SYLLABLE−BASED TEXT−TO−PHONEME CONVERSION FOR GERMAN

M. Libossek, F. Schiel*[+]

*Institute für Phonetik und Sprachliche Kommunikation (IPSK),
[+]Bavarian Archive for SpeechSignals (BAS)
Schellingstr. 3 VG/II, 80799 München, Germany

ABSTRACT

Due to the non−trivial relationship between the orthographic form and the chain of sounds in a spoken utterance in German, the text−to−phoneme conversion (TPC), as part of a text−to−speech system, is not a negligible task.

Many methods that use a fixed set of rules for TPC take into account the morphological structure of words. Even though this approach results in a high accuracy, it has one major drawback: the required morphological decomposition is difficult and error prone. In this paper we propose a new approach which uses the orthographic syllable instead of morphemes. The performance compares well with the traditional method, with the added advantage that the decomposition into syllabic units can easily be achieved by using existing hyphenation algorithms implemented in currently available word processors.

1. INTRODUCTION

Speech synthesis can roughly be structured into three major steps: Linguistic preprocessing of text input, text−to−phoneme conversion (TPC) and the actual synthesis of sounds. In some cases (e.g. Spanish) the TPC presents little or no difficulty since there is an almost one−to−one relationship between letters and sounds. Other languages however exhibit a non−trivial relationship between the orthographic form and the chain of sounds in the spoken utterance. This paper deals with the general problem of automatically generating a so−called canonical pronunciation from orthographic input by exploiting the syllable structure in the German language.

Many rule−based methods for automatic text−to−phoneme conversion take into account the morphological structure of words (e. g. [WOTHKE]). For most systems this approach results in a highly accurate TPC. The performance usually lies around 94 − 96% correctly transcribed words. Nevertheless, it has one major drawback: the required morphological decomposition is a tedious and error−prone task. Furthermore, errors that are introduced at this step are usually propagated throughout the whole speech synthesis process.

In this paper we propose a new approach which uses the orthographic syllable instead of morphemes. It turns out that the syllable in German proves to be a highly effective speech unit for TPC, resulting in a performance comparable to morpheme−based approaches. In addition, the decomposition into syllabic units can easily be achieved by using existing hyphenation algorithms implemented in currently available word processors. Since the syllable as the main unit for accent placement is needed anyway, its use for translation purposes facilitates the whole process, as the morphological decomposition can be cut out.

In the first part of this paper we will list some of the specific problems encountered in mapping German orthography to phonemes.

The second part gives a concise overview of the P−TRA system developed at Bonn university that was used for the generation and application of the different TPC rule sets.
The third part describes the four major rule types used in the syllable−based system: context−free rules (default rules), context−sensitive rules, affix rules and exception rules, and gives examples for each type respectively.

A comparison of the new approach using orthographic syllables (SYLL) with two other methods is given in the fourth section. Also, the definition of the bench mark test used to evaluate the quality of different approaches to TPC is given here. The other approaches are TPC considering only the graphematic input (BASE) and TPC taking into account the morpheme boundaries as the standard method (MORPH). The pros and cons of the different rule sets are discussed and quantitative results with regard to the bench mark test are given.

2. TPC PROBLEMS IN GERMAN

As in English diachronically motivated spellings combine with words in various states of Germanization to produce for some phonemes quite a variety of spellings.

There are, e. g. four possibilities to express an /i:/ none of which is not to some extent ambiguous:

<i>	−>	/i:/, /i/, /I/				
<ie>	−>	/i:/	vs.	<ie>	−>	/i@/
<ieh>	−>	/i:/	vs.	<ieh>	−>	/i:h/
<ih>	−>	/i:/	vs.	<ih>	−>	/i:h/

For a linguist there may be no doubt as to which version belongs to which surrounding, but for a naive system, such as an automatic conversion this represents an embarassment of riches. The problems arising from this abundance of combinations are further aggravated by the possibility to combine virtually any nouns, adjectives and adverbs into one single word, known as a compound.

2.1. Compounds

This procedure results in words where consonant or vowel combinations, typical for word beginnings can no longer be distinguished from those at the end of words since they all appear in the middle. Furthermore, quite a number of pseudoclusters are created that lead to a faulty conversion as they are translated by rules that wouldn't be applied had the words been separate.

An experiment undertaken to determine the effect of a separation of the compounds into their constituents before conversion by a morpheme−based rule set resulted in a gain of 4% accuracy. In view of this result we decided to precede the

TPC with a module for compound separation, marking word boundaries in the input string for the conversion process.

