Phonological Processing of Speech Variants

Julie CARSON-BERNDSEN

Universität Bielefeld,
Fakultät für Linguistik und Literaturwissenschaft,
Postfach 8640,
D-4800 Bielefeld 1.

Abstract

This paper describes a strategy for the extension of the phonological lexicon in order that non-standard forms which arise in fast speech may be processed by a speech recognition system. By way of illustration, an outline of the phonological processing of standard wordforms by the phonological parser (PhoPa) is given and then the extension procedure which is based on this phonological parser is discussed. The lexicon extension procedure has two stages: phonotactic extension which involves the introduction of additional restrictions into the phonotactic network for the standard language in the form of metarules describing phonological processes, and specialised word model construction whereby for each standard phonemic wordform a verification net which contains all variants of this standard form is compiled. The complete system serves as a phonologically oriented lexicon development tool, and its theoretical interest lies in its contribution to the field of speech variant learning.

1. Introduction

This paper is concerned with a particular aspect of computational phonology, namely the processing of non-standard forms which may arise in fast speech. Since no native speaker of a language consistently adheres to a given standard pronunciation in normal conversation, it is an important attribute of any speech recognition system from a robustness point of view that it be able to process such non-standard forms. These non-standard forms will be referred to in this paper as speech variants. Speech variants are systematic, and may arise as a result of a phonological process (e.g. assimilation) or may express some dialect characteristic of the speaker. In the proposal presented below, the standard form is taken to be the phonemic representation in the lexicon and speech variants are taken to be systematically learned and lexicalised. It is shown that variants of a standard form may be learned on the basis of a metarule describing a particular phonological process. This avoids the necessity for a lookup-table of the variants belonging to particular forms since these are generated according to the restrictions given in the metarule. Thus, during analysis, both standard and non-standard forms may be processed.

2. Phonological Processing

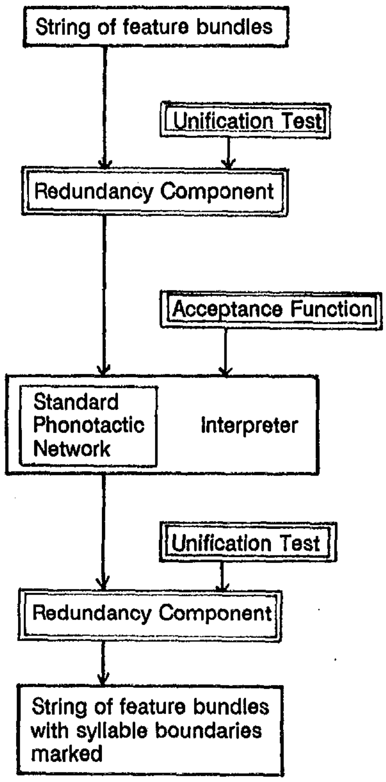
The phonological parser, PhoPa, described in Carson (1988) uses as its linguistic knowledge base a network representation of the phonotactics of a particular language in order to parse the phonetic sequence into phonemic syllables. The phonotactic network is feature-based and serves as a phonological word template consisting of nonreduced and reduced syllables.

Following Church (1987), allophonic information is considered to be important for distinguishing syllable boundaries and thus a canonicalisation step is necessary in order to filter out the variant information which is not relevant at the phonemic level. In PhoPa a feature-based transduction relation is responsible for translating between the allophonic and the phonemic domains. A transition label in the network consists of two feature bundles, an input transition bundle and an output transition bundle each containing C-features (Carson, 1988). C-features are those features which are needed to characterise sound classes which participate in a particular phonetic process.

Using a restricted form of unification, an acceptance function tests whether a particular allophonic input string of feature bundles may be processed. Since the input feature bundles may be underspecified, a redundancy component consisting of feature cooccurrence restrictions tests for feature-value consistency and attempts to optimise the the information in the feature bundles.

The processing strategy used in PhoPa is thus phonetic data driven or variant-invariant. With the core of PhoPa only wordforms in the standard language can be analysed.

Phonetic Level



Phonemic Level

Fig. 1: Structural Overview of PhoPa

For the purpose of the discussion which follows, it is important to note that the phonotactic transduction network used in PhoPa is in theory non-directional, that is to say, the transduction interpreter can be applied to either the allophonic (phonetic or variant) or the phonemic (invariant) level. Thus, a processor which uses the variant-invariant strategy performs analysis and a processor which uses the invariant-variant strategy performs synthesis. The synthesis process therefore generates variants of canonical phonemic forms. If we allow a standard phonemic form to have more than one variant then an exhaustive synthesis process would generate all possible variants of the standard wordform.

