BMBF

Verbmobil

Verbundvorhaben

The Treatment of Compounds in a Morphological Component for Speech Recognition

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

Universität Bielefeld

Vm

Report 170
September 1996

September 1996

Frederek Althoff, Guido Drexel, Harald Lüngen,
Martina Pampel, 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
15.6 Interaktive Phonologische Interpretation

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 system for German dealing with the construction of complex word form hypotheses out of a lattice of simplex forms. Our example is the recognition of compounds from their individual components. Evaluation results are presented for speech recognition with and without morphologically based word recognition.

Dieser Aufsatz beschreibt eine Morphologiekomponente in einem Spracherkennungssystem für das Deutsche, welche die Konstruktion von komplexen Worthypothesen aus einem Wörtergitter von Simplizia am Beispiel der Erkennung von Komposita aus ihren Einzelbestandteilen behandelt. Evaluationsergebnisse für morphologisch und nicht–morphologisch basierte Worterkennung werden vorgestellt.

1 Goals and motivation

This paper[1] proposes a strategy for partially satisfying the growing demands on speech recognition systems, e.g. large vocabulary recognition, few domain restrictions, robustness, and unknown word recognition by integrating morphological knowledge into the speech recognition process. Current stochastic word recognizers have, for example, certain difficulties with compound word forms. Compounds can be defined as words which are built compositionally from other words or stems of words that can occur as free forms. Examples of German compounds are *Arzttermin* (constituents: *Arzt, Termin*), *Arbeitsamt* (constituents: *Arbeit, Amt*), *Wochenendtermin* (constituents: *Woche, Ende, Termin*). Compounding is a frequent phenomenon in spontaneous speech: In the current VERBMOBIL transliteration corpus of 172672 wordform tokens and the related lexical database of 4514 wordform types, the token frequency of compounds is 11%, the type frequency amounts to 36%.

Both compounds and their individual constituents were included in the recognition dictionary, and most of the compounds as well as their individual constituents (but in almost all their possible inflected forms) occurred in the output lattice of the stochastic word recognition system (cf. Hübener et al., 1996). A dictionary of this kind is highly redundant; large dictionaries reduce the speed of the stochastic word recognition, and in view of the infinite number of potential out–of–vocabulary compounds, an exhaustive lexical listing is simply not feasible.

For the task of recognizing out–of–vocabulary words, the employment of phonotactic constraints on well–formed syllable structures has already been tested, see e.g. Jusek et al. (1994). Since complex words consist of units which are members of a finite set of *morphs*, it is also possible to specify morphotactic rules which operate on this finite morph lexicon to derive complex word forms. It is obvious that the set of *actual* morphs (those which are lexicalized in a morph lexicon) is only a subset of the set of *potential* morphs (those which satisfy the phonotactic constraints). Thus an integration of morphological knowledge leads to more specific constraints on out–of vocabulary complex word forms.

Occurrences of discontinuous ('split') word forms are a further problem in recognizing spontaneous speech. These often cannot be detected by speech recognition systems because their phonological material is torn apart by slips of the tongue, repetitions, pauses or other insertions. An analysis of split word forms in our corpus demonstrated that most are compounds split at morphological boundaries. Although split compounds are not easily recognized by stochastic

[1]This paper was originally published in Dafydd Gibbon (ed.): *Natural Language Processing and Speech Technology. Results of the 3rd KONVENS Conference. Bielefeld, October 1996*, pp. 71–76. Berlin, etc.: Mouton de Gruyter.

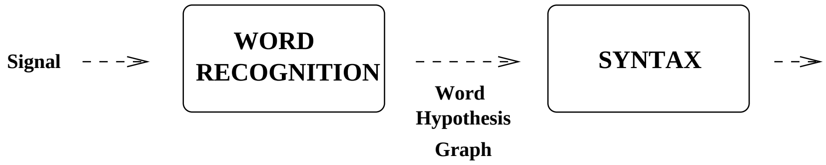

Figure 1: Standard incremental speech recognition architecture.

