Using two-level morphology as a generator-synthesizer interface in concept-to-speech generation

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

Abstract

In a project for the development of a concept-to-speech system for German, we apply extended two-level-morphology (Trost 1991) to provide a unified solution to 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.

*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.*

1 Introduction

In this paper we describe an integrated morphological and phonological component based on two-level-morphology (Koskenniemi 1984) 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 Realization in the Framework of Two-Level Morphology

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 (Trost&Matiasek 1994) 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 (Trost 1991). 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., Russi 1992) 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 stress assignment.

To perform these additional tasks some adjustments to the basic algorithm were necessary. The most important one was 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.

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 – apart from the fact that this time we are working on phonemic rather than orthographic representations –, 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 → Syllables: Rules for the determination of syllable boundaries can be formulated by making reference to certain patterns in the sequence of segments (reflecting sonority-curves).
- Syllables → 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.
- Syllables → 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 the base unit of the prosodic system.

- Stress → Syllables: Laeufer (1985) 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 by Kloeke (1982) is glottal stop insertion before stressed vowels as highlighted by pronounciation variants like: *Kíosk* - *Ki*$^?$*ósk*.

This implies that a unified mechanism which 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.

Therefore, syllabification, i.e., the segmentation of the phonological string into syllables is to be seen as 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 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, judging from Laeufer's (1985) presentation, at slow-to-normal speech rates it seems to be confined basically to one particular word class, namely the unstressed personal pronouns. These pronouns frequently behave as phonological clitics also in terms of segmental reductions, such that for our present purpose we can analyze phrasal syllabification in German as a rule pertaining to some clitical elements, which are lexically marked.

As for metrical phonology, it can be said that the correct determination of the location and prominence of stresses is one of the most important tasks for both TTS and CTS if one wants to improve the quality of the acoustic output along the dimension of prosody. 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. Jessen 1994). 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 three types of suppression of lexical stress-specifications which have to be accounted for by rules of the phonological component:

1. Contrastive stressing in compounds

   While in the general case German compounds receive main stress on the word-stress bearing syllable of their first constituent – a tendency which seems especially strong in German as spoken in Vienna – there remain numerous exceptions. Benware (1987) explains these marked stress patterns by referring to specific markedness-conditions holding for the destressed first constituents in such compounds. He mentions several types of markedness at the semantic, and even at the textual level. For instance, one can make a quite reliable prediction that in spoken weather-reports the lexicalized compound *Wétter+beruhigung* (calming down of weather) will be stressed *Wetter+berúhigung*, because in such texts the word weather is, in Benware's terms, textually exophoric.

2. Stress shift across word boundaries

   Here it should be mentioned first of all that stress shift is much less frequent in German than in English. The only context where stress shift is widespread in German is in tri-partite compounds structured [A[BC]], where secondary accent is on the C constituent rather than B, while B would have to be stressed in occurences of [BC] in isolation – e.g. *Féier+tag* (holiday) vs. *Nationál+feier+tàg*. These cases are not relevant to our phonological component, since we can account for them already in the lexical specifications. What we are left with are a handful of cases of stress-shift across word boundaries.

   One environment where this can happen is when the nominal head of a direct object which is stressed on the ultimate syllable immediately precedes a particle verb with stressed particle. For example we get *Hút aufsètzen*, with stress shift applied to the verb. However, we assume that in order for this shift to happen, there must be some idiomatization involved. Stress shift without idiomatization seems to occur only with so-called elative compounds like *stein+réich* ('stonerich'), 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.

3. Stress Shift triggered 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 three 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 purely TLM-approach to the phonological processing needed in a lexeme-based Concept-to-Speech system.

4 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 Trost (1991). 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-boundary symbol on the surface level, paired with a zero-element on the lexical level.

In particular, the implementation is inspired by Yu's (1992) 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 length of the preceding syllable's vowel, which have also been taken from Yu's paper.

Because the behaviour 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 parts 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 syllables that can potentially carry main stress under appropriate conditions. We then need to percolate a feature specification denoting foregrounding down from the textual representation onto the morpheme which bears the main stress in the environment. Notice that in X2Morf, only features specified at the present morpheme are visible to the phonological filters. In the same way we percolate a backgrounding-feature down to the morpheme carrying the default main stress, and trigger deaccentuation by using another filter.

To implement stress shift, we mark the lexical phonological strings of the affected words with special phonological diacritics and define new contexts for the deaccentuation and foregrounding rules, which trigger when appropriate diacritics are found on two adjacent words, or, in the case of elative compounds, if an elative marker is detected and metrical conditions for the stress shift are satisfied.

Finally, we must create separate lexical entries for the inflectional endings that trigger the stress shift on words suffixed by /+or/ and /+on/, provide the strings for the phonological representations with diacritics and trigger rules of stress change by a combination of feature-lookup in a filter rule and checking of the phonological diacritic.

5 Conclusion

The three different tasks of segmental (morpho)phonology, syllabification and metrical phonology - represented by different rule sets - are performed in parallel in TLM. Thus we arrive at a fully declarative and modular model for morphology and phonology for our CTS.

References

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

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

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

Koskenniemi K. (1984) A General Computational Model for the Word-Form Recognition and Generation. In: *Proceedings of COLING 84*, Stanford.

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

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.

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

Trost H. & Matiasek J. (1994) Morphology with a Null Interface. In: *Proceedings of COLING 94*, Kyoto.

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