2.2. Glottal Stops

Another important problem that all TPC systems for German have to deal with is the glottal stop. In the canonical pronunciation, all spoken words starting with a vowel are preceded by a glottal stop. The same applies to vowels after prefixes, even if they belong to another prefix. The need for defining specific rules for the insertion of the glottal stop arises because it has no correlate on the orthographic level. The morpheme boundary alone doesn’t help since no glottal stop is spoken before suffixes.

2.3. Exceptions

A third problem that is by no means restricted to German TPC is the conversion of proper names and foreign words that are used in everday speech but keep their native spelling and for some of them their native pronunciation, too. A lot of English words, for example, have found their way into German and if the pronunction of these words isn’t really English, the resulting grapheme−phoneme correlation isn’t "normal" German, either. These words have to be treated as exceptions in all TPC systems regardless as to the method used for transcription. Most systems keep them in a separate list with the correct pronunciation added and for the most part they take up much more space than the normal rewrite rules. Our systems don’t prove to be exceptions to this rule, even though in our case the exceptions are stored in the normal rule set. Furthermore, only a very small number of them have been added, since the feeling was that including whole words is too easy a way of improving the performance of the system for a given list of words, before all other possible ways of improvement have been tried.

The next part gives a concise overview of the tool used for the developement and application of the rule sets.

3. THE P−TRA SYSTEM

The TPC system P−TRA (**P**honetische **TRA**nskription) developed by Dieter Stock [STOCK] at Bonn University in 1992 serves as the programming tool for the different rule sets. P−TRA consists of the interpreter and a separately stored rule file. This allows for easy modification of the rules as the program doesn’t have to be compiled every time some rule has been changed.

In accordance with Chomsky’s definition for rewrite rules, the rules are formulated in the following manner:

"left_context" ORTH. SEARCHSTRING & right_context = PHONEMIC CORRELATE.

So, an orthographic searchstring is transcribed as its phonemic correlate if the specified left and right context − both are optional − is encountered.

Within a word the rules are applied by moving from left to right through the word. To avoid the application of more general rules before more specialized ones, the rule file is ordered alphabetically with the special rules for each letter standing before the general ones.

The left and right context as well as the search string itself can consist of orthographic letters (single letters or longer parts of words) and boundary markers for word endings, syllables, compounds or morphemes. Furthermore, three features combine to turn P−TRA into a very powerful tool:

Boolean algebra: Several basic elements of Boole’s algebra are available to formulate very powerful left and right contexts. Through the use of AND, OR as well as brackets more than one possible context for a certain search string−phonemic output pair can be combined into one rule. Thus the number of lines of rules in the set is kept low and in consequence manageable. The negation (NOT) allows for the exclusion of specific contexts. Also, a wildcard is offered (Kleene operator), so that for a certain place in the context any letter is allowed, while for the following ones certain restrictions may be specified.

Using phonemic output: Taking into account the already transcribed phonemic output presents a further means for the formulation of the left context. If the conversion performance is low this procedure leads to even more mistakes since a faulty left context prevents the use of a "correct" rule. If, however, the performance has reached high levels so that a reliable transcription of the letters before the actual one(s) is achieved, this proves to be a very useful feature.

Classes of letters and phonemes: At the beginning of a rule set classes of letters and phonemes can be specified. So e. g. the class <K0> includes all orthographic consonants or <VB> which consists of the phonetic symbols of all German front vowels. The class names are used in the contexts, thus again allowing for a compact formulation.

4. THE SYLLABIC RULE SET

The first thing that had to be decided before the rule set for syllabic units could be created was whether the phonetic /phonological syllable or the orthographic syllable should be used as basic unit, as there are some discrepances between these two in German. We settled on the orthographic syllable since one of our main aims was an easy decomposition which is something existing hyphenation algorithms perform quite well. If the phonetic/phonological syllable is needed for further processing (e. g. assigning of accent) a simple script for postprocessing is sufficient. In the following we use the German SAM−PA to denote phonetic units [SAM−PA].

Within each of the alphabetically sorted rule blocks four different types of rules are used.

4.1. Context−free rules (default rules)

For every orthographic letter and graphemic combination of letters one "last" or "general" rule is given. They are applied if none of the more specific rules fit and are thus placed at the end of the rules for each letter in the set.

Examples for this type of rule:

A = a
SCH = S

The second example reads: if the combination of the orthographic letters <S>, <C>, <H> is encountered, transcribe it as /S/. This rule is included because <SCH> stands for one German phoneme only and thus has to be transcribed as a whole rather than as its separate parts.

