ISCA Archive
http://www.isca-speech.org/archive

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

USING TWO-LEVEL MORPHOLOGY AS A GENERATOR-SYNTHESIZER INTERFACE FOR CONCEPT-TO-SPEECH

Georg Niklfeld, Hannes Pirker, Harald Trost*
Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Wien, Austria
e-mail: {georgn, hannes, harald}@ai.univie.ac.at

10.21437/Eurospeech.1995-305

ABSTRACT

In a project for the development of a concept-to-speech system for German, we apply extended two-level-morphology [8] to provide a unified solution for the tasks of morphotactics, segmental (morpho)phonology, syllabification and assignment of stress. Starting from a lexeme-based lexicon, we show that a declarative two-level-implementation of a single rule-corpus complemented with feature filters is sufficient for a comprehensive account of the various mutual influences holding between separate phonological dimensions in the phonology of German.

Information from higher levels of linguistic structure, up to textual representation, can be exploited in our system by performing a look-up of relevant feature-values through the filter conditions.

1. INTRODUCTION

In this paper we describe an integrated morphological and phonological component based on two-level-morphology [4] for a German concept-to-speech (CTS) system. The aim of this component is to mediate between the semantic/syntactic description produced by the linguistic generator and the phonetic/acoustic description of the speech synthesizer. The component deals with morphotactics, morphophonology, syllabification and stress.

The overall framework is an experimental CTS for German which consists of a unification-based linguistic generator and a demi-syllable based synthesizer. The component described in this paper bridges the gap between the feature-based descriptions produced by the generator on the one side and the synthesizer on the other side. In particular, it applies morphological and phonological knowledge to produce a phonological description that conveys segmental and prosodic information that can subsequently be used to drive the speech synthesis component.

In contrast to text-to-speech (TTS) systems CTS provide linguistic as well as pragmatic information about the message to be uttered. This rich amount of information should be made use of in order to improve the quality of the synthesized utterances. In particular, the available linguistic and pragmatic information can help to produce a more distinct intonation.

CTS prototypically are applied to restricted domains with a small vocabulary. Therefore, a CTS can rely on a small but information-rich lexicon. Furthermore, since there is no need for a graphemic representation all lexical entries are given in phonemic form. In our system a lexeme-based lexicon is used, i.e. derived words and compounds are part of the lexicon. Inflection is handled by the morphology though. Since German is a highly inflecting language, using full word forms would inflate the lexicon and make maintenance harder.

2. A TWO-LEVEL MORPHOLOGY APPROACH

Technically, the component is realized in the framework of two-level morphology (TLM) which is a proven technology with the advantages of declarativity and the availability of very efficient implementations. In our system we use X2MorF [9] which has the advantage of being fully integrated into the grammar formalism used by the generator. Also, in X2MorF linguistic information can influence the application of two-level rules [8]. This feature is used at various places in our phonological description, because we have to account for phenomena triggered at levels as high up as textual representation.

TLM has already been used in TTS (e.g. [7]) for the morphological analysis and synthesis of words. In our CTS TLM is exploited in a more extensive way. In addition to the treatment of morphotactics and morphophonology our system also performs genuinely phonological tasks, in particular syllabification and control of stress placement.

To perform these additional tasks some adjustments to the basic algorithm were necessary. The most important one is to apply the component to whole phonological phrases instead of single words as in traditional TLM. The component produces an annotated string of phones including information about word stress and syllable boundaries. This description then serves as input to the synthesizer component.

*This research has been sponsored by the Fonds zur Förderung der wissenschaftlichen Forschung, Grant No. P9755-PHY. Financial support for the Austrian Research Institute for Artificial Intelligence is provided by the Austrian *Bundesministerium für Wissenschaft, Forschung und Kunst.*

3. PHENOMENA HANDLED

Apart from describing segmental alternations triggered by morphological conditions, which is done very much in the spirit of existing TLM implementations of German–except that this time the representations are phonemic rather than orthographic–we also provide rules for the determination of syllabification and for certain necessary modifications of the lexically prespecified stress patterns.

