MORPHOLOGICAL ANALYSIS FOR A GERMAN TEXT-TO-SPEECH SYSTEM

Amanda Pounder, Markus Kommenda
Institut für Nachrichtentechnik und Hochfrequenztechnik
Technische Universität Wien
Gusshausstrasse 25, A-1040 Wien, Austria

ABSTRACT

A central problem in speech synthesis with unrestricted vocabulary is the automatic derivation of correct pronunciation from the graphemic form of a text. The software module GRAPHON was developed to perform this conversion for German and is currently being extended by a morphological analysis component. This analysis is based on a morph lexicon and a set of rules and structural descriptions for German word-forms. It provides each text input item with an individual characterization such that the phonological, syntactic, and prosodic components may operate upon it. This systematic approach thus serves to minimize the number of wrong transcriptions and at the same time lays the foundation for the generation of stress and intonation patterns, yielding more intelligible, natural-sounding, and generally acceptable synthetic speech.

1. INTRODUCTION

Many applications of computer speech require unrestricted vocabulary. In particular, voice output units of this kind permit the linkage of the common telephone network to a central computer, thus enabling access for a large public. "Karlchen", the Frankfurt talking railway timetable, and other automatic information services are based on this principle.

If a written text serves as input to a speech synthesis system with unrestricted vocabulary (text-to-speech synthesis), the derivation of a correct and natural-sounding pronunciation and intonation must be provided for. The software module GRAPHON (GRAPHeme-PHONeme-conversion) has been developed to convert any given German text into its phonetic transcription (I.P.A.), enriched by some prosodic markers.

The text-to-speech system is being implemented on an HP 9816 workstation system with a 68000 CPU and 768 kbyte of RAM. At present a SSI 263 phone synthesizer serves as acoustical output unit; a simplified articulatory model used to control a refined digital vocal tract synthesizer is under development. The software is written in PASCAL and operation of the whole system is expected to be almost real-time. (For further implementational details cf. [1].)

While text-to-speech systems for the English language are fairly advanced, there is much room for development for German speaking systems. It is possible only to a limited extent to profit from work in the field of English. Obviously, German pronunciation rules differ from those of other languages; however, the mere replacement of a given grapheme-to-phoneme conversion rule by another is inadequate to meet the demands of the very different principles on which two writing-systems are founded. This also applies to the structural levels of morphology and syntax.

2. MOTIVATION FOR A MORPHOLOGICAL COMPONENT

The application of an English pronunciation rule is lexically determined, that is to say, is restricted to a generally arbitrary subset of the lexicon (compare, for example, the values of <ea> in the sets {*bread, head, thread...*} and {*knead, bead, heat...*}). It is for this reason that many English-based systems include very extensive dictionaries, for example the pioneering work of Allen [2] with a 12000 morpheme lexicon. On the other hand, German rules have in general a much wider scope of application, which has led researchers working in the field of German to consider large lexical inventories unnecessary. The inventories in e.g. SAMT [3] or SPRAUS-VS [4] are thus restricted to function words needed for the syntactic analysis (prepositions, pronouns, articles, etc.). Similarly, our earliest efforts in this area were based on a small lexicon and an extensive rule catalogue; however, numerous incorrect transcriptions at morphological boundaries and the frequent recurrence to ad-hoc rules (cf. [1]) made the lack of some sort of morphological indicator apparent.

However more closely German spelling may reflect pronunciation than is the case in English, difficulties arise in producing a correct pronunciation automatically if knowledge available to the human speaker, such as the internal structure of a given word or its native as opposed to foreign origin, is not made use of. The following examples should suffice to demonstrate the relationship between morphology and the values of the written symbols:

- One fundamental rule is that vocalic quantity is determined by the number of following consonants: the first rule given in the DUDEN Aussprachewörterbuch [5] states that <a> is to be pronounced /a:/ when followed by only one consonant grapheme before the stem boundary, so that the inflectional form *rast* of the verb *rasen* ("rush") becomes /ra:st/, whereas the simplex noun *Rast* ("rest") becomes /rast/.
- Consonant or vowel groups may be assigned digraph or trigraph value only when they appear within morphological boundaries; compare for example the different values of <sch> in *löschen* /ʃ/ ("extinguish") und *Höschen* /sç/ (dim. of "pants"), or of <ei> in *Geier* /ai/ ("vulture") and *geirrt* /əˈɪ/ ("erred").
- The first stem syllable in German (native stock) receives the primary word stress, a rule which implies this stem's being identifiable; compare *geben* /'ge:bn̩/ ("give") and *Gebein* /gə'bain/ ("bones").

