Visualising Lexical Prosodic Representations for Speech Applications

Julie Carson–Berndsen
Department of Computer Science
University College Dublin
Belfield, Dublin 4
Ireland
Julie.Berndsen@ucd.ie

Dafydd Gibbon
Fakultät für Linguistik & Literaturwissenschaft
Universität Bielefeld
Postfach 100 131, 33501 Bielefeld
Germany
gibbon@spectrum.uni-bielefeld.de

Abstract

This paper presents a technique for generic lexical description and feature geometry based graphical visualisations of phonological and prosodic properties for speech technology applications combining work on prosodic inheritance with a proposal to use nonsegmental phonological representation in lexical description in order to compile lexica for specific speech applications from a more general lexical representation. Prosodic inheritance is a theory of generalisations about prosody expressed as a connected system of implications conforming to the structure of an inheritance graph, rather than as individual implication rules or constraint sets. In line with developments in computational phonology, the computational core of the generic lexicon is based on finite state transducers and graphical visualisations are generated for lexical entries as are text representations of the feature geometry descriptions in terms of bracketed structures.

1 Introduction

This paper presents a technique for generic lexical representation of phonological, in particular prosodic properties in the lexicon (cf. [Gibbon, (1998)]), which can have applications in various areas of speech technology. The relevance of this work to language and vision lies in the visualisation of feature geometry and multilinear representations of prosodic phonologies such as autosegmental phonology. Multilinear phonological representations were originally proposed in the theory of autosegmental phonology by Goldsmith [Goldsmith, (1976)] in order to cater for tonal phenomena which could not be integrated into a purely segmental system. Multilinear representations consist of separate tiers (e.g. tonal tier, nasality tier) each of which has its own synchronisation function and melody (similar to the musical score of an orchestra). Furthermore, feature geometry introduces the dimension of the internal organisation of sounds into a multilinear representation.

The approach presented in this paper combines work on prosodic inheritance (cf. [Reinhard and Gibbon, (1991)] [Gibbon and Ahoua, (1991)]) with a recent proposal to use nonsegmental phonological representation in lexical description in order to compile lexica for specific speech applications from a more general lexical representation (cf. [Cahill et al., in press]) using the notion of *time maps* as defined in [Carson-Berndsen, (1998)]. The generic lexicon is modelled in the DATR lexical representation language and generates a range of representations, including graphical objects.[1] The proposal in this paper differs from previous work in that it uses feature geometry as a basis for the representation rather than autosegmental inheritance ([Gibbon and Ahoua, (1991)]) or the multivalued feature vectors proposed in [Carson-Berndsen, (1998)]. The advantage of this over feature vectors is that the generic lexicon can then be used not only for those applications suggested in [Cahill et al., in press] but also in connection with a testbed for phonological descriptions in more theoretical work and in connection with web-based tutorials for students of phonology and phonetics. This paper will present the generic lexical description and the graphical visualisations.

2 Prosodic Inheritance: a DATR model

Prosodic inheritance is a theory of generalisations about prosody expressed as a connected system of implications conforming to the structure of an inheritance graph, rather than as individual implication rules or constraint sets. In line with developments in computational phonology, the computational core of the generic lexicon is based on finite state transducers.

The representation language which has been chosen for the generic lexicon is DATR (cf. [Evans and Gazdar (1989)] and more recently the discussion in [Evans and Gazdar (1996)]). DATR is a language which has been specifically designed for the representation of lexical entries and generalisation hierarchies over these entries. The language permits definition of nonmonotonic inheritance hierarchies using path/value equations, allowing lexical regularities, subregularities and idiosyncrasies to be captured in a principled manner. DATR is not restricted to nonmonotonic inheritance, and subsumption–style hierarchies can also be represented if required. Numerical, string–processing and file–handling extensions to DATR have been added in the *Zdatr* implementation used in the present study ([Gibbon and Strokin, (1998)]); with these extensions quantitative information, such as duration, frequency and weightings can be included in the lexicon and numerical functions of these can be defined. In [Gibbon and Ahoua, (1991)] phonological features are explicitly used in a prosodic framework where default inheritance is used to model both phonological markedness and the extraction of autosegmental tiers. In [Cahill et al., in press] *time maps* from autosegmental structures to phonetic coordinates are defined, allowing symbolic and numerical representations of tier melodies to be calculated from a general lexicon representation using finite state transducers. This work has been extended to include enhancements permitting numerical and symbolic representations which are suitable for use in speech systems (cf. [Carson-Berndsen, (1999a)]; [Carson-Berndsen, (1999b)]), and graphic visualisation techniques. This paper concentrates on the graphic visualisation techniques.

