BMBF

Verbmobil
Verbundvorhaben

Morphology and Speech Technology

Harald Lüngen, Martina Pampel, Guido Drexel, Dafydd Gibbon, Frederek Althoff, Christoph Schillo

Universität Bielefeld

Vm

Report 169
September 1996

September 1996

Harald Lüngen, Martina Pampel,
Guido Drexel, Dafydd Gibbon,
Frederek Althoff, Christoph Schillo

Universität Bielefeld (UBI)
Fakultät für Linguistik und Literaturwissenschaft
Universitätsstr. 25
Postfach 10 01 31
33501 Bielefeld

Tel.: (0521) 106 - 3510
Fax: (0521) 106 - 6008
e-mail: `morphy@Spectrum.Uni-Bielefeld.DE`

Gehört zum Antragsabschnitt: 5.3 Morphologie
15.3 Interaktive Phonologische
Interpreations

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 B 2 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

Abstract

This paper describes a morphological component in a speech recognition architecture for German dealing with the recognition of compounds from their individual constituents. The specification of our morphological model allows for variation in functionality, e.g. the reconstruction of split compounds, of lexicalised, and of non-lexicalised (unknown) compounds. An implementation and evaluation results for split compounds are presented.[1]

[1] This paper was originally published in *Proceedings of the 2nd ACL-SIGPHON Workshop on Computational Phonology and Speech Technology*, University of California, Santa Cruz, 1996, pp. 25-30

1 Motivation and Goals

This paper deals with the design of a morphological component for a speech recognition system originally comprising a stochastic word recognition component plus a prosody component that provides word hypotheses graphs (WHGs, or word lattices according to the specification in [12]) as its output to a syntax/semantics component. The recognizer was designed for spontaneous continuous speech in appointment scheduling dialogues in the context of the VERBMOBIL speech-to-speech translation system [15], [4]. Our motivation for integrating a morphological component into it resulted from the observation that speech recognizers have certain difficulties dealing with complex German word forms, especially those morphologists call *compounds*. The model underlying our morphological component relates closely to the paradigm of finite state phonology and morphology (discussed in e.g. [10], [14], [13]).

Compounds are defined for present purposes as words which are built compositionally from other words that can occur as free forms on their own. Examples of German compounds are

- Arzttermin: Arzt, Termin
- Aprilwoche: April, Woche
- Dezemberhälfte: Dezember, Hälfte

1.1 Split Compounds

An interesting problem with recognizing compounds in speech is a phenomenon we call *split compounds*. Split compounds are compounds which often cannot be detected by speech recognition systems because their phonological material is torn apart by slips of the tongue, repetitions, pauses and/or other insertions e.g. *Mitarbeiter_<äh><P>_besprechung*. One important observation we made in this context is the following: *The splitting of compounds mostly takes place at morphological boundaries.* Although split compounds are not easily recognized by the stochastic word recognition components, the constituents of these compounds are often recognized and transmitted to a syntactic component via the WHG, provided they are listed in the lexica of the word recognizer. The syntactic component is not always able to treat these split compounds correctly, i.e. as one word.

1.2 The Size of Recognition Lexica

Beyond the problem of recognizing split compounds, the recognition of compounds in general demands large lexica.

- In principle, morphological composition rules in German permit the generation of an infinite set of compounds.
- Whereas the parts of e.g. English compounds are frequently spelled out separately, German compounds are almost always written as single orthographic words.

Current stochastic word recognizers depend on the definition of an orthographic word as they have access to lexica which are automatically generated from given orthographic transliterations using blanks as word separators. So, given a fixed corpus, the vocabulary covering it has to be much larger for German than it has to be for English, which reduces the speed of the stochastic word recognition [6]. Because of the infinite number of potential compounds, an exhaustive lexical listing is simply not feasible for German (and many other languages).