4.2. Context−sensitive rules

In contrast to other TPC systems the biggest part of the rule set consists not of exception rules but context−sensitive rules which are used in most of the conversions. These are made up of search strings of varying length − from one letter to four or five letters with an equally variable amount of context. Here the advantages of considering relatively small (8 letters at most) and, in contrast to morphemes, few distinct units are clearly visible: The combination of different syllables is rather restricted, so that the concept of classes can be widely used. This results in contrast to the other rule sets developed with P−TRA (see part 5) in a smaller number of this type of rules.

An example:

"−##" E & <K0>(#<K0>, <K0>) = E

The above mentioned concepts of "AND", "OR" and "NOT" are here represented by:

- AND: Just by writing the elements of the context next to each other without space in between.
- OR: The Boole’s operator "OR" is represented by the comma (",").
- NOT: Represented as can be seen in the left context through the minus sign ("−").

The example reads: if the search string <E> does not stand at the beginning of a word (−##) and is followed by any of two consonants (<K0>) which are either next to each other or have a syllable boundary (#) between them, then translate it as /E/.

4.3. Affix rules:

While developing the first rule set with P−TRA, Stock found that even for a rule set using as restricting units only word boundaries, special rules for prefixes lead to an improved performance [STOCK]. Affixes in general need special rules and strongly influence the pronunciation − accent placement mostly − of the words they are combined with. This in turn influences the quality and quantity of the vowel(s) following the prefix. Thus the performance of the syllable−based system, too, benefited enormously from rules for the conversion of affixes per se, as well as from the use of affixes as restricting contexts.

An example for this kind of rule:

"##, <K0>#" VER & #<K0;S>, #S−(I#O) = fE6

This rule reads: if <VER> stands at the beginning of a word (##) or after a consonant (<K0>) followed by a syllable boundary (#) and if it is followed by a syllable boundary and any consonant except S (signalled through the use of the semicolon (;)), transcribe it as /fEA/. If however the left context applies and it is followed by a syllable boundary and an <S> and the <S> is not (−) followed by <I>, syllable boundary (#), <O>, then, too, transcribe it as /fE6/. This is necessary to prevent the transcription of <VERSION> (same in English) as /fE6zjo:n/ as this would be incorrect. Correct: /vE6zjo:n/.

4.4. Exception rules

This last type of rules usually plays a very important role in the rule sets since the more exceptions are added, the better the performance becomes. As will be seen in the next part, the syllable−based rule set in contrast to the other sets needs the least exception rules.

On the one hand, exception rules give the transcription of widely used foreign words e. g. <SOFTWARE> that deviate from the normal German pronunciation. On the other hand several German words − Germanizations from Latin or Greek mostly − that are nowadays regarded as "German" words often need a special translation: e. g. the word <MUSIK> (engl. music) requires for both its syllables a special rule. In cases like this, it is easier to formulate one exception rule:

MU+SIK & ## = mu+zi:k

The "+" stands for the syllable boundary in both the orthographic and the phonemic string. So, if <MU+SIK> is followed by a word boundary (##) transcribe it as /mu+zi:k/.

5. EXPERIMENTS AND RESULTS

As mentioned in 4.2, the first rule set developed with the P−TRA system by Stock [STOCK) considered as restricting units only word boundaries. The idea behind this procedure was to find out, how well a system would perform that uses only a very limited amount of linguistic knowledge. The resulting rule set was further refined at our institute. This BASE system consists of 1304 rules. The aforementioned four types of rules appear, too, although in a somewhat different distribution.

Using this system as a basis, another set was developed that follows the traditional way of considering morphemes for transcription. An experiment about the usefulness of considering compound boundaries proved to be so successful that several rules were added to utilize this information. The resulting rule set (MORPH) lies with 1327 rules slightly above the BASE system.

Compared with these two, the newly developed syllable−based system (SYLL) lies with 650 rules dramatically lower. Responsible for this difference is the lesser number of context−dependent rules and exception rules (BASE: 200, MORPH: 150, SYLL: 61). [LIBOSSEK]

The benchmark test: To be able to compare the performance of these three methods a testlist was compiled in such a way that for every combination of three, four and five letters that occured at least 50 times (for the first two groups) and 20 times (for the last) in a corpus of 1Mio words at least one word was included. The resulting list of 4685 words was manually transcribed according to [DUDEN] and morpheme boundaries or syllable boundaries were added for the MORPH system and the SYLL system respectively.

5.1. Results and Discussion

The results of a comparison of the three systems, using the benchmark test are shown in table 1.