The work presented in this paper has been implemented and tested using the *Zdatr* implementation described in [Gibbon and Strokin, (1998)]. The graph visualisation system `daVinci`, an X-Window visualisation tool for drawing directed graphs which is being developed at the University of Bremen, is used for the visualisation of the feature geometry structures in a more user-friendly format.

Queries to the *Zdatr* lexicon control the extraction of the following kinds of information:

- standard segmental phonemic representations with/without temporal annotations;
- standard segmental feature bundle representations with/without temporal annotations;
- tier melodies with precedence constraints (‘relative time map’);
- tier melodies with temporal annotations (‘absolute time map’);
- inter–tier overlap constraints;
- syllable structure;

---

[1] The language described in this application is German. Phonemic representations are in SAMPA and typewriter font for machine readable examples and in conventional IPA roman font in the text.

- `daVinci` objects in tree–structured feature geometry graphs, with a post–processor for defining re–entrancies.

3 An inheritance hierarchy for feature geometry

The feature geometry description used here is based primarily on Clements and Hume [Clements and Hume, (1995)] but also uses some conventions from Pullyblank [Pullyblank (1995)]. A feature geometry representation is a graph with the following clearly distinguished components:

1. Rooted trees which are isomorphic up to depth 3 for each C or V segment in a sequence, labelled with articulatory features, and representing constraints on the articulation of speech sounds;
2. Operations over nodes in successive trees, for example generating re–entrancies between segments in order to represent assimilations and other phonological relations.

The following shows general definitional statements for consonants formulated in DATR, together with specific definitions for the consonants /t/, /d/, /s/, /f/ and /v/.

```
C:
  <>             == Null
  <rootfeats>    == "<sonorant>" "<approximant>" "<vocoid>"
  <sonorant>     == [-sonorant]
  <approximant>  == [-approximant]
  <vocoid>       == [-vocoid]
  <laryngeal>    == [spreadglottis]
  <oral_cavity>  == [-continuant]
  <nasal>        == [-nasal]
  <atr>          == [-atr]
  <place>        == [coronal] "<[coronal]>"
  <[coronal]>    == ’(’ [+anterior] ’)’ ’(’ [-distributed] ’)’
  <featural>     == ’(’ root "<rootfeats>"
                    ’(’ laryngeal ’(’ "<laryngeal>" ’) )’
                    ’(’ "<nasal>" ’)’
                    ’(’ oral cavity ’(’ "<oral_cavity>" ’)’
                       ’(’ c_place ’(’ "<place>" ’) )’ ’)’ ’)’
  <davinci>      == Davinci_Node:<name_root lightgreen>.


C_t: <>            == C
     <segmental>   == t.
C_d: <>            == C
     <laryngeal>   == [voice]
     <segmental>   == d.
C_s: <>            == C
     <oral_cavity> == [+continuant]
     <segmental>   == s.
C_f: <>            == C_s
     <place>       == [labial]
     <segmental>   == f.
C_v: <>            == C_f
     <laryngeal>   == [voice]
     <segmental>   == v.
```

An individual consonantal segment such as /d/ defined above inherits everything which is not specified at `C_d` from the more general `C` node. Vowels are defined similarly with each specific vowel inheriting from a more general `V` node.