1.3 Unknown Word Recognition

This leads to the problem of recognizing unknown (out-of-vocabulary) words. The recognition of out-of-vocabulary items has been tested using phonological knowledge (actually phonotactic knowledge) about well-formed syllable structures [7] and [9]. These phonotactic constraints are useful for the recognition of simple, morphologically unstructured out-of-vocabulary items, but except in these contexts they do not represent strong enough constraints on their own to recognize morphologically complex unknown words. Since complex words consist of units which are members of a finite set of *morphs* or formatives, one can specify morphotactic rules, which operate on a finite morph lexicon to derive complex word forms. It is obvious, that the set of *actual* morphs - those, which are attested in a corpus and lexicalized in a morph lexicon - is only a subset of the set of *potential* morphs - predicted items which satisfy the phonotactic constraints of a language. Thus an integration of morphological constraints in a speech recognition system will lead to more specific constraints about unknown complex word forms.

1.4 Challenges for Speech Recognition Systems

The growing demands on speech recognition systems like

- large vocabulary word recognition

- few domain restrictions
- robustness
- unknown word recognition

have led to the idea of integrating morphological knowledge into the speech recognition process. Our morphological component is designed to achieve the following goals:

- Improvement of the word recognition rate through the reconstruction of split compounds.
- Reduction of word recognizer lexica through the recognition of lexicalized compounds from representations of their individual constituents.
- Expansion of the functionality of speech recognition systems by making a contribution to the recognition of unknown (non-lexicalized) words.

2 Online Morphology: Architecture and Interfaces

Morphological knowledge has already been used by speech technologists to improve word recognition systems, though mostly offline, that is, before and not during the speech recognition process, i.e. mostly in the making of system lexica [6] and lexical databases [5].
Our goal was to develop an online morphology which is active during the speech recognition process. To convince speech technologists of the usefulness of an online morphology we decided to integrate our morphological component *MORPHY* into the speech recognition architecture as an optional morphological filter over word lattices. The word hypotheses graph filter *MORPHY* performes compound reconstruction, located between the word recognition component and the syntactic component. *MORPHY* has the following features:

- The interfaces of the morphology correspond exactly to the existing interfaces between the stochastic word recognition component and the syntactic component, which is a WHG; consequently no specifications of associated components have to be changed.
- Operations on the WHG lead to the addition of new information by inserting new word hypotheses.

- No reduction of information is undertaken.

This conception permits easy comparison between recognition results with and without morphology (see section 6.2).

3 Compounding in Spontaneous Speech

3.1 Type and Token Frequencies of Compounds in German

The high frequency of occurrences of compounds in a VERBMOBIL subcorpus of spontaneous German speech (172672 tokens) and the associated lexical database (4514 types) is shown in Table 1.

Morphological Category	Type Freq.	Token Freq.
Compounds	36%	11%
Non-compounds	64%	89%

Table 1: Type and Token Frequencies of German Compounds

3.2 Analysis of Occurrences of Split Compounds in Transliterations of German Time Scheduling Dialogues

An analysis of occurrences of 57 split compounds in spontaneous speech (using the orthographic transliterations of 266 VERBMOBIL time scheduling dialogues) led to the following classification of possible insertions between constituents of binary compounds (7 cases are miscellaneous other types):

Category A: Inconspicuous paralinguistic events. Phonetically rather inconspicuous events include <P> (pause), <A> (breath taking), or <Z> (hesitation) (29 of 57)
Example: *April_<P>_hälfte*

Category B: Hesitation markers. Hesitation markers in German include *<ähm>, <äh>*, or *<hm>* (5 of 57)
Example: *Mitarbeiter_<äh><P>_besprechung*

Category C: Phonetically similar events. Phonetically similar insertions of the following kinds may occur:

(a) the second constituent as a whole

(b) the first syllable of the second constituent

(c) a sequence that is phonetically similar (especially the initial segment) to the first syllable of the second constituent, partly in combination with an item from the first two classes (16 of 57)
Examples: *Juli_+/*woche/+<hm>_wochen, ab_+/ser=/+_sagen, Termin_/ka=/+_kalender*

3.3 Empirical Analysis of Word Hypotheses Graphs

Subsequently, sections of the WHGs generated by the word recognizer were examined that corresponded to the time frames in which the split compounds were uttered. The single constituents of the compounds had almost always been recognized in one or more of their inflected forms, provided, of course, they were listed in the lexicon of the speech recognition system. The questions were:

- What was recognized in place of the different sorts of insertions (including morphophonologically ill-formed insertions)?