Note that in the phonology of German these three dimensions mutually interact in various ways:

- Segments → Syllabification: Rules for the determination of syllabification can be formulated by making reference to certain patterns in the sequence of segments (reflecting sonority-curves).
- Syllabification → Segments: E.g. the segmental rule of final devoicing in German applies to consonants in coda-positions, thus being logically dependent on specified syllable structure.
- Syllabification → Stress: For any descriptions of the partial regularities governing main stress assignment in German words, syllable structure (i.e. syllable weight) is of crucial importance. Furthermore, theories of metrical phonology take the syllable as base unit of the prosodic system.
- Stress → Syllabification: [5] shows that in German the phenomenon of syllabification across word boundaries is restricted to contexts where the second word begins with an unstressed vowel. In this case then, syllabification depends on information about stress.
- Stress → Segments: Finally, segmental properties can be conditioned by stress. One of the examples given in [3] is glottal stop insertion before stressed vowels as highlighted by pronounciation variants like: *Kíosk* - *Ki$^?$ósk*.

This implies that a unified mechanism that can cover all three interacting dimensions of phonological representation in one single layer of parallel, declarative constraints can be regarded as favourable.

As motivation for the use of prosodic representations like syllables, notice further that in our system, enriching the segmental string with syllable-structure information is necessary not only because syllables play a role in the linguistic description: Preliminary experiments with the demi-syllable based component used for speech-synthesis have shown that indeed the quality of the acoustic output can be influenced positively if we make the rules selecting demi-syllables from the inventory sensitive to syllabification information. Syllabification thus is a prerequisite for further phonological and phonetic processing.

Though we rely on a rich lexical representation, the effects of syllabification cannot be entirely precoded in the lexicon because syllable-boundaries may shift when inflectional affixes are added. Furthermore, as mentioned above, syllabification in certain cases applies across word boundaries. In order to be able to model these processes, we have to perform syllabification on domains larger than single morphological words, which forces us to introduce the concept of phonological phrase in the phonological component.

However, the phenomenon of (re)syllabification on a phrasal level is highly constrained in German: It is restricted to contexts where the second word starts with an unstressed vowel and furthermore, at slow-to-normal speech rates it seems to be confined basically to one particular word class, namely the unstressed personal pronouns (cf. [5]). These pronouns often undergo segmental reductions as well and can therefore be analyzed as phonological clitics, such that for our present purpose we can analyze phrasal syllabification in German as a rule pertaining to some clitical elements, which are marked in the lexicon.

As for metrical phonology, it is clear that if the quality of the acoustic output of TTS- and CTS-systems is to be improved along the dimension of prosody, the correct determination of the location and prominence of stresses is a crucial task to be accomplished. Yet, in general the values of prosodic features are only partially determined by phonological influences. Pitch accent placement, for instance, is influenced by semantic, pragmatic and syntactic factors. The assignment of pitch accent to a constituent thus has to be decided by the linguistic generator. On the other hand, controlling the actual placement of the pitch accent within such a constituent is one of the tasks of the phonological processing.

As opposed to phrasal prominence relations, all word internal prominences have to be assigned in the phonological component itself.

German is commonly described as being a language with lexically marked stress, although many interesting partial regularities can be observed, which mirror influences from different parts of the grammar (cf. [2]). However, since it is not possible to exhaustively predict stress positions by rule, we have to mark stress in the lexicon. If our lexicon was morpheme-based, we would then still have to account for a wide variety of interaction-patterns between lexical stress-specifications of individual morphemes in the course of derivations. Starting from the lexeme-level, as in our system, one is left with two types of suppression of lexical stress-specifications which have to be accounted for by rules of the phonological component:

- Contrastive stressing in compounds

  As [1] argues, the frequent exceptions to the rule that German compounds receive main stress on their first constituent can be explained by referring to specific markedness-conditions on the semantic and textual level, which hold in such compounds for the destressed first constituent. E.g. in spoken weather-reports the lexicalized compound *Wétter+beruhigung* (calming down of weather) gets stressed *Wetter+berúhigung*, because in such texts the word weather is, in the terminology of [1], textually exophoric.