The `<featural>` and `<davinci>` paths in these general definitions refer to the feature geometry. The `<featural>` path refers to the text representation of the feature geometry description in terms of bracketed structures. From these node definitions, we can now infer the following feature representations for the segment /t/, for example:

```
C_t:<featural> =
 ( root [-sonorant] [-approximant] [-vocoid]
   ( laryngeal ( [spreadglottis] ) )
   ( [-nasal] )
   ( oral cavity ( [-continuant] )
                 ( c_place ( [coronal] ( [+anterior] ) ( [-distributed] ))))) .
```

The `<davinci>` path inherits the graph structure from `Davinci_Node` in a format which can be read by the `daVinci` interpreter. The `daVinci` format defines nodes and edges which can be specified with respect to name, color and pattern attributes. In the `<davinci>` path of the `C` node definition, the root level nodes are defined to be lightgreen. However, since the details of the `daVinci` representation are not relevant to the discussion which follows, let it suffice to say that in addition to the generation of the text format, other information must be inherited which is used to define nodes. For example, `daVinci` does not allow two nodes to have the same name; therefore, it is not sufficient to say that a segment /t/ has a laryngeal node and that another segment /m/ also has a laryngeal node but we must define these in the `daVinci` notation to be laryngeal_t and laryngeal_m. These extensions are inherited using global inheritance via the `<segmental>` path of the individual segments. The part of the `Davinci_Node` node definition which defines the graph from the root down is therefore as follows:

```
Davinci_Node:
  <>             == Null
  <name_root $C> == Davinci_Edge:<dashed>
                   'l("root_' "<segmental>" '" ,
                     n("Node",[a("OBJECT","' root "<rootfeats>" '"),
                               a("COLOR","' $C '")], ['
                    Davinci_Edge
                     <name_laryngeal lightyellow> ']))),'
                    Davinci_Edge
                     <nasal lightyellow> ','
                    Davinci_Edge
                     <name_oral_cavity lightyellow> '])))]))'.
```

This node also contains paths for `<name_laryngeal lightyellow>`, `<nasal lightyellow>` and `<name_oral_cavity lightyellow>`. `$C` is a DATR variable representing any colour. The syllable entries are defined, in line with those given in [Cahill et al., in press], as follows:

```
S_mIt:  <>                 == Syllable
        <phn onset first>  == "C_m:<>"
        <phn peak first>   == "V_I:<>"
        <phn coda first>   == "C_t:<>".
```

This node definition for the syllable /mɪt/ states that the first position of the syllable onset inherits specific properties from the consonant definition for /m/, the first position in the syllable peak inherits from the vowel definition for /ɪ/ and the first position in the syllable coda inherits from the consonant definition for /t/. All other information is inherited from the `Syllable` node.

From the node definitions for the syllable entries, taken together with the axioms for syllable structure, we can now infer the following `daVinci` graph representation of the feature geometry representations for the syllable /mɪt/ (only a subsection of the complete graph is given in the figure):

