TAGGING SYLLABLES

*Brigitte Krenn**
Department of Computational Linguistics
University of the Saarland, Saarbrücken, Germany
krenn@coli.uni-sb.de

Abstract

Syllabification is viewed as a tagging task. Phonemes constituting a syllable are treated like words in a sentence. Each phoneme is annotated with information representing the phoneme itself, and its position within a syllable. Within a number of tagging experiments, the specificity of linguistic information represented in the tag set is varied. The annotation scheme which encodes an onset-nucleus-coda model is shown to lead to the best tagging results.

1 Motivation

In speech synthesis, knowledge of the syllable structure is indispensable, because:

1. Speech rhythm is related to syllable structure.
2. Stress and accent influences the realization of all phonemes in a syllable.
3. Unaccented syllables following and preceeding accented ones are potential targets for tonal movements.
4. Final lengthening affects all phonemes of accented syllables, and the coda of unaccented syllables.
5. [@] [r] sequences with and without syllable boundary are differently realized; [@-r] → [@r], [@r] → [6] (r-schwa in SAMPA).

For syllabification, the interplay of phonological and morphological aspects must be considered, such as:

Sonority: The sonority values of the sounds that constitute a syllable peak in the nucleus, and decrease towards the syllable edges (cf. [11]).

Phonotactics: Languages are subject to constraints on permissible phoneme sequences at specific syllable positions (cf. [11], [14]).

Morphology: Morphological processes influence syllabification, e.g.: Syllables cannot span over prefix or word boundaries. Suffixation may alter syllable boundaries (cf. [14]).

In statistical approaches, as opposed to rule-based approaches[1], there is no need for explicit modeling of phonological principles such as maximal onset and sonority, phonotactic rules, and morphological boundaries. Comparable generalizations are induced from specifically annotated training data. Thus, the amount of linguistic engineering reduces to the construction of a training corpus. This can be done automatically, provided a sufficiently large word list with syllable boundaries marked is available.

The work presented here is part of a concatenative text-to-speech system for German which is currently under implementation.

In the following, the underlying syllable model is characterized (section 2). A Markov Model for tagging syllables is specified in section 3.4. Tagset and training corpus are described in sections 3.3 and 3.2. Variations on annotation scheme and training material, and their influence on tagging results are discussed in section 4.

2 The Syllable

The approach presented is based on the following characterization of syllable:

- A syllable is a phonological unit organized around a syllabic peak, such as vowel, diphthong, schwa or syllabic consonant (=C). During syllabification, however, a syllabic consonant is represented as schwa consonant sequence, see for instance the German word [ze:-g@l] (Segel; En: sail. After syllabification [@l] changes to [=l].

*The author wishes to thank Ralf Benzmüller for valuable comments on the contents of the paper, and Thorsten Brants for kindly providing a HMM-based part-of-speech tagger.

[1] See for instance [4], [15], [8].

- As basic, surface-oriented syllable model we assume a C* V C* sequence where C stands for consonant and V for vowel. * is the Kleene star. For German, however, C* does not exceed 3 consonants preceeding and 4 following the nucleus (cf. [7]).
- On a more abstract level the C* V C* sequences are grouped into onset-nucleus-coda (onc) sequences with vowels, diphthongs and schwa at nucleus position. For discussions of onc-models see [7], [14].
- Ambisyllabic consonants, i.e. intervocalic consonants that belong to two syllables, are incorporated either into the coda of the preceeding syllable or the onset of the following. As an example see [fal@n] (fallen; En: to fall). Here [l] phonologically belongs to [fa] as well as to [@n]. Thus, we have either [fal-@n] or [fa-l@n].
- Extrasyllabic elements are treated as coda elements. For a discussion of extrasyllabicity see again [14].

3 Learning Syllabification

3.1 Tagging Syllable Structure

Similarly to the assignment of syntactic category to the words in a sentence (part-of-speech tagging, cf. [3]), the phonemes of a word are annotated with labels representing the phoneme itself and its position within a given syllable, such as nucleus, onset, coda. In this model, syllable boundaries are identified between coda-onset, coda-nucleus, nucleus-onset, and nucleus-nucleus tags.

3.2 Construction of the Corpus

Training and test data are automatically generated from CELEX-2, a lexical database for Dutch, English and German. Phoneme strings with syllable boundaries marked are automatically transformed into phoneme tag sequences, with tags representing the phoneme *ph* and its position within a given syllable. For illustration, five alternative annotations of the German phoneme string [fal@n] (*fallen*, En.: to fall) are given in table 1.[2]

The German corpus comprises a total of 320163 phoneme strings (word forms), consisting of 1149471 syllable tokens which reduce to 12159 syllable types. Average syllable length as computed from syllable types is 3.8 phonemes.