- Was there a connected path through the WHG linking the constituents of the compounds?

The results are as follows.

- Insertions of category A were generally mapped onto <*NIB*>, a WHG arc label meaning *non-interpretable frame.*

- Insertions of category B were mapped onto arcs with the corresponding labels <*ähm*>*,* <*äh*>*,* <*hm*>. In addition, they were mapped just as frequently on arcs with labels such as *er, ich, er, ja, ah, ab, wie*, which can be roughly characterized as phonetically unstressed, monosyllabic function words and discourse particles

- Insertions of category C which were rarely autonomous words, were for the most part mapped onto autonomous words (naturally, since the WHGs were the output of a *word* recognizer) whose first syllables were phonetically similar to the sequence of phonemes in the original insertion. For example, for the item *ka=* in *Termin_/ka=/+_kalender*, words such as *Kalbe, Koelln, koennten, halten, konnte* were recognized (see Figure 3).

A complication could be observed in cases where the insertion was phonetically very close to the second constituent of the compound: word recognition frequently recognized both constituents of the compound (and the whole compound as well in some cases) already in the frame covering only the first constituent and the insertion. In the frame covering the utterance of the second constituent proper, some other word, which was phonetically similar to the second constituent, was recognized so that it became impossible for the syntax to find a connected path through the WHG matching the whole utterance. Example: For *an_+/gemes<Z>=/+_gemessene* the word form *angemessene* was recognized in the frame covering only *an_+/gemes<Z>=/+*, whereas in the frame covering the item *_gemessene*, words such as *Messe, müßte, mußten, müssen, Essen, ersten, nächsten* were recognized (see Figure 4).

3.4 A Model of the Representation of Split Compounds in WHGs

The four Finite State Networks in the Figures 1, 2, 3 and 4 below model the occurrences of binary (two-constituent) split compounds as connected paths in WHGs according to the three cases described in the last section. Since the morphological component 'mends' the split compounds, a 'reconstruct'-function, taking the word forms related to the constituents of the compound as arguments, is indicated as well.

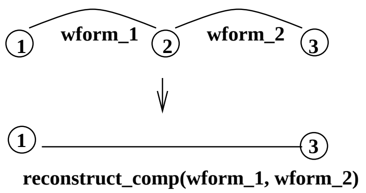

Figure 1: No insertion

The idea is, whenever a sequence of connected arcs in a WHG matches a path through one of the networks, the morphological component inserts a new arc into the WHG provided that the conditions formulated above are met. The items recognized are *wform* (word form), *insertwform*, *INSERT* (inserted forms). The regular expression

$$(\ wform_i\ (insertwform_i)\ INSERT^*\ )_{i=1}^{n}\ wform_{n+1}$$

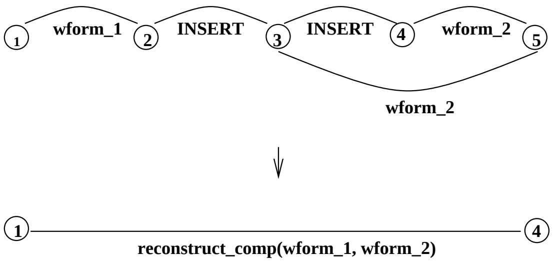

Figure 2: Insertion of category A or B

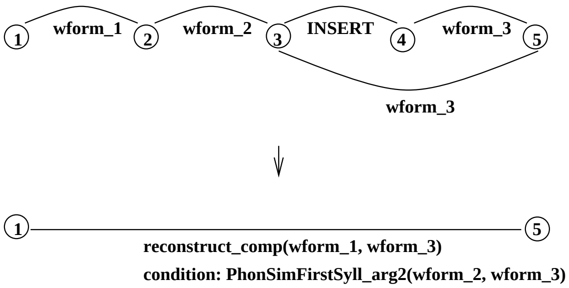
1
wform_1
2
wform_2
3
INSERT
4
wform_3
5
wform_3
1
5
reconstruct_comp(wform_1, wform_3)
condition: PhonSimFirstSyll_arg2(wform_2, wform_3)

Figure 3: Insertion of category C (3)

is a generalization of the four networks, defining no limit on the number of constituents and INSERTs; these can be limited by parametrizable constraints. The INSERTs are removed from the reconstructed compounds; however, they are retained in the WHG as they may be significant at higher levels of analysis.

