous speech, producing very convincing, natural sounding results that can be evaluated against the original natural production.

PROLAB also provides a formal check program for the correct syntax of PROLAB labels in label files. Something comparable does apparently not exist for ToBIG.

4.4. Data bases

Within the German ToBI field a number of machine readable data bases exist, but with widely differing prosodic scopes and aims. In preparation of this workshop a small data base has been labelled in the unified ToBIG system on the orthographic and intonation tiers. This could be the start for building up a more homogeneous labelled data base in the ToBI environment.

In addition to the data prepared for this workshop, 7 VERBMOBIL dialogues, containing 115 turns, have been labelled using the full range of PROLAB categories. These prosodic label files are supplements to the signal and segmental label files (g071-7a) on the Kiel CD2 and are available in 7 bit ASCII on diskettes to anyone who has acquired the CD. Through the application of filters the comprehensive label files can be reduced to any degree of detail desired. If the segmental information is not required or even regarded as distracting, it can be removed – except for word boundaries and the lexically stressed vowels in sentence-stressed words, which are needed for attaching the stress and intonation categories; and since the label files always have an orthographic representation in the header, orthographic words can be

inserted at the respective boundary markers instead. Phrasing markers and sentence-stresses will always have to be kept, but any other prosodic labels can (progressively) be eliminated to suit any specific application.

5. Conclusions: Mapping of ToBIG and PROLAB

It has already been shown in 4. that the correspondences between ToBIG and PROLAB are neither biunique nor unique in one direction. The following tables list the PROLAB/ToBIG and ToBIG/PROLAB equivalents and their frequencies of occurrence in the data set that was selected for transcription in preparation of this workshop. As can easily be seen a mapping is impossible. For example, out of 47 H* pitch accents only 30 (64%) correspond to any type of 'medial peak', 7 (15%) to 'early peak', 5 (10.5%) to 'late peak' and 5 (10.5) to 'valley'; on the other hand out of 21 L+H* as many as 13 (62%) also correspond to 'medial peak', 3 (14%) to 'late peak', 4 (19%) to 'valley' and 1 (5%) to 'early peak'. This means that both systems will have to continue to be used side by side; they cannot be conflated.

Table 1:
PROLAB/ToBIG equivalents and their frequencies

medial peak:	1.	H*	9		L+H* L-L%	1
narrow		H* <	1	early valley:) ,	H* H-	1
		H* L-	3	narrow	L+H* H-	3
		H* H-	4			
		H* !H-L%	1	late valley: (,	H*	1
		L+H*	4	narrow	?H*	1
		L+H* H-	1		H* H* H-	1
		L+H* L-L%	1		L* L-H%	1
		L+H* < !H-	1			
		L+H* !H-L%	1	early valley:) ?	H* H-H%	1
				wide	L-H%	1
medial peak:	2.	H* L-L%	4		L+H* H-H%	1
wide		H* !H-	1			
		L+H* L-L%	3	late valley: (?	L* H-H%	1
		L+!H*	1	wide		
medial peak:	0.	H*	3			
flat		!H*	1			
		H* H-	2			
		!H* H-	1			
		L+H* H-	1			
early peak:)	1.	H* L-	2			
narrow		H* H-	1			
		!H* L-L%	1			
		H+!H*	1			
		H+!H* L-	1			
early peak:)	2.	H* L-	1			
wide		H* L-L%	1			
		!H* L-L%	1			
		H+!H* L-L%	2			
		!H+!H* L-L%	1			
		H+L* L-L%	2			
		!H+L* L-L%	1			
		L+!H* L-L%	1			
		L* L-L%	1			
late peak: (1.	H*	1			
narrow		H* <	3			
		H* !H-L%	1			
		L*+H	1			
		L+H*	1			
		L+H* H-L%	1			
late peak: (2.	L*+H L-L%	2			
wide		L*+H L-	1			

Table 2:
ToBIG/PROLAB equivalents

H*	1.	H* H-H%)?
	0.	H* L-L%	2.
	(1.)2.
	(,	!H* L-L%)1.
H* <	1.)2.
	(1.	H* !H-L%	1.
!H*	0.		(1.
?H*	(,	L* H-H%	(?
		L* L-L%)2.
L+H*	1.	L* L-H%	(,
	(1.		
L+!H*	2.	L+H* L-L%	1.
H+!H*)1.		2.
L*+H	(1.		(2.
		L+H* H-L%	(1.
H* H-	1.	L+H* !H-L%	1.
	0.	L+H* H-H%)?
)1.	L+!H* L-L%)2.
),	H+!H* L-L%)2.
!H* H-	0.	!H+!H* L-L%)2.
H* !H-	1.	L*+H L-L%	(2.
H* L-	1.	H+L* L-L%)2.
)1.	!H+L* L-L%)2.
)2.		
H+!H* L-)1.		
L*+H L-	(2.		
L+H* H-	1.		
	0.		
),		