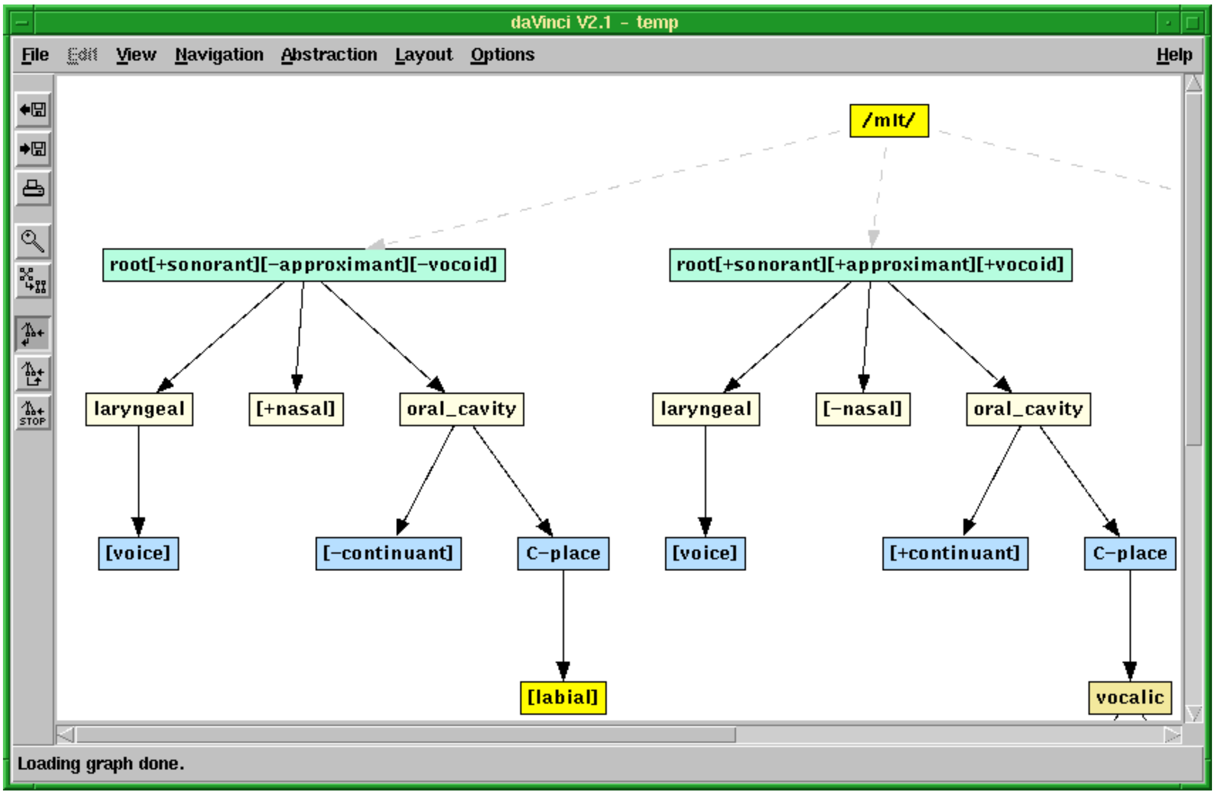


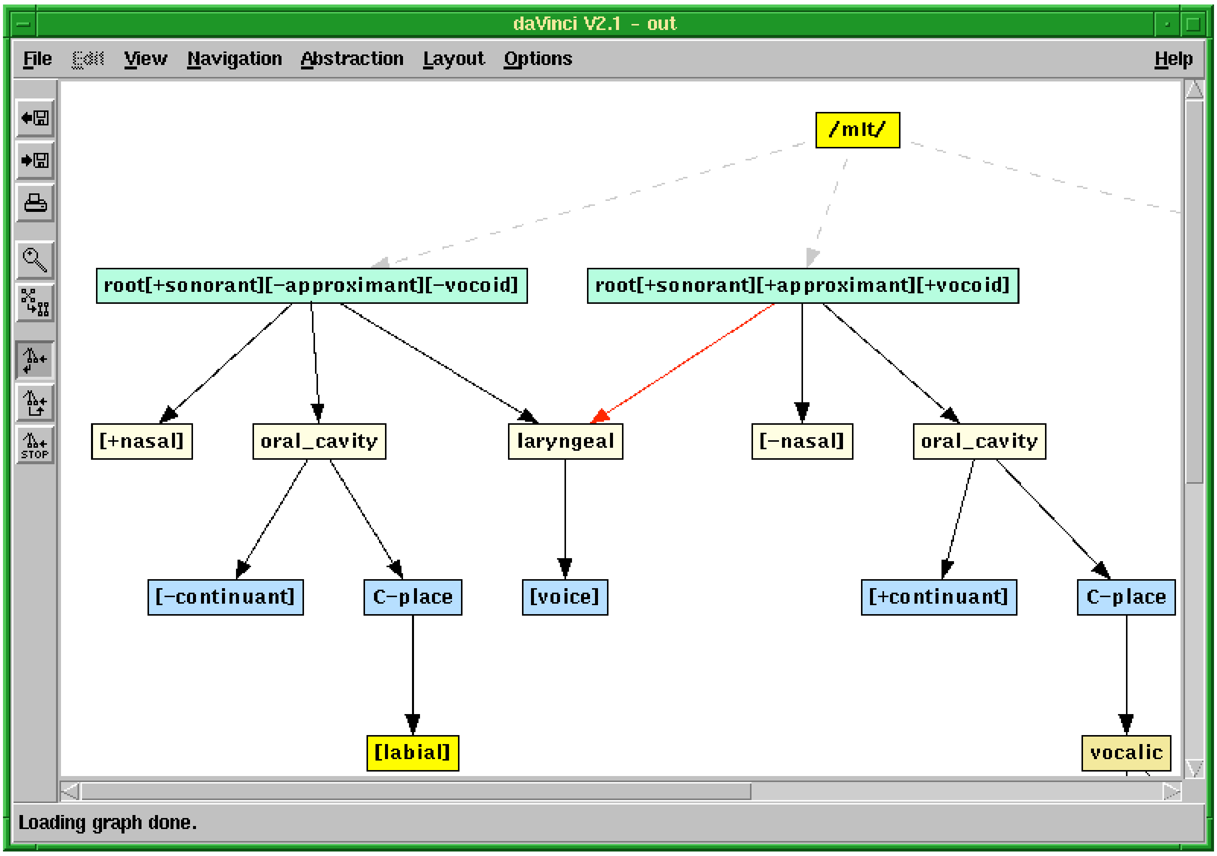
Figure 1: A subsection of the DaVinci representation for the syllable /mɪt/