These phenomena play a role in the domain of derivation and inflection, which has been dealt with in several systems, e.g. SYNTEX [6] or REDE [7]; these do contain lists of common prefixes and suffixes to permit affix-stripping, although they are predominantly rule-based. The same problems are found in the field of composition; their import is heightened by the very great frequency of this process in the German language. Still, Rühl [6] proposes a decomposition algorithm which relies on distributional criteria and on lists of consonant clusters in initial and final position (based on Kästner [8]). Other authors too prefer to minimize the lexical component: "The attempt to incorporate this problem into a mainly rule-based system seemed to us to require a rather great and thus undesired step towards a kind of dictionary approach" ([9], p.226).

It is however certainly possible to make a case for a morphological analysis containing a morph-lexicon of some depth. The conversion program presented here makes extensive use of such an analysis component (see fig. 1) and thus in our opinion profits from the following advantages:

- inflection, derivation, and composition can be treated simultaneously, more economically, and with a reduced number of incorrect segmentations; this latter is achieved by specifying the respective environments of potential elements;
- simple and efficient treatment of exceptions, for instance the pronunciation of foreign words; this and the preceding result in a reduced transcription error rate and in simplified and more transparent grapheme-to-phoneme conversion rules;
- correct placement of word-internal boundaries, labelling of the constituents and the lexically stored information concerning native vs. foreign status favour accurate word stress assignment;
- the lexicon-based approach prepares the ground for word classification and extraction of certain syntactic constraints, providing the input for an elementary sentence parser.

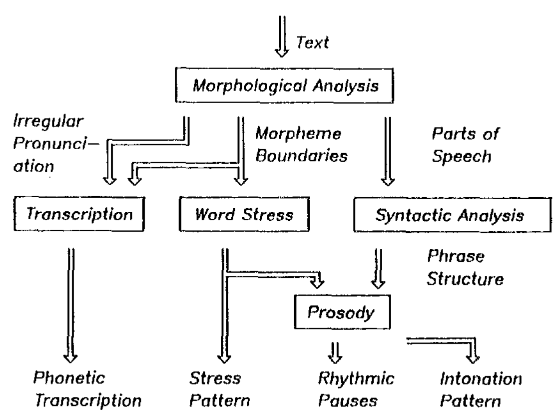

Fig. 1: The role of the morphological component within GRAPHON

3. SKETCH OF THE MORPHOLOGICAL COMPONENT

3.1. Lexical Inventory

Morphological analysis in our system relies on a single lexicon rather than on separate lists of, say, prefixes, stems, junctures etc. The entries in this lexicon are morphs and not morphemes in that stem variation, i.e. processes such as umlaut (e.g. *Apfel* - *Äpfel* "apple"), ablaut (*lauf* - *lief* "run") and e-deletion (*trocken* - *trockn-* "dry") are not covered by rule but by storage of allomorphs. As we are not concerned with generation, this appears to be the most practical method. Forms that are in some way irregular are then naturally provided with individual entries, for example anomalous verb forms (*sein* - *bin* - *war* - *wär* - ... "be") or forms of the definite article (*der, die, das, dem,...*). We have chosen to set up the most basic forms wherever possible, e.g. NAM- as opposed to NAME (nominative singular), which permits an economical treatment of derivation and inflection. As a matter of fact, the overriding principle governing the decision what exactly should constitute an entry is a pragmatic one: for example, rather than taking sides on linguistic, historical, or psychological grounds in such controversial cases as *antwort-* vs. *ant + wort-* ("answer"), *himbeer-* vs. *him + beer-* ("raspberry"), or *verlier-* vs. *ver + lier-* ("lose"), we choose the solution favouring the ideal functioning of the system as a whole.

3.1.1. Structure of a Dictionary Entry

A dictionary entry consists of the lemma, i.e. graphemic representation of the morph, on the one hand and an information-tree, serving to characterize its phonological, morphological and syntactic value on the other.

A number of practical conventions has been set up for the form of the lemma: a given morph is represented by a maximum of ten lower-case letters ; the diacritic sign " (umlaut) is made use of (cf. other systems which decompose the vowels in question as <ae>,<oe>,<ue>); likewise, the sign <ß> is not replaced by <ss> either in the input text or in the lexicon. An orthographic rule of German states that <ss> becomes <ß> before a consonant or a word-boundary, so that the latter sign's usual function as an indicator of vowel length is neutralized in these positions (compare *Flüsse* "rivers" vs. *Füße* "feet" with *Fluß* (/ʊ/) vs. *Fuß* (/u:/)); this "defect" (cf. [10], p.108) can be got round by maintaining the opposition between <ss> and <ß> in the lemma.