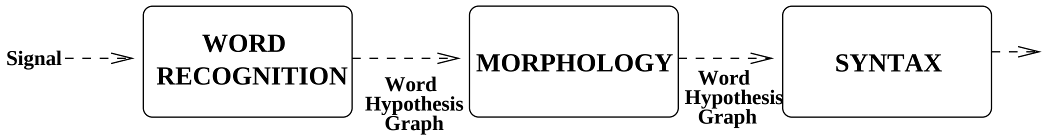

Figure 2: Standard incremental speech recognition architecture with morphology.

word recognition systems, their constituents are, and they can be recombined using morphological and phonological knowledge (cf. Lüngen et al., 1996).

Thus, our morphological component is designed to achieve the following goals:

1. To reduce of the size of the word recognizer dictionary through the recognition of lexicalized compounds from their individual constituents,
2. To prepare the ground for robust morphologically based recognition of out–of–vocabulary words.

2 Speech recognition with online morphology: Architecture and interfaces

In order to explore the use of morphological decomposition in the speech recognition process, two different architectures were tested. Figure 1 shows the speech recognition architecture without morphology, and Figure 2 presents the integration of our morphological component.[2]

The interfaces of the online morphology, word hypothesis graphs (WHGs), correspond exactly to the existing interface between the stochastic word recognition component and the syntactic component, no interface specifications of associated components had to be changed. Operations on the WHG in the morphological

[2]The experimental communication model for speech recognition used in the VERBMOBIL subproject 15 is INTARC, cf. Amtrup (1995).

component add new information by inserting new word hypotheses containing new compounds and confidence values, and the resulting WHG is transmitted to the higher components.

3 Reducing the recognition dictionary

Geutner (1995) mentions a degrading of the acoustic part of her word recognizer when using morph dictionaries containing affixes; this was predictable, however, since these are phonetically very small and often unstressed units. Though our morphological model (see Section 4) allows for the additional treatment of inflection and derivation, we have initially restricted our attention to morphological composition since word–sized linguistic units are involved.

The test vocabulary covered by the speech recognition system, a small wordlist of 470 wordforms, was reduced by 20% by splitting the 142 compounds into their constituents. This resulted in a list of 389 simplex words as potential compound constituents. This reduction rate will increase with increasing corpus size. The word recognition component was subsequently trained on the reduced dictionary.

4 Modelling

Finite–state automata have been established as adequate and efficient models for describing phenomena in the area of morphology (cf. Kay, 1987; Kaplan and Kay, 1994; Sproat, 1992). Our compositional morphotactics is encoded in a Finite State Network (FSN). Since a WHG is, in effect, an FSN, the task of a morphological lattice parser is simply to find an intersection between two FSNs (cf. Kaplan and Kay, 1994).

In the current network designed for the construction of compounds from their individual constituents, the arcs of the network are labelled with the stem forms of the compound constituents. We thus generalize over all possible inflected forms of one stem including those found in compounds (i.e. *Modifier+Interfix*).

The employment of an independent lexical knowledge source permits the requirement of strict string identity of a path label in the compositional morphology network with a path label in a WHG to be relaxed. Compare the following WHG extract containing inflected forms

```
<BEGIN> <END> <LABEL>    <SCORE> <FRAME_BEGIN> <FRAME_END>
   0      1    Termine    1.0      300           500
   1      4    Kalenders  1.0      800          1200
```

with a network extract containing stem forms

```
<BEGIN> <LABEL> <END> <SELECT>
   0    termin    1   OrthStem
   1    kalender  2   OrthStem.
```

The lexicon defines entries for fully inflected word forms of compounds as well as simplexes. Each lexical entry has different attributes for orthographic inflected form, associated stem forms, morphosyntactic category, and morphological category.[3] The entries for compounds contain an additional attribute <Constituents> whose value is a list of the stems of their constituents, which serve as the key to find the fully inflected form of a lexicalized compound, after a network path has been successfully traversed (see Table 1).

Table 1: Selected lexical information for *Terminkalenders*, *Termin*, and *Kalenders*.

<Orth>	<OrthStem>	<SynCat>	<MorphCat>	<Constituents>
Terminkalenders	terminkalender	N,gen,sg,mask	compound	termin,kalender
Kalenders	kalender	N,gen,sg,mask	simplex	
Termine	termin	N,nom,pl,mask	simplex	

In procedural terms, the parser finds a lexical entry via the arc label of the WHG (an inflected form), tries to traverse the current arc in the network and finds the information whose attribute value of the lexical entry is to be unified with the network arc label via the SELECT Parameter associated with each network arc (see the network extract above). Given this matching procedure, the network's arc labels may represent different linguistic units such as orthographic or phonological representations of words or morphs, syntactic and 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.