It is in fact the case that the network context provides exacty the defined search space which is necessary for the automatic speech variant learning extension and furthermore it allows for structurally based heuristics which reduce this search space during synthesis. After synthesis, speech variants are integrated into the lexicon for efficient later recognition; this is clearly expensive on storage, however.

3. *Classification of Speech Variants*

Speech variants can occur either as a result of phonological processes, for example elision, epenthesis or assimilation, or they can arise in line with a regional or dialectal sound change. Speech variants can, however, be classifed according to three abstract processes based on segments: deletion, insertion and substitution. Each of these abstract processes has a corresponding abstract rule type. A **deletion rule** deletes a whole segment and can also modify feature values in neighbouring segments. An **insertion rule** inserts a whole segment and a **substitution rule** is applied to single features in particular segments having the effect of substituting one segment for another. All rule types require a context consisting of directly neighbouring segments. However, this context can in some cases be empty.

Each of the abstract processes can only occur within a particular range of the syllable. On the basis of German data, the following are the "most probable" ranges for the three processes. Deletion and insertion occur only in the rhyme (peak and/or coda) of the syllable and substitution has the whole syllable as its range (i.e. it can occur in the onset or peak or coda). These facts allow for a structure-based heuristic which defines the application range for each process type and thus limits the search space required for the extension. Since the syllable structure is directly represented in the phonotactic network (see Fig. 3 below), the search through the network can be restricted to a particular sub-structure (onset, rhyme etc. - for both nonreduced and reduced syllables). Considering the syllable structure in terms of a tree, the heuristic defines the optimal starting point for the search and the search proceeds in a depth-first fashion through the syllable tree.

Thus, since the application range for an insertion rule is the rhyme, the optimal starting point for the search is the peak. If the search is unsuccessful in the peak then the coda is searched.

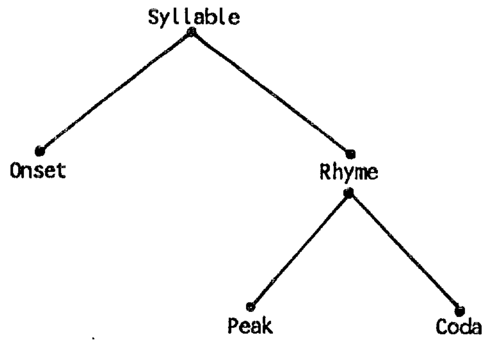


Fig. 2: Syllable Tree

Speech variants are described by declarative metarules which describe a particular phenomenon. The metarules have a left hand side and a right hand side, each of which consists of feature bundles, and they must belong to one of the abstract rule types mentioned above. Thus, epenthesis will be described by a metarule of insertion rule type. The epenthesis of a homorganic voiceless plosive between a nasal and a an apical fricative in German, for example, would be described by the following metarule:

$$\begin{bmatrix} +\text{nas} \end{bmatrix} \begin{bmatrix} +\text{cont} \\ +\text{strid} \\ +\text{cor} \\ +\text{ant} \end{bmatrix} \quad \text{->}$$

$$\begin{bmatrix} +\text{nas} \\ \alpha\ \text{ant} \\ \beta\ \text{cor} \end{bmatrix} \begin{bmatrix} -\text{cont} \\ -\text{strid} \\ -\text{son} \\ -\text{voice} \\ \alpha\ \text{ant} \\ \beta\ \text{cor} \end{bmatrix} \begin{bmatrix} +\text{cont} \\ +\text{strid} \\ +\text{cor} \\ +\text{ant} \end{bmatrix}$$

This caters for the forms [gAmps] for /gAms/, German: <Gams>; [gAnts] for /gAns/, German: <Gans>; [ge#zAg^ks] for /ge#zAg^s/, German: <Gesangs>.

As can be seen, it is possible to describe what would normally be thought of as two processes, namely an epenthesis process and an assimilation process, in terms of a single metarule.

4. The Extension Procedure

The extension procedure consists of two stages: phonotactic extension, whereby additional restrictions are added to the phonotactic network, and specialised word model construction which results in an extended phonological lexicon.