The information-tree contains classificatory data pertaining to the morph itself and to those it may immediately select; they concern morphological status (lexical stem - particle - derivational morph - inflectional morph - juncture -...), native or foreign status, and combinatorial restrictions. In addition, the lexicon allows the introduction of information for the assignment to parts of speech and, wherever necessary, indications as to exceptional pronunciation or stress pattern.

3.1.2. Extent of the Lexical Inventory

At present the lexical inventory comprises some 2000 entries, the choice of which was based on Ortmann [11], itself compiled from four frequency lists. As for the contents of the entries, we relied on Augst [12], Mater [13], and Wahrig [14]. For the ongoing testing, revision, and supplementing of this primary list we depend on the frequency list in Meier [15] as well as on sample texts from various random sources. Inasmuch as affixes, particles, and junctures (at least native ones) constitute closed classes, they should be represented exhaustively in the inventory. This is unfortunately not the case as soon as one turns to foreign elements, to whose number are always being added new candidates. Moreover, it is very difficult if not impossible to establish general principles according to which foreign suffixes in particular may be isolated and the dividing line found between stem and suffix.

Proper nouns are represented only to a very limited extent; their range should be adapted to the requirements of the task at hand. In fact, the compilation of the inventory has been carried out with the aims of expandability and maximum flexibility.

It is of course not to be expected that the lexicon would ever cover the entire vocabulary of a native speaker, nor is that our intention; consequently, we foresee a "joker morph" which can stand for any stem that may happen to occur. This is made possible by the generalization that a German stem conforms to a number of structural principles: for example, every stem must contain a vowel and the variety of consonant clusters in initial, medial, and final position is restricted (cf. [8]). An even more general canonical description can be exploited in the case of foreign elements. Such a device has not yet been implemented.

For the time being, 64 kbyte have been reserved to accomodate the lexical inventory. Note that all lexical data as described above are coded so as to achieve maximum storage efficiency.

3.2. Word Parser

The segmentation of a given (complex) word is carried out automatically in a series of steps; the process is bound from the very first of these to the dictionary, as stated above. Just as the human speaker seeks familiar units in his identification of a word, the automatic analysis considers for further attention only those segments which correspond to forms available in the lexicon, such that the segments are contiguous and no letters are left unaccounted for. Thus a segmentation such as *mein + un + g* for *Meinung* ("opinion") could not be produced in the first place, as *+g+* has no representation in the lexicon. The number of potential analyses is further reduced by the fact that no boundaries are searched for in a word corresponding identically to a single unit in the lexicon, for example *der* would not be analyzed as *d + er* or *d + e + r* . For reasons of run-time efficiency, a strategy is used which "prefers" the longest segments, starting from the beginning of a given word; thus *deck + en* ("cover") would be the first segmentation proposed before *d + eck + en*. The usefulness of this principle can be seen from an example like *Eintritt* ("entrance"), where the order of segmentations would be: *ein + tritt, ein + t + ritt, ei + n + tritt, ei + n + t + ritt, e + in + tritt, e + in + t + ritt, e + i + n + tritt, e + i + n + t + ritt.* The first decision proposed by the parser can be proved to be the correct one in the overwhelming majority of cases, which allows us to delay requiring a second proposal until the first has been rejected on structural grounds in the following step of the analysis procedure.

In this second step the proposed segmentations are examined as to their conformity to the principles of German morphological structure. The following structural formula describes every German word, whether of native or foreign origin:

$$[P_0^2 + S + D_0^5 + J]_0^\infty \ \# \ P_0^2 + S + D_0^5 + I_0^1$$

whereby:

- X_a^b ... there may be between a and b segments of this type in a given structure
- +,# represent morphological boundaries of different strengths (differentiation relevant for the context of certain phonological rules)
- P... Particle (in general equivalent to inseparable prefixes, e.g. *+ent+*, *+prä+*)
- S... Stem
- D... derivational morph, always a suffix (e.g. *+ig+*, *+ung+*)
- I... inflectional morph, always a suffix (e.g. *+en+*, *+em+*)
- J... juncture morph (e.g. *+es+* in *Bundesbahn* "national railway")

The segmentation is assigned a structural description by matching the combinatorial features of each unit with the morph status information of its neighbour as given in the respective lexicon entries. A morph may be specified according to the following properties and in turn select certain values for these properties in its neighbour:

- native or foreign status,
- lexical functionality (this property is manifested by the capacity to receive inflection),
- morphological status (as in the above structure definition with additional detailed classification), and
- lexical class, i.e. part of speech as reflected in the inflectional ending.