4 A Model of a Morphological Component for Speech Recognition

The specification of the model has been kept as general as possible in order for the morphological component to allow for the variation in functionality expressed in 1.4.
The recognition of lexicalized compounds is regarded as an additional constraint on the more general function of recognizing non-lexicalized compounds by filtering through a lexicon and employing a suitable network.

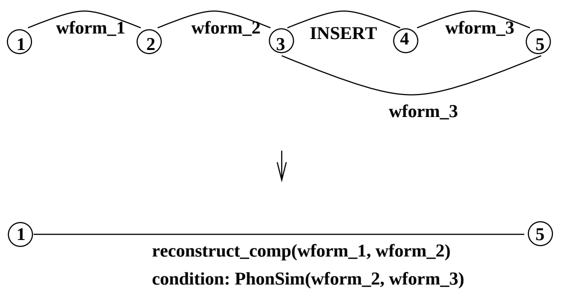
1
wform_1
2
wform_2
3
INSERT
4
wform_3
5
wform_3
1
5
reconstruct_comp(wform_1, wform_2)
condition: PhonSim(wform_2, wform_3)

Figure 4: Insertion of category C (4)

4.1 Modelling Composition in Finite State Networks

Since a WHG is a directed acyclic graph, a model of the representation of compounds in WHGs can always be represented in an Finite State Network, or, equivalently, as a regular expression. The task of a morphological lattice parser is then simply to find an intersection between two FSNs (see [10]). This task is highly general in comparison with the original goal of detecting split compounds; consequently, different functionalities of the morphological component can be realized simply by substituting different networks and ways of integrating morphology with neighbouring finite state system components can easily be imagined.

4.2 Network Design and the Role of Lexical Knowledge

Through the employment of an independent lexical knowledge source, the concept of strict matching of a path in the network describing the compositional morphology with a path in a WHG is relaxed for various purposes. The arc labels of the network can therefore represent different linguistic units, such as orthographic or phonological representations of words or morphemes, syntactic/semantic categories (such as *N, V, verbal particle* or *nominalization*), or morphological categories like *prefix, stem, root*, regardless of what kind of categories and representations are to be found in the WHG. During traversal of the network, the question of what value of an attribute of a lexical entry is to be matched with the label of a network arc is answered by a SELECT Parameter associated with the arc. In the current network designed for the reconstruction of split compounds, for example, stem forms of constituents of the compounds and the categories of the possible INSERTs are found at the arcs of the network:

Network:

```
0  arbeit  1 SELECT:orthstem
1  INSERT  2 SELECT:orthinfl
2  amt     3 SELECT:orthstem
```

The lexicon contains entries for fully inflected word forms of compounds as well as constituents and INSERT units. A lexical entry has different attributes for

Lexicon: Lexical entry for Arbeitsamt[2]:

```
orthinfl      = Arbeitsamt
orthstem      = Arbeitsamt
phoninfl      = ?'a6.baIt#+s#?''amt
phonstem      = ?'a6.baIt#+s#?''amt
conststems    = arbeit;amt
morphcat      = compound
```

The essential function of the lexicon in this application is to provide a mapping from the inflected forms in the WHGs onto stem forms and from there to the forms found in compounds:

WHG:

```
0 1 arbeiten     1.0 300  500
1 2 /hesitation/ 1.0 500  600
2 3 /pause/      1.0 600  800
3 4 Amt          1.0 800 1200
```

Moreover, additional phonological and prosodic constraints (such as phonological similarity of initial syllables) are defined, which are not available either in the WHG or in the network itself. Syntactic or semantic constraints can also be incorporated if they are encoded in the lexicon [1].