As can be seen from Figure 1, this is a purely segmental representation of the feature geometry of the syllable /mɪt/ in the sense that the synchronisation function is the same across tiers. In order to move to a nonsegmental description, information on tier melodies is required.

In the DATR representation we have information about the individual tiers and their melodies. We can, therefore, use this information to generate a new graph description for the tier and feature values which can be substituted in the graph representation format. Individual tier melodies are represented in DATR using finite state transducers. These provide a *smoothing function* modelling Leben's *Obligatory Contour Principle (OCP)* over feature value trajectories on specific tiers. As demonstrated in [Cahill et al., in press], such finite state transducers can be represented elegantly by variables in DATR.

The approach taken in this paper is similar in that variables are also used. For example, in the syllable /mɪt/, the melody on the `laryngeal` tier consists of the feature value [voice] preceding the feature value [spreadglottis]. For the graph visualisation, however, information about the melody alone is not sufficient and information about the feature geometry of the individual segments is also required.

We therefore want to construct a representation like the following for the laryngeal tier: `[voice] m ɪ/ [spreadglottis] t`. This determines that the arc from the root of the segment /ɪ/ to the node `laryngeal` is the one which must be manipulated. The actual substitution in the graph representation format is performed on the output from the *Zdatr* component currently using a postprocessor UNIX script with regular expression substitutions; this postprocessor will be replaced by *Zdatr* operations. The graph representation in Figure 2 is generated by specifying that smoothing should be applied on the `laryngeal` tier.



Loading graph done.

Figure 2: A subsection of the DaVinci representation for the syllable /mɪt/ with smoothing over laryngeal tier

The new arc, from the `root` node of /ɪ/ to the `laryngeal` node, which has been inserted into the feature geometry representation, is highlighted using the colour red (although this is not apparent in the greyscale representation). The second `laryngeal` node from Figure 1 has been deleted.

In the current experimental model, the `laryngeal` tier and the `[continuant]` branch of the `oral_cavity` tier can be processed in this way. As already noted, in contrast to the treatment of melodies in [Cahill et al., in press], where the feature values are assumed to be simplex in the phonological domain (cf. [Carson-Berndsen, (1998)]), feature geometry representations assume a more complex internal organisation of speech sounds (cf. [Clements and Hume, (1995)]). We must, therefore, distinguish in this context between simplex and complex nodes; the `oral_cavity` node, as can be seen in the representation in Figure 1 dominates the `C_place` and the `[±continuant]` nodes. We cannot simply replace the arc to the `oral_cavity` node unless the complete subtree is identical. Currently one level of node complexity has been integrated: it is recognised that `oral_cavity` is a complex node and therefore only the `[continuant]` branch undergoes the nonlinearity constraint.