	BASE	MORPH	SYLL
Correct words	3845	4523	4424
Excep. errors	468	130	204
Glottal stops	192	1	1
Other errors	180	31	56
Correct %	82.1	96.5	94.4

Table 1: Comparison of the performance of the three rule sets

The errors that each system produces can be classified into three major groups:

Exception errors: 1700 of the 4685 words of the benchmark test list fall under the heading of foreign words. Among them we have words of technical jargons, new and old foreign words with high frequency as well as some proper names. In relation to all committed errors exception errors gain steadily in importance over the three sets, starting with 55.71% of all errors for the BASE system, rising to 78.16% for the SYLL set and reaching a maximum of 88.44% for the MORPH set. This trend exists because the other errors just as steadily loose in importance. This point is one of the big weaknesses of the BASE system: even though a great many exceptions are included in the set, the performance still is the worst of the three sets. The reason lies probably in the fact that when the words are separated into smaller units as in the other two systems, affixes as "special" parts of words can be more reliably distinguished. This allows for a better conversion, even for foreign words.

If we keep in mind that the SYLL system uses the least exceptions of the three, it compares quite well with the MORPH system. Here the fact that every system can be improved by adding exception rules is stressed once again. So this is one of the screws that can still be adjusted to perfect the new SYLL system.

Wrongly placed or missing glottal stops: One of the interesting facts of these experiments proved to be the extinction of this type of mistake through the use of either morpheme and compound boundaries or syllable and compound boundaries. The combined information of a clearly marked word or wordpart boundary as well as the use of well−defined affix rules, reduced an error that occured in 17.29 percent of all mistakes for the BASE set to a neglegable 0.68% in the latter two systems. For the BASE set this error can be split into two groups. On the one hand, there are words that are correctly transcribed but are missing their glottal stop.

On the other hand, as there is a phonological rule (final devoicing) to turn voiced plosives and some voiced fricatives into their voiceless counterparts at word endings − even if they appear in the middle of a compound, a missing glottal stop is very often also accompanied by a failure to perform this conversion. We consider this combination of missing glottal stop and incorrectly performed final devoicing as a second version of this type of error.

Other mistakes: A third group of errors includes two main kinds that can’t be easily separated:

- Application of incorrect rules.
- Some German words require exception rules.

The main reason for the first type lies again in the missing boundaries. It results mostly in an incorrect transcription of affixes at compound boundaries.

The second type leads to an incorrect conversion of seemingly "normal" German words. Some words deviate from the pronunciation expected from the spelling and accordingly need a special rule. As expected, the BASE system is the most susceptible of all three to this error, with a sharp reduction for the other two sets. The rise in errors for the SYLL system can again be explained with the lesser number of rules. It is highly likely that in the more than twice as many rules of the MORPH system some of the German exceptions are included, probably in the context−sensitive rules. This is another screw that can be adjusted to improve the SYLL system.

6. SUMMARY AND CONCLUSION

The aim of our paper is to draw attention to one method of TPC for speech synthesis that has been rather neglected so far: syllable−based conversion. Most systems today first perform a decomposition into morphemes, then apply rules for the transcription of these units [WOTHKE] or even look them up in a morpheme lexicon [HEUVEN, POLS] and then apply rules to combine the morphemes into words.

In a comparison of three rule sets − one using only word boundaries as linguistic information, the second utilizing morphemes as restricting units and the new method of considering syllables for conversion − it could be shown that both of the systems that use smaller units for conversion i. e. the MORPH and the SYLL system perform significantly better for all types of grapheme−phoneme problemes of the German language.

Comparing the results for a benchmark test for these two systems however, very few differences in performance can be found. Obviously morpheme boundaries as well as syllable boundaries, in collaboration with marked compound boundaries enable a similiarly high level of accuracy for conversion. The big advantage of syllables as opposed to morphemes is the ease with which it can be extracted through the use of existing hyphenation modules in current word processors.

7. REFERENCES

1. [DUDEN] Mangold, M. "Wörterbuch der deutschen Standardaussprache," Dudenverlag, Mannheim, Wien, Zürich, 1990.

2. [HEUVEN, POLS] Heuven, Vincent J. Van; Pols, Louis C. W. (eds.) "Analysis and Synthesis of speech," Walter de Gruyter, Berlin, 1993.

3. [LIBOSSEK] Libossek, M. "Automatische Graphem−Phonem Übersetzung unter Berücksichtigung linguistischer Einheiten im Deutschen," Magisterarbeit, LMU München, 2000.

4. [SAM−PA] www.phon.ucl.ac.uk/home/sampa/home.htm.

5. [STOCK] Stock, D. "P−TRA − Eine Programmiersprache zur phonetischen Transkription," in: Hess, W.; Sendlmeier, W. (eds): Beiträge zur Angewandten und Experimentellen Phonetik, Franz Steiner, Stuttgart, 1992.

6. [WOTHKE] Wothke, K. "Automatic phonetic transcription taking into account the morphological structure of words," in: Technical Report, IBM Germany, Scientific Center, Heidelberg, 1991.