In procedural terms, the parser finds a lexical entry via the arc label in the WHG (an inflected form), tries to traverse the current arc in the network and finds the information which attribut value of the lexical entry is to be matched with the label of the arc of the network through the SELECT-Parameter associated with each network arc.

[2] *orthinfl* - orthographic inflected form; *orthstem* - orthographic stem form; *phoninfl* - morphoprosodically transcribed inflected form; *phonstem* - morphoprosodically transcribed stem form; *conststems* - associated stem forms, compounds only; *morphcat* - morphological category.

5 Implementation

The morphological component *MORPHY* is fully implemented for the functionality of reconstructing split compounds [2]. *MORPHY* was developed and tested under SOLARIS 2.4 and LinuX 1.3.56 and should be run on a SPARC Station with at least 48 MB main memory (the required amount of memory depends on the size of the network and wordgraphs) or a PC with 32 MB memory. The program is written in ANSI-C++. *MORPHY* supports inter-process communication via Amtrup's Intarc Communication Environment (ICE) v1.4 [3].

6 Evaluation

6.1 Software Evaluation

An initial evaluation with different networks was conducted with 1211 WHGs with an average size of 109 hypotheses and an average duration of 740 msec (see Table 2). Network 1 turned out to be the most satisfying one in runtime behaviour.

net- work	# nodes	# arcs	# compounds	runtime [msec]	realtime factor
0	1339	2405	1.505368	99.418	0.135
1	923	1585	1.437655	85.808	0.116
2	1633	2993	1.606936	91.452	0.123
2.hk	2266	1237	0.417011	86.136	0.116

Table 2: Statistics for the test runs. From left to right: Name of network, number of nodes and arcs for this network, average values of reconstructed split compounds per WHG, mean runtime and real time factor.

6.2 Quantitative Evaluation

Recognition performance of the component involves two aspects: first, standard overall recognition rate and accuracy evaluation; second, and more relevant, relative evaluation of the proportion of split compounds which were successfully reconstructed.

The effect of the morphological component on the word recognition rate was calculated by standard evaluation techniques [11] for 1211 WHGs, each containing one dialogue turn, and eight conditions. The eight parameter settings refer to the number of possible intervening INSERTs between constituents of compounds, classifications of reconstructible compound types, and admissible morpho-syntactic constraint relaxations on mapping inflected forms in the WHG to constituents of compounds. The table below shows a comparison between the results of the output of the word recognition module with and without morphological post-filtering using the parameters *Network 1* and *Permissible INSERT sequence length 1 or 2*.

Attributes	Recognition without Morphology	Recognition with Morphology
Word Recognition	63.0% (2753)	63.1% (2758)
Substitutions	27.0% (1182)	26.7% (1167)
Deletions	10.0% (437)	10.2% (447)
Insertions	4.6% (202)	4.5% (197)
Errors (sum)	41.7% (1821)	41.4% (1811)
Word Accuracy	58.35%	58.58%
Reference words	4372	4372
Hypotheses per word	4.710	4.776

All of the compounds in the speech data that were split at morphological boundaries and whose constituents were recognized by word recognition, could be automatically reconstructed by *MORPHY*. In the test WHGs, 1741 compounds were reconstructed. This relatively large number is on the one hand due to the fact that the stochastic word recognition very frequently finds potential INSERT units where actually no INSERTs were uttered, on the other hand they arise from the overgeneralization of the morphological reconstruction constraints.

7 Conclusion and Prospects

The modelling, implementation and evaluation of our morphological component *MORPHY* for speech recognition demonstrated that the task of reconstructing split compounds within an existing conventional architecture also led to artificial difficulties. The reconstruction of split compounds operates on a WHG generated by the word recognition component that only contains word forms which are in the word recognition lexicon.