Applying smoothing to the `oral_cavity` tier in the syllable /ʃoːn/ currently results in the representation defined in Figure 3. Here a new arc is inserted from the `oral_cavity` node of the feature geometry representation of /oː/ to the `[+continuant]` node since both /ʃ/ and /oː/ have the same [continuant] values but the further internal organisation of their `C_place` nodes differ and thus smoothing cannot be applied here.

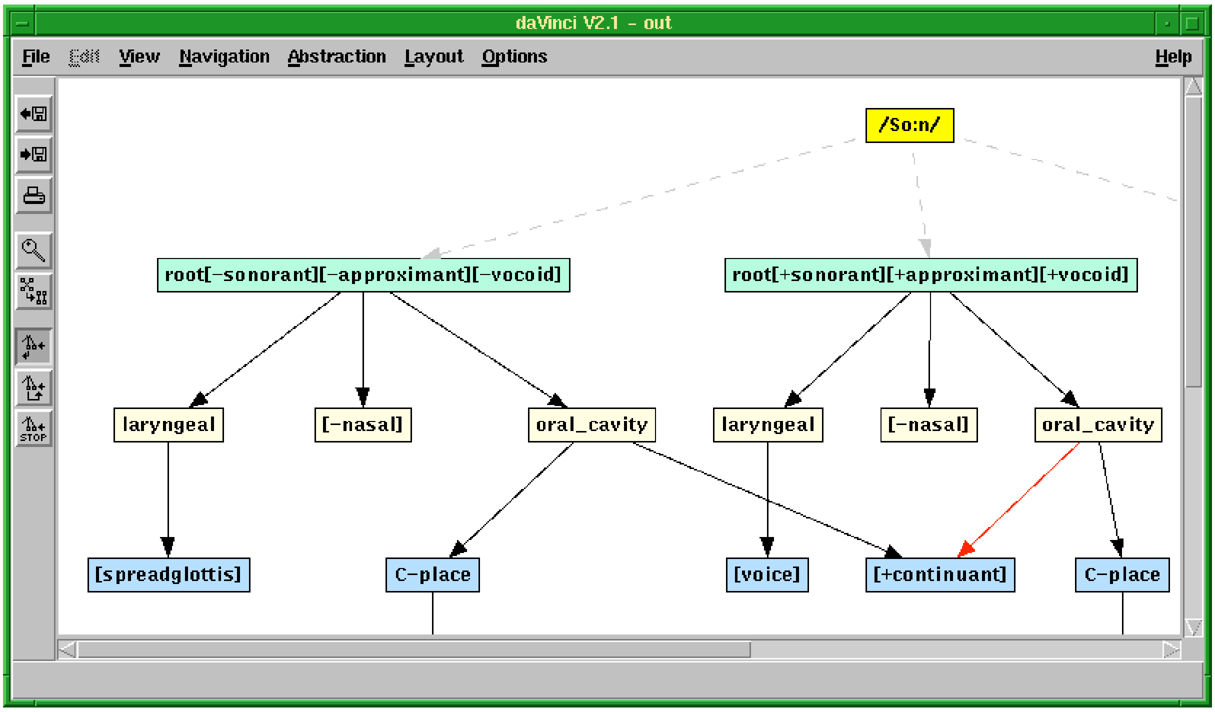


Figure 3: A subsection of the DaVinci representation for the syllable /ʃoːn/ with smoothing over oral_cavity tier

By incorporating temporal annotations based on average durations of context-dependent phonological segments, this representation can used to model speech variants and to provide coarticulation models for speech technology applications as suggested in [Carson-Berndsen, (1999a)] and [Carson-Berndsen, (1999b)].