- Stress shift

  Word-internally conditioned stress shift is rare in German and can for the most be accounted for by lexical specification of the resulting patterns. However, on the phrase level a handful of cases of stress-shift exist: Stress shift can apply when the nominal head of a direct object which has ultimate stress immediately precedes a particle verb with stressed initial particle. E.g. *Hút aufsètzen* ("hat on-put", to put a hat on), with stress shift applied to the verb. Another case are so-called elative compounds like *stein+réich* ("stonerich", very rich), which are lexically equipped with a marked stress pattern (main stress on the final constituent) that gets adapted to initial stress quite easily if rhythmical alternation can be improved that way. Finally, we have to account for stress shift triggered word-internally by inflection: With the non-native Suffixes /+or/ and /+on/, stress shift applies upon concatenation of the plural inflectional morpheme /en/: *Dóktor* - *Doktóren.*

Although none of the described stress-processes appears to be very significant in terms of number of tokens affected, we still provide mechanisms to account for them in the phonological component, thus demonstrating the adequacy of a pure TLM-approach to the phonological processing needed in a lexeme-based concept-to-speech system.

4. MAPPING OF MULTIPLE TIERS ONTO A SINGLE STRING

The described phonological component defines a mapping between two strings of symbols, representing two levels of phonological structure: a lexical and a surface level.

Conceptually, each of the two string-representations must be understood as the result of collapsing a couple of linear encodings of different aspects of the utterance into a single sequence of symbols.

In that way, the following dimensions of linguistic structure are conveyed: Segmental representation (phonemic on the lexical level, allophonic on the surface level), word-level stress, syllabification and prosodic constituency, and morphological constituency.

For the purpose of the described component, the sequence of phonological segments is designated to provide the skeletal positions on which the units from the various planes of representation are linked together.

The notational convention applied in the string representations is that every non-segmental unit is linked to the position represented by the next segment to its right. This convention is complemented with rules that specify the scope of a unit when linked to a particular segmental position. For example, the scope domain of a syllable-marker begins on the link-providing first segment of the syllable-onset and is terminated by the next link of a syllable-marker to the right. In some cases, however, it is not possible to stick to this straightforward approach of linking to the left edge of scope domains: Lexical German stress is assigned to syllables, but syllabification is computed only at the surface level of the phonological component. Therefore, at the time when at the lexical level the units of the stress plane are linked to the segmental skeleton, the left edge of their scope domain is undefined. One technical solution for this is to link lexical stress markers to syllable nuclei, i.e. vowels.

For cases in which more than one non-segmental unit is linked to the same segmental position, a linear ordering convention is required to prevent linearization ambiguities in the collapsed representations. This ordering convention is actually implemented explicitly in the component as a separate rule licensing an otherwise void linearization operator, which is inserted obligatorily in the lexicon, because implicit checking in the independently motivated rules would lead to conceptual clumsiness:

In the present system, a number of two-level rules express generalizations in German phonology which make reference to only a subset of the planes included in the representation. Such rules can very simply be made "blind" to a particular plane by specifying the pair-alphabet of that plane to be ignored throughout the rule. However, if by these means two planes are made mutually insensitive, there remains no appropriate place where the ordering constraint holding between them in the string of the collapsed representation could be formulated. Thus by the following rule we license the linearization of the units of various phonological dimensions into a single string.

L:0 ⇒ #: _ [(BoundsUnit)(MorphUnit)(SyllUnit)
(StressUnit) SegmUnit]* #: ;

5. IMPLEMENTATION

Logically, we can discriminate between three different tasks the component has to perform. First, it deals with morphophonological phenomena (e.g., schwa insertion, consonant elision, umlaut). Another example is final devoicing which is dependent on syllabification. The rules for this part of the system are modelled in the spirit of [8]. Some adaptations and extensions were necessary due to the fact that in this application we have to deal with words in phonemic rather than graphemic form.

The second task, syllabification, is achieved rather straightforwardly by implementing a series of constraints on syllable-onsets as contexts for the licensing of a syllable symbol on the surface level, paired with a zero-element on the lexical level.