In a situation like this the question arises as to which of the two systems is more adequate, either generally or for particular purposes. I think it is fair to say that the descriptive adequacy of PROLAB is higher with regard to the symbolization of sentence stress and pitch contour alignment because the same categories are always represented with the same, different categories consistently with different labels, and categorical differentiations are symbolized, whereas contextual phonetic variation is ignored at the symbolic level, only to be retrieved at the signal level in the grouping of speech analysis data under the same categories. Furthermore, PROLAB is more economical in its marking of pitch contours: 11 PROLAB peak and valley categories correspond to 32 ToBIG pitch label sequences (pitch/phrase accents plus boundary tones). And this greater economy is coupled with a superior systematicity and clarity.

These issues will be the central topic for discussion at this workshop and no doubt in months to come. If we reach a greater clarification and hopefully consensus on these points, the meeting will have reached a respectable goal. The main thing is that we communicate with each other in a scientific and constructive way that puts the essential and central emphasis on questions of subject matter, i.e. prosodic theory, prosodic modelling, prosodic representation, and relegates administrative tools and procedures to the level of adjuncts.

References

- [1a] Beckman, M. E.; Ayers, G. M. (1994), Guidelines for ToBI Labelling, version 2.0.
- [1b] Hirschberg, J.; Beckman, M. E. (1993), The ToBI Annotation Conventions.
- [2] Grice, M.; Benz Müller, R. (1995), "Transcription of German intonation using ToBI-tones - the Saarbrücken system", *Phonus 1*, Institute of Phonetics, University of the Saarland.
- [3] Kohler, K. J. (1995), "PROLAB - the Kiel system of prosodic labelling", *Proc. XIIIth ICPHS*, Stockholm.
- [4] Kohler, K. J. (1992), "Prosodisches Transkriptionssystem für die Etikettierung von Sprachsignalen", *AIPUK*, vol. 26, pp. 238-252.
- [5a] Kohler, K.J. (1991), "A model of German intonation", *AIPUK*, vol. 25, pp. 295-360.
- [5b] Kohler, K.J. (1990), "Macro and micro F0 in the synthesis of intonation", in: *Papers in Laboratory Phonology I* (eds. J. Kingston, M. E. Beckman), Cambridge: CUP, pp. 115-138.
- [5c] Kohler, K.J. (1987), "The linguistic functions of F0 peaks", *Proc. XIth ICPHS*, Tallinn, Vol. 3, pp. 149-152.
- [6a] Kohler, K.J. (1988), "An intonation model for a German text-to-speech system", *Proc. Speech '88*, Edinburgh: Institute of Acoustics, pp. 1241-1247.
- [6b] Kohler, K.J. (1991), "Prosody in speech synthesis: the interplay between basic research and TTS application", *J Phon* 19, pp. 121-138.
- [6c] Kohler, K.J. (forthcoming): "Parametric control of prosodic variables by symbolic input in TTS synthesis", *Sec ESCA/IEEE Workshop on Speech Synthesis*, Sept 1994, New Paltz, N. Y.
- [7] Kohler, K.J., Pätzold, M., Simpson, A. (1994), *Handbuch zur Segmentierung und Etikettierung von Spontansprache - 2.3. VERBMOBIL Technisches Dokument Nr. 16*, Kiel: IPDS.
- [8] Reyelt, M., Batliner, A. (1994), *Ein Inventar prosodischer Etiketten für VERBMOBIL*, VERBMOBIL Memo 33, Version 1.0
- [9] Mayer, J. (1995), *Transcription of German intonation - the Stuttgart system*, University of Stuttgart.
- [10] Pierrehumbert, J. B. (1980), *The phonology and phonetics of English intonation*. MIT PhD Thesis, published 1988 by Indiana University Press.
- [11] Féry, C. (1993), *German Intonational Patterns*, *Linguistische Arbeiten* 285, Tübingen: Niemeyer.
- [12] Grice, M. L. (1992), *The Intonation of Interrogation in Palermo Italian; Implications for Intonation Theory*, PhD Thesis UC London.