Specification of these properties is optional; however, the more information provided, the more restrictions with respect to the general structure formula are achieved, so that the number of potential labellings is reduced and the labellings themselves bear more information. Thus, it is possible to provide at least a partial treatment for words whose stems are not represented in the 2000 entry lexicon.

Should no match be obtained in this step, the process is repeated with a new segmentation until compatible sets of features are found.

```
+ ant + eil +
+ an + teil +
 f1SX fLSX    fXIN
 f1SX fLSX    fXIV

+ er + werb + st + ät + ig + en +
+ er + werb + st + ät + ig + e + n +
+ er + werb + st + ät + i + ge + n +
+ er + werb + s + tät + ig + en +
 fXPX fLSX   fXJX fLSX  fXDX fXIA
 fXPX fLSX   fXJX fLSX  fXDX fXIV
 fXPX fLSX   fXJX fLSX  fXDX fXZX fXIA
```

Fig. 2: Sample segmentations and structural specifications

f/F...native/foreign
l/L...lexical/non-lexical
X ... unspecified
P ... particle
S ... stem
J ... junction
D ... derivational morph
I ... inflectional morph
Z ... participle morph
N ... noun
A ... adjective
V ... verb

Fig. 2 presents examples of the resulting segmentations and labellings. We see that the first segmentation of *Anteil* ("portion") is rejected, as in this case the stem would be preceded by a suffix (*+ant+* being a longer segment than *+an+*, it has received "priority" up to this point). In the second segmentation, *+an+* is correctly recognized as a non-lexical stem, upon which a lexical stem may follow. It is not possible to specify the lexical class selected by *+an+*, as it combines with all parts of speech; and as *+teil+* can function as a noun or a verb stem, there result two potential labellings. The ambiguity cannot be resolved at this stage.

The following example is somewhat more complicated. Crucial here is the boundary between the two stems of the compound *Erwerbstätigen* ("employees"): the phonological consequences of an error (/ʃt/ instead of /st/) are quite serious. After the correct segmentation has finally been found, three possible interpretations are proposed. Note that *+en+* can serve as a participle morph (Z), so that the word would syntactically function as an adjective.

The third step consists of additional checks and finer specifications in order to isolate the correct structure and part-of-speech assignment for the whole word. For instance, if a suffix has been identified as a possible past participle morph, this could be verified by searching for a corresponding prefix (cf. *teil + t* "shares" vs. *ge + teil + t* "shared"). Another check could exploit certain restrictions on the sequence of lexical and non-lexical stems in a complex word. Such tests have not as yet been implemented.

The lexical class of a German word is, generally speaking, determined by its last element, so that the classification algorithm makes use of the results of the matching process at the end of the word. Some derivational morphs, e.g. *+ung+*, *+keit+*, *+isch+*, permit unambiguous classification. Unfortunately the same cannot be said of inflectional endings in particular and many other elements as well, taken alone. By exploiting the combinatorial information, however, many ambiguities are eliminated; moreover, capitalization can be treated as a signal for the lexical class noun.

Each text unit is now provided with a structural specification such that the phonological, syntactic and prosodic components may operate on it. Fig. 3 shows segmentations and lexical class assignment for a sample sentence; based on these, the phonological component already in place determines the correct pronunciation and generates the I.P.A. transcription, also given in fig. 3.

(R)	+die+	[diː]
(A,V)	+richt+ig+e+	[rɪçtigə]
(N)	+zer+leg+ung+	[tsɛɐ̯leːgʊŋ]
(R)	+von+	[vɔn]
(N,A)	+wört+ern+	[vœrtɐ̯n]
(V)	+ist+	[ɪst]
(A,V)	+wicht+ig+	[vɪçtiç]
(R)	+für+	[fyːɐ̯]
(R)	+die+	[diː]
(N)	+be+stimm+ung+	[bəʃtɪmʊŋ]
(A,N,V,A)	+ihr+er+	[iːrɐ̯]
(N,V)	+aus+sprach+e+	[aʊsʃpraːxə]
(R)	+und+	[ʊnt]
(N)	+be+ton+ung+	[bətoːnʊŋ]
(R)	+und+	[ʊnt]
(R)	+für+	[fyːɐ̯]
(R)	+die+	[diː]
(N)	+er+zeug+ung+	[ɛɐ̯tsɔigʊŋ]
(R)	+der+	[deːɐ̯]
(N)	+satz+melod+ie+	[zatsmelodiː]

Fig. 3: Sample segmentations, lexical class assignment and resulting I.P.A. transcription

N...noun; A...adjective; V...verb; R...other

4. CONCLUSION

Although extensive tests on large corpora have not as yet been carried out, experiments with our current system permit evaluation of following aspects of the morphological analysis component in GRAPHON:

- The development of the phonological component has shown that the setting up of a catalogue of pronunciation rules became simpler and more systematic, and at the same time, the rate of transcription errors could be greatly reduced.