In particular, the implementation is inspired by [10], a detailed investigation of possible word-initial and word-medial onset clusters in German. This implementation of the well-attested Maximal Onset Principle is complemented by some additional conditions depending on the vowel quantity in the preceding syllable, also taken from [10].

Because the behavior of morpheme-boundaries with respect to syllabification is specified in the lexicon, syllabification is always computed over the appropriate domain, namely the phonological word.

In the same way the cases of syllabification on the phrasal level are accounted for: Unstressed personal pronouns beginning with a vowel carry a special diacritic that makes them look like part of the preceding phonological word for the syllabification rules.

Turning now to stress rules, we account for contrastive stressing in compounds by first introducing a new lexical symbol into our representations which marks one syllable in a morph that can potentially carry main stress under appropriate conditions.

We then need to percolate the feature `foregrounding` down from the textual representation onto the morph which bears the main stress in the environment. Note that for the feature-filters of X2MorF, only those features are visible that are specified at the morph containing the lexical:surface pairing to be licensed. In the same way we percolate a `backgrounding` feature down to the morph carrying the default main stress, and trigger destressing there by using another filter.

To implement stress shift, we provide appropriate values for a `prosodic-class` feature in the lexical feature structures of affected words (particle verbs, elative compounds, or/on-suffixed nouns) and define new contexts for the destressing and foregrounding rules, which trigger iff

- one of the feature-filters on `prosodic-class` is satisfied in the feature specification of the morph being processed and
- in the string representation, the metrical conditions for the stress shift are satisfied.

For example, the lexical word stress marked on the second constituent of an adjective specified as `[prosodic-class: elative-compound]` gets paired with surface secondary stress if surface word stress is detected on the initial syllable of the next word to the right. In parallel to this, a second rule checks the same feature-environment and pairs the lexical potential-word-stress marker on the first constituent of the elative compound with surface word stress if destressing on the second constituent is detected.

6. CONCLUSION

We have described a TLM-based phonological component for a concept-to-speech system. Interactions between the phonological dimensions of segmental representation, syllabification and stress are accounted for, as well as influences on phonology stemming from higher levels of linguistic representation. We arrive at a declarative model for the phonological and morphological facts that can be processed efficiently. A full listing of the rules of the described component is given in [6].

REFERENCES

[1] Benware W.A. (1987) "Accent Variation in German Nominal Compounds of the Type (A(BC))", *Linguistische Berichte* (108), 102-187.

[2] Jessen M. (1994) *A survey of German word stress*, ms., University of Stuttgart.

[3] Kloeke, W.U.S. van Lessen (1982) *Deutsche Phonologie und Morphologie. Merkmale und Markiertheit*, Niemeyer, Tuebingen.

[4] Koskenniemi K. (1984) "A General Computational Model for Word-Form Recognition and Generation" in: *Proceedings of COLING 84*, Stanford.

[5] Laeufer C. (1985) *Some Language-Specific and Universal Aspects of Syllable Structure and Syllabification: Evidence from French and German*, PhD-Thesis, Cornell Univ.

[6] Niklfeld G. (1995) *Phonologische Datenstrukturen in einem Concept-to-Speech System: Segmentale und Prosodische Phonologie*, ms.

[7] Russi T. (1992) "A framework for morphological and syntactic analysis and its application in a text-to-speech system for German" in: Bailly & Benoit (eds.), *Talking Machines*, North-Holland, 163-182.

[8] Trost H. (1991) "X2MORF: A Morphological Component Based on Augmented Two-Level Morphology" in: *Proceedings of IJCAI-91*, Morgan Kaufmann, San Mateo, CA.

[9] Trost H. & Matiasek J. (1994) "Morphology with a Null Interface" in: *Proceedings of COLING 94*, Kyoto.

[10] Yu S.-T. (1992) "Syllable Final Clusters and Schwa Epenthesis in German" in: P. Eisenberg et al. (eds.): *Silbenphonologie des Deutschen*, Gunter Narr, Tuebingen.