However, the lexicon was originally designed for a stochastic word recognizer and not for the compound reconstruction task, and constituents of compounds were only included if they occured independently in the corpus. For example, whereas the words *Juliwoche*, *Juli*, and *Woche* appeared in the lexicon, only the word *Junihälfte* and not its constituents *Juni* and *Hälfte* were included in the list. Therefore it could not be ensured, that the set of compound constituents available to the morphology is a proper subset of the word recognizer vocabulary. As a result only those split compounds whose constituents are more or less accidentically known to the word recognizer are reconstructed.

Therefore we are investigating the use of lexica systematically extended by the constituents of compounds, which will also permit progress beyond the split compound reconstruction task towards a more phonologically and morphologically based compositional recognition of both lexicalized and non-lexicalized compounds on the basis of their constituents.

In respect to the second goal mentioned in section 1.4, we are working on the reduction of recognition lexica size. For example, for a given word list, a morphologically based reduction of the recogniton lexicon about 11% is feasible by splitting a predefined set of recognizable compounds into their parts. This reduction will increase in proportion to increases in corpus size.

Evaluation is currently in progress of the use of the morphological component with a stochastic word recognizer with morphologically fully analyzed vocabulary with respect to words of category “compound”. In the speech recognition process, these will be combined to compounds in the morphological component, and the results will be transmitted to the higher components (syntax/semantics). The testing and evaluation of this scheme is described in [8] and [1].

References

[1] Frederek Althoff, Guido Drexel, Harald Lüngen, Martina Pampel, Christoph Schillo. 1996. *The Treatment of Compounds in a morphological component for speech recognition.* Verbmobil Report (to appear), Universität Bielefeld.

[2] Frederek Althoff, Christoph Schillo. 1996. *Morphology Component: Morphy Manual Version 1.1.* Verbmobil Technisches Dokument 39, Universität Bielefeld.

[3] Jan W. Amtrup. 1995. *ICE–INTARC Communication Environment Users Guide and Reference Manual Version 1.3.* Verbmobil Technisches Dokument 14, Universität Hamburg.

[4] Doris Bleiching, Guido Drexel, Kerstin Fischer, Dafydd Gibbon, Harald Lüngen, Martina Pampel. 1995. *Morphologie im Forschungsprototypen.* Verbmobil Memo 83, Universität Bielefeld.

[5] Doris Bleiching, Guido Drexel and Dafydd Gibbon. 1996. *Ein Synkretismusmodell für die deutsche Morphologie* Verbmobil Technisches Dokument (to appear), Universität Bielefeld.

[6] Petra Geutner. 1995. "Using Morphology Towards Better Large-Vocabulary Speech Recognition Systems." In *Proceedings of the ICASSP-95.*

[7] Kai Hübener & Julie Carson-Berndsen. 1994. "Phoneme Recognition Using Acoustic Events." In: *Proceedings of the 3rd International Conference on Spoken Language Processing, Vol. 4.*

[8] Kai Hübener, Uwe Jost & Henrik Heine. 1996. "Speech Recognition for spontaneously spoken German dialogs". To appear in: *Proceedings of the ICSLP-96.*

[9] Andreas Jusek et al. 1994. "Detektion unbekannter Wörter mit Hilfe phonologischer Modelle." In *Mustererkennung 94, 16. DAGM-Symposium Wien*, Springer Verlag, Berlin.

[10] Ronald M. Kaplan & Martin Kay. 1994. "Regular Models of Phonological Rule Systems." In *Computational Linguistics*, 23:3.

[11] Michael Lehning. 1994. *Ein Programmsystem zur Evaluierung der signalnahen Spracherkennung.* Verbmobil Technisches Dokument 9, Technische Universität Braunschweig.

[12] Elmar Nöth & B. Plannerer. 1993. *Schnittstellendefinition für den Worthypothesengraphen.* Verbmobil Memo 2, Friedrich-Alexander-Universität Erlangen/Nürnberg.

[13] Graeme D. Ritchie et al. 1992. *Computational Morphology.* MIT Press, London.

[14] Richard Sproat. 1992. *Morphology and Computation.* MIT Press, Cambridge, Massachusetts.

[15] Wolfgang Wahlster et al. 1992. *Wissenschaftliche Ziele und Netzpläne für das VERBMOBIL-Projekt.* Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken.