

# MULTI-LEVEL ANNOTATION FOR SPOKEN LANGUAGE CORPORA

Philippe Blache & Daniel Hirst  
LPL-CNRS, Université de Provence  
email: {pb; [daniel.hirst@lpl.univ-aix.fr](mailto:daniel.hirst@lpl.univ-aix.fr)  
[www.lpl.univ-aix.fr](http://www.lpl.univ-aix.fr)

## ABSTRACT

The constitution of multi-level databases integrating, for example, both prosodic and morphosyntactic levels of representation presents a number of problems, some specific to the individual domains, and others concerning the integration of the two domains. It is argued that the formalism of annotation graphs provides an adequate solution to these problems, which can be implemented in an XML representation. It is further argued that a generic query language, DQL, currently being developed, will provide a satisfactory framework both for querying and for manipulating documents of this type.

## 1. BACKGROUND

The annotation of linguistic information, and, more generally, the constitution of electronic linguistic resources, has for a long time been restricted to the morphosyntactic level. The representation of this type of annotation presented a certain number of challenges, in particular that of the representation of ambiguous or incomplete information. Now that these problems are mastered, it is important to move forward to the next step where it will be possible to envisage the annotation of information coming from different levels of linguistic analysis and in particular phonetic, phonological, syntactic and semantic information. To this end it is necessary to adopt an adequate formalism for representation since it is not always possible to project the different types of information onto a single level. In particular the annotation of spontaneous spoken corpora incorporating both prosodic and syntactic information is likely to present two types of problem for classical annotation systems.

### 1.1 Prosodic representations

It is well-known that prosodic and morphosyntactic information are not always directly superimposable as can be

<p>C'est un œuf /sE . tø) . nøf/ 'it's an egg'</p>
--

shown from the following example from French (where "." in the phonetic transcription indicates a syllable boundary):

In this example, the phonemes of four lexical items 'ce' /s' / 'est' /Eɛt/ 'un' /ø) n/ 'œuf' /øɛf/, are mapped onto three syllables /sE/ /tø) / /nøf/, none of which correspond to a single word. It is not possible to propose a multi-level representation in terms of a strict hierarchy or immediate constituents including both the morphosyntactic and the phonological structure, although this is of course possible for each of these levels considered separately. The problem can become even more acute when more complex phonological structures are considered as in figure 1 (from [8]) which combines tonal and segmental information in a prosodic structure which, once again, is not isomorphic with morphosyntactic structure and cannot consequently be represented with most classical annotation systems.

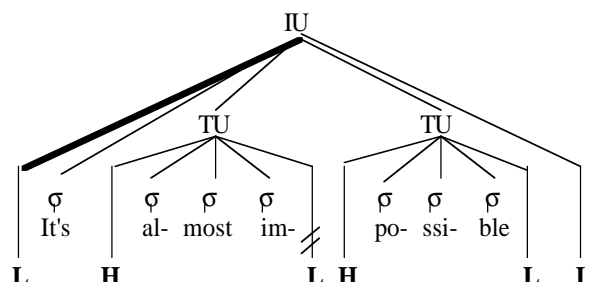
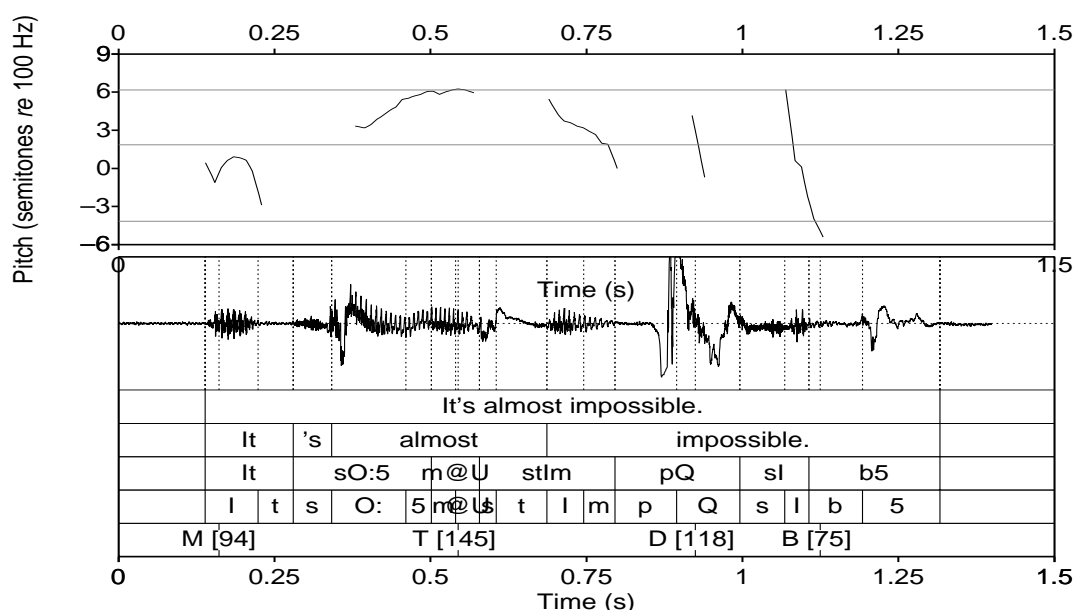


Figure 1. One model of the phonological structure of the English sentence "It's almost impossible" combining tonal and segmental specifications.

Figure 2 illustrates a TextGrid label file of a recording of the same sentence, displayed with the *Praat* software [7], and incorporating tonal information output from the *Momel* and *INTSINT* algorithms ([9], [10] and [11]) which first convert a raw  $f_0$  curve to a sequence of target points and then provide an optimised symbolic coding of these target points. Phonematic segments are here represented using the machine-readable SAMPA phonetic alphabet [15].

The possibility of automatic extraction of values of fundamental frequency targets using algorithms like those mentioned above, together with values of segmental duration using automatic alignment techniques such as [14], means that it is now possible to obtain large quantities of acoustic data for



**Figure 2.**  $f_0$  display (top) and TextGrid label file with sentence, word, syllable, phoneme and tonal (INTSINT) labels (bottom) from a recording of the example illustrated in Figure 1.

a number of languages, a result which will undoubtedly prove beneficial for testing prosodic modelling on a multilingual basis. In order to evaluate the adequacy of different models of prosodic structure for predicting this sort of fine phonetic detail, it is crucial to be able to combine phonological representations such as that of figure 1 with acoustic data such as that in figure 2 into a uniform system of annotation.

## 1.2 Syntax of spoken corpora

The syntactic annotation of spoken corpora also poses a number of specific problems. It has been shown for one thing [5] that it is often impossible to dissociate different levels of linguistic analysis, syntactic descriptions often needing for example to refer to intonation as distinguishing criteria. The simultaneous representation of a number of different levels of annotation for a given corpus would consequently be extremely useful.

There are, furthermore, a number of specific constructions and devices in this type of corpora which cannot be naturally represented with a tree-structure. Among these are several phenomena associated with dysfluencies and repairs which, it has been proposed, are best analysed by a paradigmatic rather than a syntagmatic representation (cf. for example [5], [6]). Thus in the following

- (1). dès l'arrivée sur cette frontière {qui est blafarde | qui est {sinistre | véritablement sinistre}} comme toutes les frontières  
(as soon as we arrived on the border which is pale, which is sinister, really sinister, like all borders)

The successive elements (in curly brackets separated by the symbol '|') can be considered paradigmatic variants, the last of which is taken, in this case, as the intended message. Cases like this need to be distinguished from other superficially similar examples such as

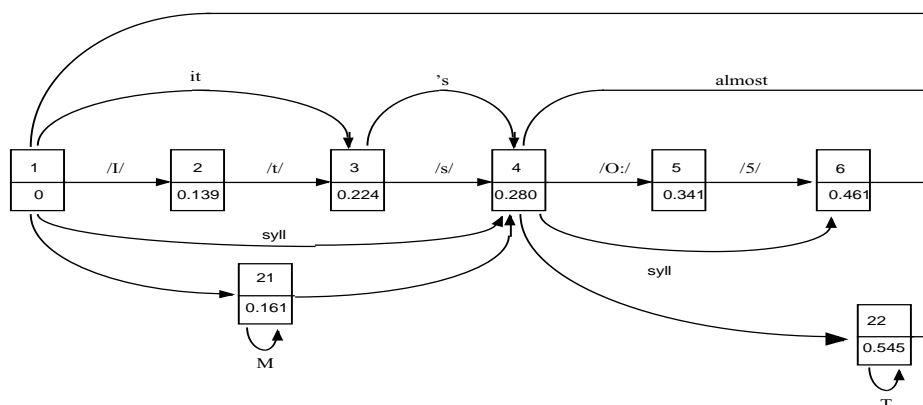
- (2). on réduit, on réduit, il arrive un moment où on ne plus réduire  
(we reduce, we reduce, there comes a moment when we can't reduce any more)

where the repeated elements are part of an expressive device, each element reinforcing the previous one rather than replacing it as in example (1).

## 2. ANNOTATION GRAPHS

The formalism of Annotation Graphs proposed by [1], [2] provides a satisfactory solution to the difficulties of annotation described in the preceding section. The same input can be annotated by different subsets of arcs corresponding to different levels of annotation (prosodic, syntactic etc.). A specific level of linguistic representation thus corresponds to a subset of the general graph.

Since the representation is a graph and not a tree there is nothing to stop association lines from crossing, making it possible to represent levels which are not directly superimposable. The only constraint for annotation is that events be describable in terms of a set of discrete linearly ordered (or at least partially ordered) moments which constitute the nodes of the graph. These moments in turn can be indexed by an offset reference to a basic timeline



**Figure 3.** Annotation graph combining information from the phonological structure in figure 1 and acoustic information from the TextGrid label file in Figure 2.

associated, for example, with a physical object such as a speech signal, or to some more abstract specification of linear order.

Figure 3 shows an example of part of an annotation graph combining information from the phonological representation in figure 1 and acoustic data from the TextGrid label file illustrated in Figure 2.

The acoustic signal provides a common reference for the alignment of the tonal segments M, T etc with respect to the phonematic segments or with respect to the more abstract levels of structure (syllables, words, tonal units, intonation units or whatever other higher level prosodic units might be used in the annotation).

Tonal targets, which unlike the other prosodic and syntactic categories constitute temporal points rather than intervals, are represented in figure 3 by an arc with identical start and end nodes. This makes it possible to code this type of information while maintaining the general strategy of encoding content on the arcs rather than on the nodes of the graph. There are of course other ways of encoding this information (cf. [1] for discussion).

Another advantage of this type of annotation is the possibility of specifying information concerning just a part of the input data without necessarily building a complete structural analysis. In the case of syntax, this means it is possible to associate an arc with a set of properties characterising the corresponding part of the data. This type of syntactic annotation is particularly useful in the case of non-derivational formalisms such as that of Property Grammars which have the specificity of providing partial analyses.

### 3. XML CODING OF ANNOTATION GRAPHS

The XML language specifically excludes the possibility of cross embedding. Thus a representation such as:

```
<word><syll><phone>s</phone></word>
<phone>E</phone></syll>...
```

would be refused by any XML parser as ill formed. Despite the fact, annotation graphs can be represented with XML. One solution is to code the nodes and the arcs which constitute the annotation graph as independent empty XML elements.

A node would then consist of an empty three-argument XML element:

```
<node id="n1" time="0.000"/>
<node id="n2" time="0.139"/>
<node id="n3" time="0.224"/>
...
<node id="n21" time="0.161"/>
<node id="n22" time="0.545"/>
```

while an arc could consist of an empty six-argument element:

```
<arc id="a1" begin_node="n1" end_node="n2"
category="phone" content="l"/>
<arc id="a2" begin_node="n2" end_node="n3"
category="phone" content="t"/>
...
<arc id="a31" begin_node="n21" end_node="n21"
category="tone" content="M"/>
<arc id="a32" begin_node="n22" end_node="n22"
category="tone" content="T"/>
```

More general solutions have recently been proposed within the framework of the ATLAS architecture [2].

## 4. DQL: A GENERIC QUERY LANGUAGE

In a large number of projects concerned with high level annotation, specific tailor-made query languages have been developed. We suggest that a more interesting direction might be the use of a generic query language which allows not only data retrieval from labelled documents but also the direct manipulation of the documents themselves thus providing a complete tool for annotation.