Lexical entries with semantic attributes are planned for future work in order to formulate semantic constraints on the composition of stems (see e.g. Fischer, 1993) in the network. The lexicon itself is a constraint which filters lexicalized compounds from the set of potential compounds.

[3] Phonological stem and inflected form are included as well, but not shown in the example. With such a lexicon, constraints about potential insertions between compound parts in spontaneous speech could be formulated (see Lüngen et al., 1996).

Table 2: Recognition rates with and without morphology.

Recognition	without Morphology	with Morphology
Word Recognition	92.4% (1223)	92.2% (1220)
Word Accuracy	92.29%	92.06 %
Hypotheses per word	346.076	349.494

5 Results and conclusions

An evaluation of the morphological component software has been performed with 103 WHGs, each containing one turn of a dialog. The runtime behaviour of the stochastic word recognition component was reduced by about 7% by employing the morphologically reduced dictionary. Furthermore, it was shown that the runtime behaviour of the morphology is linear, so that a parallelization of the two components using one processor for each will lead to an improvement of the runtime behavior of the whole morphologically based word recognition system shown in Figure 2.

In the recognition evaluation carried out with the 103 WHGs for the two architecture settings, the word recognition rate was calculated with standard evaluation software (cf. Lehning, 1994). The left column in Table 2 shows the rates for recognition without morphology (cf. Figure 1); the right column gives the rates for recognition with morphological decomposition of compounds in the dictionary and morphological postprocessing as shown in Figure 2 above.

It is remarkable that the word recognition and accuracy rates do not exhibit any statistically relevant differences, contrary to the claim of Geutner (1995), who suggests that smaller recognition units (e.g. simplexes as opposed to compounds) lead to a degrading of the acoustic recognition. Simplex words or lexical roots are acoustically stable enough to serve as recognition units. Also the number of hypotheses per word does not differ significantly.[4]

This suggests that the morphological decomposition of words will be an object worth pursuing in speech recognition. It is a pre–condition for the recognition of out–of vocabulary words on the basis of smaller meaningful constituents below word–level. We predict that among the bound morphs, i.e. affixes of the language, a subset can be defined that also fulfills the requirement of acoustic stability, and that our results can be generalized from the recognition of compounds to the recognition of derived words.

[4]The relatively high number of hypotheses per word with and without morphology is due to incremental processing.

References

J. W. Amtrup (1995). ICE–INTARC communication environment users guide and reference manual. Version 1.3. Verbmobil Technisches Dokument 14. University of Hamburg.

K. Fischer (1993). *Kompositionelle Semantik im Lexikon am Beispiel denominaler Komposita.* Master's thesis, Universität Bielefeld.

P. Geutner (1995). Using morphology towards better large-vocabulary speech recognition systems. In: *Proceedings of the ICASSP-95.*

K. Hübener, U. Jost and H. Heine (1996). Speech recognition for spontaneously spoken German dialogs. In: *Proceedings of the ICSLP-96.*

A. Jusek et al. (1994). Detektion unbekannter Wörter mit Hilfe phonologischer Modelle. In: *Mustererkennung 94, 16. DAGM-Symposium Wien*, Berlin. Springer Verlag.

R. M. Kaplan and M. Kay (1994). Regular models of phonological rule systems. *Computational Linguistics* 23:3.

M. Kay (1987). Nonconcatenative finite-state morphology. In: *Proceedings of the Third Conference of the European Chapter of the Association for Computational Linguistics*, Copenhagen, Denmark. University of Copenhagen.

M. Lehning (1994). Ein Programmsystem zur Evaluierung der signalnahen Spracherkennung. Verbmobil Technisches Dokument, Nr. 9. Technische Universität Braunschweig.

H. Lüngen, M. Pampel, G. Drexel, D. Gibbon, F. Althoff and C. Schillo (1996). Morphology and speech technology. In: *Proceedings of the ACL-SIGPHON Workshop on Computational Phonology in Speech Technology*, Santa Cruz, California. University of California, ACL.

R. Sproat (1992). *Morphology and Computation.* MIT Press, Cambridge, Massachusetts.