4 Conclusion

This paper has presented a nonsegmental feature geometry based lexical description of syllables within the framework of a generic lexicon representation. In line with developments in computational phonology, the computational core of the generic lexicon is based on finite state transducers The lexicon, implemented in DATR, can be queried so as to output a range of types of information required both in linguistic analysis (for example different phonological structures) and in speech research (for example duration information for signal annotation). Any of these information types can be visualised on the fly, providing both feature geometry representations and temporal annotations for the individual features. The main emphasis of this paper was on generating graphical visualisations and text representations of the feature geometry descriptions in terms of bracketed structures for the lexical entries. However, the generic techniques described here could also be applied to complex multimodal tasks, for example that of integrating lexical information with the visual synthesis of articulatory movements synchronised with acoustic output.

The relationship of this work to to speech applications lies in the representation of nonlinear phonology in highly structured lexica allowing modelling of coarticulation phenomena and temporal interpretation of spoken forms. These aspects are discussed further with respect to a linguistic word recognition system in [Carson-Berndsen, (1999a)] and [Carson-Berndsen, (1999b)].

References

[Cahill et al., in press] Cahill, L. & J. Carson-Berndsen & G. Gazdar (in press). Phonology-based Lexical Knowledge Representation. In: van Eynde, F. & D. Gibbon editors, *Lexicon Development for Speech and Language Processing.* Kluwer Academic Publishers, Dordrecht.

[Carson-Berndsen, (1998)] Carson-Berndsen, J. (1998). *Time Map Phonology: Finite State Models and Event Logics in Speech Recognition.* Kluwer Academic Publishers, Dordrecht.

[Carson-Berndsen, (1999a)] Carson-Berndsen, J. (1999) A Generic Lexicon Tool for Word Model Definition in Multimodal Applications. In: *Proceedings of EUROSPEECH 99*, 6th European Conference on Speech Communication and Technology, Budapest, September 1999.

[Carson-Berndsen, (1999b)] Carson-Berndsen, J. (1999) A Feature Geometry Based Lexicon Model for Speech Applications. In: *Proceedings of IDS 99*, Interactive Dialogue in Multimodal Systems, ESCA Tutorial and Research Workshop, Kloster Irsee, June 1999.

[Clements and Hume, (1995)] Clements, G. N. & E. V. Hume (1995). The Internal Organization of Speech Sounds In: Goldsmith, J. A. editor, *The Handbook of Phonological Theory.* Blackwell Publishers, Cambridge MA: 245–306.

[Evans and Gazdar (1989)] Evans, R. & G. Gazdar (1989). Inference in DATR. In: *Proceedings of the Fourth Conference of the EACL*: 66–71.

[Evans and Gazdar (1996)] Evans, R. & G. Gazdar (1996). DATR: A language for lexical knowledge representation. In: *Computational Linguistics*, 22,2.: 167–216.

[Gibbon, (1998)] Gibbon, D. (1998). German Intonation. In: Hirst, D. & A. Di Cristo (eds.): *Intonation Systems; A Survey of Twenty Languages* Cambridge University Press: Cambridge.

[Gibbon and Ahoua, (1991)] Gibbon, D. & F. Ahoua (1991). DDATR: un logiciel de traitement d'heritage par defaut pour la modelisation lexicale. *Cahiers Ivoiriens de Recherche Linguistique (CIRL)* 27: 5–59.

[Gibbon and Strokin, (1998)] Gibbon, D. & G. Strokin (1998). *ZDATR Version 2.0 Reference Manual Version 1.0* University of Bielefeld.

[Goldsmith, (1976)] John Goldsmith (1976). *Autosegmental Phonology.* Indiana University Linguistics Club, Bloomington,Indiana.

[Pullyblank (1995)] Pullyblank, D. (1995). Feature geometry and underspecification. In: Durand, J. & F. Katamba editors, *Frontiers of Phonology.* Longman, London and New York: 3–33.

[Reinhard and Gibbon, (1991)] Reinhard, S. & D. Gibbon (1991). Prosodic Inheritance and Morphological Generalisations. In: *Proceedings of the ACL European Chapter Conference, Berlin*, April 1991.