[2] We distinguish the following position tags: phoneme at the beginning (b), the end (e), in the middle (m), onset (o), nucleus (n), coda (c) of a syllable, ambisyllabic phoneme (a).

Phoneme	M1	M2	M3	M4	M5
f	fb	fb	fo	fo	fo
a	am	ae	an	an	an
l	le	lb	lc	lo	la
@	@b	@m	@n	@n	@n
n	ne	ne	nc	nc	nc

Table 1: Phonemic transcription of *fallen* with five different annotation models M1-5

3.3 The Tagset

In the case of the onc-model, a set of 66 tags is specified for the phonemes identified for German. If ambisyllabic elements are extra marked, the tagset increases to 80 tags. Hence the maximum number of tags per phoneme is three, the minimum number is one (cf. table 2).

3.4 The Statistical Model

Currently, Markov Models are the most successful approach to tagging. For the task at hand, a standard HMM tagger has been applied. Tags correspond to the states of the Markov Model, phonemes to the emitted signals. The probability of a particular annotation is defined as the maximized probability of the tag sequence t_i emitting the phoneme sequence ph_i. Thus, we have

$$\underset{t_i}{\operatorname{argmax}} \prod P(t_i|t_{i-2}, t_{i-1}) P(ph_i|t_i)$$

Particular features of the tagger are described in [1]. For a more general introduction to statistical approaches in natural language processing see [10].

4 Variation in Learning

In supervised learning, the statistical generalization process is guided by explicitly annotated linguistic information. Thus, deliberate selection of information represented in the annotation of the training data is of crucial importance for the learning result.

In experiments on PoS-tagging, it has been shown, that the kind of linguistic information available to the learner significantly influences the tagging result. (Cf. [6], [2], [13].) In order to test the validity of this claim for syllabification, training items and annotation scheme have been systematically varied.

4.1 Variation on the Lexicon

In order to vary the lexical training material, two strategies have been pursued:

Phonemes	Position Tag(s)	Ambiguity Classes	Max Number of Tags
2:,9,@,{,E:,I,O,U,Y,a,a:,e:,i:,o:,u:	n	M3-5:n	1
aI,aU,OY	n	M3-5:n	1
tS	o,c	M3-5:o,c	2
N	o,c,a	M3:c; M4:o,c; M5:c,a	2
b,dZ,d,g,h,j,v,z	o,a	M3:o,c; M4:o; M5:o,a	2
C,S,Z,f,k,l,m,n,p,pf,r,s,t,ts	o,c,a	M3,M4:o,c; M5:o,c,a	3

Table 2: German phonemes (SAMPA notation) and associated position tags according to the onc-model

1. Training and test material is equally distributed over the alphabetically sorted corpus.
2. For testing, portions of 20000 adjacent phonemes are extracted from the corpus. The rest of the corpus is used for training.

4.2 Variation on the Tagset

As for the annotation scheme, a linguistically motivated onc-model and a simple positional syllable model and are compared.

Onset-Nucleus-Coda Model: The phonemes constituting a syllable are assigned to onset (o), nucleus (n), and coda (c) positions. The following restrictions hold: each syllable must have exactly one nucleus element (recall: vowels, diphthongs and schwa are permissible nuclei); consonants preceeding the nucleus are part of the onset; consonants following the nucleus are part of the coda. With respect to ambisyllabic elements three variations are tested:

1. The ambisyllabic element is attached to the preceeding syllable (see M3 in table 1),
2. The ambisyllabic element is attached to the following syllable (see M4 in table 1).
3. As extra condition, ambisyllabic elements are explicitly marked (see M5 in table 1).

Simple Positional Model:
For control experiments, a simple positional annotation scheme is applied where the first (b) and the last (e) phoneme of a syllable, and the phonemes in the middle (m) of a syllable are distinguished. Ambisyllabic elements are attached to the preceeding (see M1 in table 1) or the following syllable (see M2 in table 1).

4.3 Training Results

With respect to tagset variation, the onc-model shows the best average[3] tagging results for ambisyllables tagged as coda or onset elements; cf. table 3, M3 and M4 respectively. Additional tagging of ambisyllables further reduces the tagging result, cf. M5. Tagging results become even worse when the model distinguishes only phonemes at the beginning, the end, and in the middle of the syllable, cf. M1, M2.

Extracting adjacent portions of 20000 phonemes from the training set decreases the accuracy as well; 98,12% mean accuracy for the onc-model with ambisyllables attached to the coda.[4]

Model	% correct	min	max
M3	98,34	98,23	98,46
M4	98,26	98,15	98,37
M5	94,17	93,99	94,48
M1	93,56	93,37	93,78
M2	92,64	92,44	92,87

Table 3: Tagging results

4.4 Discussion

There are two main reasons why the best tagging results have been achieved by applying the onc-model, namely: little ambiguity in the tagset, and adequacy of the annotation scheme.