- A relatively limited number of lexical entries is capable of handling a considerable quantity of running text. The morphological information stored in each entry has proved to be relevant and in general sufficient for correct segmentation. However, in order to increase accuracy in determining lexical class, as required by the syntactic analysis, it would be advantageous to expand the number of categories represented in the lexicon entries. As it was not clear before the present tests exactly which additional classification would be useful, we chose to start from a minimum and provide for easy future expansion. For example, the experiments confirm our assumption that it would be desirable to specify the potential junctures for a given stem and to differentiate several inflectional paradigms within a lexical class, in particular strong and weak verbs. These data would have resolved the ambiguities encountered for the sample words in Fig. 2.

- As the aims of our system do not include any attempt to incorporate semantics and as moreover there is no feedback from the syntactic component planned, a unique structural specification cannot be expected in the case of ambiguities requiring reference to these structural levels. Since such ambiguities do not necessarily lead to incorrect grammatical specification and only rarely to incorrect pronunciation, this is only a relative limitation.

Correctness of the phonemic transcription certainly accounts for a great part of the quality and acceptability of a text-to-speech system. Nevertheless it is often claimed (e.g. [6]) that synthetic speech should be evaluated along further dimensions, such as intelligibility, listening comprehension and naturalness. One goal of the approach presented here is to lay the ground for the incorporation of rules for the assignment and realization of stress and intonation patterns not only on the word but also on the sentence level. Thus the basic phonetic transcription will be extended and modified so as to give a representation closer to natural speech.

REFERENCES

[1] Kommenda, M.: "GRAPHON - ein System zur Sprachsynthese bei Texteingabe". In: H. Trost and J. Retti (Eds.), Österreichische Artificial Intelligence-Tagung. Springer, Berlin, 1985.

[2] Allen, J.: "Synthesis of Speech from Unrestricted Text". Proc. IEEE, vol. 64 (1976), pp. 433-442.

[3] Wolf, H.E.: "Sprachvollsynthese mit automatischer Transkription". Der Fernmelde-Ingenieur, vol. 38 / no. 10 (1984), pp. 1-42.

[4] Mangold, H.; Stall, D.S.: "Principles of Text-Controlled Speech Synthesis with Special Application to German". In: L. Bolc (Ed.), Speech Communication with Computers. Carl Hanser, München, 1978, pp. 139-181.

[5] DUDEN Ausspracheworterbuch. Bibl. Inst., Mannheim, 1974.

[6] Rühl, H.-W.: Sprachsynthese nach Regeln für unbeschränkten deutschen Text. Dissertation Ruhr-Universität Bochum, Germany, 1984.

[7] Müller, B.S.: "Regelgesteuerte Umsetzung von deutschen Texten in gesprochene Sprache für das Sprachausgabegerät VOTRAX". In: B.S. Müller (Ed.), Germanistische Linguistik, vol. 79-80 (1985), pp.83-112.

[8] Kästner, W.: Automatische Phonemisierung orthographischer Texte im Deutschen. Helmut Buske, Hamburg, 1972.

[9] Menzel, W.: " A Grapheme-to-Phoneme Transformation for German". Comp. & AI, vol.3, 1984, pp. 223-234.

[10] Philipp, M.: Phonologie des Deutschen. Kohlhammer, Stuttgart, 1974.

[11] Ortmann, W.: Wortbildung und Morphemstruktur eines deutschen Gebrauchswortschatzes. Goethe-Institut, München, 1983.

[12] Augst, G.: Lexikon zur Wortbildung. Forschungsberichte des IdS, vol. 24.1-4. Gunter Narr, Tübingen, 1975.

[13] Mater, E.: Rückläufiges Wörterbuch der deutschen Gegenwartssprache. Bibliographisches Institut Leipzig, 1983.

[14] Wahrig, G.: Deutsches Wörterbuch. Gütersloh, 1983.

[15] Meier, H.: Deutsche Sprachstatistik. Georg Olms, Hildesheim, 1978.