One language of this type: *DQL* (Document Query Language), which seems particularly appropriate for the tasks we envisage, is currently being developed as a successor to *SgmlQL* [12], [13], based on *OQL* (the object oriented version of *SQL*) and which allows the manipulation of structured documents (*SGML*, *HTML*, *XML*...). *DQL* is an evolution integrating the *Xpath* language, allowing access to the components of an *XML* documents.

*DQL* allows an implementation of all the standard queries addressed to a structured document (cf. [3]) but also provides essential document manipulation operations such as bracketing (substitution of a set of sub trees of the same type by a single tree) and its reverse (replacement of a single tree by a set of sub trees), as well as compression (set of leaves), extraction of sub trees or suppression of branches.

*DQL* will be adapted to the specific task of manipulating annotation graphs. This presents specific problems concerning the cases where non-tree structure involve crossing association lines, as for example in the case of discontinuous constituents or paradigmatic phenomena illustrated in §1.1 above. It is anticipated, however, that the generic nature of *DQL* will make it easily adaptable to follow the evolution of the annotation formalism since it can be used with any form of structured representation. This, added to the above-mentioned functions presents a major advantage.

### Acknowledgement

Part of the research described here benefited from support by the European COST action n°. 258 *Improving the quality of speech synthesis*.

## REFERENCES

1. Bird S. & M. Liberman 1999. A formal framework for linguistic annotation. Technical Report MS-CIS-99-01. Dept of Computer and Information Science, University of Pennsylvania. to appear in *Speech Communication*.
2. Bird S., D. Day, J. Garofolo, J. Henderson, C. Laprun, & M. Liberman 2000. ATLAS: A flexible and extensible architecture for linguistic annotation. in proceedings of the *Second International Conference on Language Resources and Evaluation*..
3. Bird S., P. Buneman & W. Tan 2000. Towards a query language for annotation graphs. in proceedings of the *Second International Conference on Language Resources and Evaluation*..
4. Blache, P. (2000). Property grammars and the problem of constraint satisfaction. in *Proceedings of the ESSLLI-2000 Workshop on Linguistic Theory and Grammar Implementation*..
5. Blanche-Benveniste C., J. Deulofeu, J. Stefanini, K. van den Eynde 1984. *Pronom et syntaxe – L'approche pronominale et son application au français*, Selaf
6. Blanche-Benveniste C. 1997. *Approches de la langue parlée en français*, Ophrys
7. Boersma, P. , Weenink, D. 2000. *Praat: a system for doing phonetics by computer*.  
<http://www.fon.hum.uva.nl/praat/>
8. Hirst, D.J. 1998. Intonation in British English. in D.J.Hirst & A. Di Cristo (eds) *Intonation Systems. A Survey of Twenty Languages*. Cambridge; Cambridge University Press.56-67.  
<http://www.lpl.univ-aix.fr/~hirst/books.htm>
9. Hirst, D.J. 2000. ProZed. A multilingual prosody editor for speech synthesis. in *Proceedings IEE Colloquium on State-of-the-Art in Speech Synthesis*. London, April 2000.
10. Hirst, D.J. 2000. Optimising the INTSINT coding of F0 targets for multi-lingual speech synthesis. in *Proceedings ISCA Workshop: Prosody 2000 Kraków*, October 2000.
11. Hirst, D.J. , Di Cristo, A., Espesser, R. in press. Levels of representation and levels of analysis for the description of intonation systems. In Horne, M. (ed.) in press. *Prosody: Theory and Experiment*, Dordrecht; Kluwer Academic Publishers.
12. Le Maitre J., E. Murisasco, M. Rolbert 1997. From Annotated Corpora to Databases: the SgmlQL Language, in *Linguistic Databases, CSLI Lecture Notes #77*.
13. Le Maitre J., E. Murisasco 1999. Controlled Hypertextual Navigation in the SgmlQL Language. in *Proceedings of DEXA'99, LNCS #1677*, Springer.
14. Malfrère, F. & T. Dutoit 1997. High Quality Speech Synthesis for Phonetic Speech Segmentation, *Proceedings Eurospeech. '97*. 2631-2634.
15. Wells, J.C., Barry, W., Grice, M., Fourcin, A., Gibbon, D. .1992. Standard computer-compatible transcription. *Esprit project 2589 (SAM), Doc. no. SAM-UCL-037*. London, Phonetics and Linguistics Department, UCL