While in the simple positional model almost all phonemes can be assigned three different tags, in the onset-nucleus coda model 15 phonemes have exactly one tag, only 14 phonemes are three times ambiguous in case ambisyllabic elements are extra marked. There is a clear correlation between the increase of ambiguity in the tagset and the decrease in tagging accuracy.

The interrelation between syllable structure and sonority values of phonemes is best represented by

[3] The average is computed from the percent correct resulting from 50 training and test runs that partition the alphabetically sorted corpus into 50 different pairs of training and test sets.

[4] Differences in tagging accuracy are highly significant, as for all pairwise comparisons of the outcome the percent correct of the one is consistently higher than of the other.

the onc-model, as sonority peaks and nucleus positions coincide. Phonotactic constraints which are also related to the syllable onset or coda are captured in the tri-grams.

In general, syllabification is well suited for statistical modeling, as the number of phonemes and their possible positions within syllables is small, and the variety of permissible syllables is restricted by phonotactic constraints. On the other hand, large amounts of training data can be easily constructed from word lists with phonemic transcription and syllable boundaries marked.

5 Conclusion

Syllabification is treated as tagging problem. It is shown that for the specific task state of the art tagging techniques lead to accuracy results that are at least comparable to or outperform other learning models. For a comparison of connectionist and symbolic approaches to learning cf. [5].

Learning models are preferable to rule-based approaches to syllabification as phonological and morphological regularities are induced from linguistically interpreted training data. Thus, explicit modeling of maximal onset and sonority principles, phonotactic rules, and the interaction of morphology and syllable boundaries is not necessary.

An annotation scheme which distinguishes onset, nucleus and coda positions and that treats ambisyllables as coda elements is shown to lead to the best training results (98.34 mean accuracy). This model is also comparable to the set of universal constraints determining syllable structure as specified within Optimality Theory, cf. [9], [12].

References

[1] Thorsten Brants. Tnt – a statistical part-of-speech tagger. Technical report, University of the Saarland, Department of Computational Linguistics, 1996.

[2] Jean-Pierre Chanod and Pasi Tapanainen. Creating a tagset, lexicon and guesser for a french tagger. In *Proceedings of the ACL Sigdat Workshop*, pages 58 – 65, Dublin, Ireland, 1995.

[3] Douglass Cutting, Julian Kupiec, Jan Pedersen, and Penelope Sibun. A Practical Part-of-Speech Tagger. In *Proceedings of the 3rd Conference on Applied Natural Language Processing*, pages 133–140, 1992.

[4] Walter Daelemans. Automatic hyphenation: Linguistics versus engineering. In F. Steurs F.J. Heyvaert, editor, *Worlds behind Words*, pages 347 – 364. Leuven University Press, 1989.

[5] Walter Daelemans and Antal van den Bosch. Generalization Performance of Backpropagation Learning on a Syllabification Task. In *TWLT*, Enschede, 1992.

[6] David Elworthy. Tagset design and inflected languages. In *Proceedings of the ACL Sigdat Workshop*, pages 1 – 10, Dublin, Ireland, 1995.

[7] Tracy A. Hall. *Syllable Structure and Syllable Related Processes in German.* Niemeyer, Tübingen, 1992.

[8] Georg Niklfeld, Hannes Pirker, and Harald Trost. Using Two-Level Morphology as a Generator-Synthesizer Interface in Concept-to-Speech. In *Proceedings of the 4th European Conference on Speech Communication and Technology*, pages 1223–26, Madrid, Spain, 1995.

[9] Alan Prince and Paul Smolensky. Optimality theory: Constraint interaction in generative grammar. Technical report, CU-CS-696 Dept. of Computer Science, University of Colorado at Boulder, 1993.

[10] Christer Samuelsson and Brigitte Krenn. A linguist's guide to statistics. Technical report, University of the Saarland, Dept. of Computational Linguistics, 1996.

[11] Andrew Spencer. *Phonology.* Blackwell, Oxford, Cambridge, 1996.

[12] Bruce Tesar. Computing optimal forms in optimality theory: Basic syllabification. Technical report, CU-CS-763-95 Dept. of Computer Science, University of Colorado at Boulder, 1995.

[13] Evelyne Tzoukermann, Dragomir R. Radev, and William A. Gale. Combining linguistic knowledge and statistical learning in french. In *Proceedings of the ACL Sigdat Workshop*, pages 51 – 58, Dublin, Ireland, 1995.

[14] Richard Wiese. *The Phonology of German.* [The Phonology of the World's Languages.] Clarendon Press, Oxford, 1996.

[15] Si-Taek Yu. Syllable final clusters and schwa epenthesis in german. In P. Eisenberg et al., editor, *Silbenphonologie des Deutschen*, pages 172 – 207. Narr, Tübingen, 1992.