The phonotactic extension is concerned with the automatic extension of the phonotactic network by introducing new transitions. The extension procedure has two input components, the metarule and the lingistic knowledge base, namely the phonotactic network. Since in its original form the metarule cannot be directly incorporated into the phonotactic network, the extension procedure first applies a metarule interpreter which produces a graph representation for the metarule. The graph representation corresponds to the network representation exactly, that is to say that each transition in the graph consists of two feature bundles: an input specification and an output specification. The output transistion specification is always the phonemic form.

When the graph representation has been produced, a possible unification is sought for the network and the left hand side of the metarule. This involves a search through the network within the range defined by a heuristic. In the case of a substitution rule, for example, a unification would be sought first among the transitions of the onset and then among the transitions of the rhyme. In fact all possible unifications are sought within the application range of the process type, since a metarule may be underspecified and thus may apply in more than one fully-specified context.

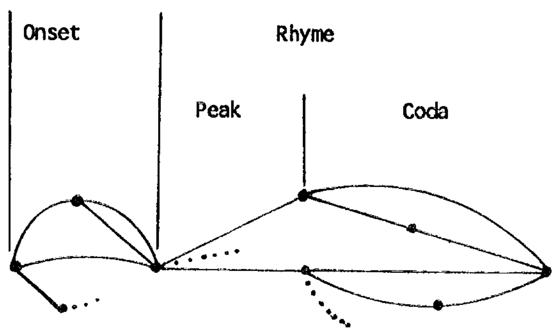


Fig. 3: Network Structure

If the left hand side of the metarule is unifiable with the network, the right hand side of the metarule is inserted into the network at the relevant place, by unifying the first and last states of the rule with the relevant states in the network. This phase of the extension procedure results in additional restrictions being added into the phonotactic network.

When the phonotactic extension is complete, specialised word models are compiled on the basis of the extended phonotactic network. The task of the compiler is to construct for each lexicon entry (i.e. standard phonemic form) a corresponding word model containing all variants of the entry. The compiler uses an invariant-variant processing strategy, and thus transduction takes place in reverse, that is to say, a translation between the phonemic and the allophonic domains. The compiler notes the paths which are consulted in the extended phonotactic network and produces a verification net for the input word. Word models are therefore subnets of the complete network. Thus, on the basis of a metarule describing a particular phenomenon, the phonological lexicon can learn new variants of a standard wordform.

5. Conclusion

The extension procedure and the phonological parser constitute part of the lexical component of a speech processing system. They provide the possibility of analysing some types of non-standard forms which arise in normal conversation. The extended phonological lexicon provides top-down information about the structure of the word. By adding appropriate metarules, the resulting extended network may be used for parsing non-standard forms by transducing them into a standard phonemic form.

Since the long-term aim of a speech recognition system is to be able to cope with an unlimited vocabulary and to be speaker-independent, being able to process such speech variants plays an increasingly important role in speech recognition research.

The phonological parser and the extension procedure have been implemented in Arity PROLOG V5.1.

The extension procedure described here was developed as part of the research project Phonological Rule Systems at the University of Bielefeld which was financed by the Research Institute of the Deutsche Bundespost.

6. Bibliography

Carson, J. (1988): Unification and Transduction in Computational Phonology. In: *Proceedings of the 12th International Conference on Computational Linguistics*, 106-111.

Carson, J.; Gibbon, D.; Knäpel, K. (1989): Interim Report 31.03.89 *Forschungsprojekt: Entwicklung phonologischer Regelsysteme und Untersuchungen zur Automatisierung der Regelerstellung für Zwecke der automatischen Spracherkennung*. Research Project financed by the Deutsche Bundespost.

Carson-Berndsen, J.; Gibbon, D.; Knäpel, K. (1989): Final Report 30.09.89 *Forschungsprojekt: Entwicklung phonologischer Regelsysteme und Untersuchungen zur Automatisierung der Regelerstellung für Zwecke der automatischen Spracherkennung.* Research Project financed by the Deutsche Bundespost.

Church, K. (1987): *Phonological parsing in speech recognition*. Boston: Kluwer Academic Publishers.

Gibbon, D. (1985): Prosodic Parsing in English and Akan. Paper given at the *21st International Conference on Contrastive Linguistics*. Blazejewko, Poland.