AUTOMATED PHONOTYPICAL TRANSCRIPTION THROUGH THE
GEPH PHONOLOGY EXPERT-SYSTEM

G. PERENNOU, M. DE CALMES, I. FERRANE, J. THONI
IRIT-CERFIA Laboratory UA 824 CNRS
118 route de Narbonne
31062 TOULOUSE Cedex FRANCE

ABSTRACT

Within the scope of the Man-Machine Communication PRC (Programme de Recherches Concertées) Speech topic, we have developed the GEPH System, which is meant to enhance BDLEX, the lexical database of both spoken and written French. GEPH is supposed to derive the permissible pronunciations of a given statement, from its underlying representation, as this is secured from both analysing the statement for syntax and consulting BDLEX for a phonological representation of words.

Within the scope of the ESPRIT SAM Project (Speech Assessment Methodologies), the semi-automation of the annotation tasks, applied to speech corpora, involves as a sub-task an automatic phonotypical transcription. In the context of an expert system in phonology, this makes the problem easier for only one typical pronunciation is supposed to be derived. Furthermore, this pronunciation is to remain abstract and will not require applying low-level rules. In our paper, we intend to give the configuration of the automatic transcription system worked out from GEPH and made to operate on compatible IBM-PC—the system is written in C language.

1. INTRODUCTION

In speech processing systems, the phonological representation of lexical items necessitates taking into account dialectal and idiolectal variants. These variants have much importance in the phonological component: for a given speaker, the rules do not apply with the same frequency, nor in the same contexts.

From this point of view, it is necessary to distinguish between abstract rules that relate the lexical phonological representations with the phonotypical transcriptions and the broad phonetic notations (see the current terminology of the GRECO-PRC Pôle Parole) and the phonological rules applied towards the end (see for example how they are defined by Morin in 6).

An phonological expert system in phonology must be able to use the knowledge of phonologists in different types of tasks. The simplest consist in verifying, as a phonologist would, whether a set of rules allows to predict the attested pronunciations from phonological representations; more complex tasks, but also more interesting for speech recognition, are those aiming at generating a set of pronunciations deriving, for a given dialect, from the given phonological representations.

The phonological expert system GEPH was developed with these tasks in mind. It is part of the project for the lexical database of written and spoken French, BDLEX, currently underway at the GRECO-PRC CHIM Pôle Parole.

The first goal of this project was to load into a computer coherent sets of phonological rules for 'standard French' and different regional variants (starting with the Marseille dialect).

The second goal was to give appropriate lexical and phonological components to the speech encoding and decoding systems. These components must reproduce the mechanisms described by the phonological rules while taking into account the requirements imposed by concrete situations. There was therefore no reason why the models developed by phonologists for purely linguistic reasons should be the best adapted for this type of interface.

Under these conditions, the interest in phonological expert systems such as GEPH resides in the construction of knowledge bases provided by expert phonologists. This knowledge can then be converted into appropriate models for speech recognition or synthesis.

A difficulty inherent to this approach resides in the fact that it differs from that of phonologists, that is to say: discovery and analysis of properties common to particular languages, properties which can be used in order to propose new theoretical models. In this search, it is common to focus on certain phonological aspects, so that it is not necessary to be exhaustive for each language. In order to build a phonological knowledge base, we are thus obliged to gather fragments from many different sources, which can bring about problems of coherency.

To our knowledge, the first work that fixed itself such an goal was that of Y.C. Morin in 1971, which the author presents as follows: "This report presents a description of the low-level phonology of Parisian French and its computer implementation on the phonological grammar tester described by Friedman and Morin".

The second one is that of Cohen and Mercer (1), deliberately oriented towards automatic speech recognition. The authors present their objectives thusly: "To deal with this problem in the speech-recognition system being constructed at the Thomas J. Watson Research Center of IBM, we have developed a phonological rule component which employs: 1- A lexicon of American English phonemic base forms; 2- A set of phonological rules to account statistically for phonemic and major variation resulting from idiolect, dialect, style and pace; and, 3- An algorithm for applying the rules to the base forms for the generation phonological variants".

This work, in advance with regards to the real needs of then existing recognition systems, has not to our knowledge been used in the systems under development at IBM.

Currently, the projects under study for dictation machine use large vocabularies (more than 100,000 words for IBM France's PARSYFAL system) and generally quite complex linguistic models, in which a phonological component really has its place. Projects such as GEPH and the other expert systems just mentioned are now becoming of real interest.

In this communication, we will briefly present the components of the project BDLEX that have been developed for converting the phonological mechanisms necessary for
predicting pronunciation from broad phonetic notation. We will then examine the incidence of dialectal and idiolectal variations on the different classes of rules used in our model.

2. STRUCTURE OF THE GENERATIVE PHONOLEGICAL COMPONENT

The generative phonological component must convert surface utterances, provided by the semantic-syntactic component, into articulatory phonetic structures. However, speech recognition relies more on the auditory phonological component, which is harder to define and which certainly cannot be reduced to a simple adaptation of the generative component. In particular, the auditory phonetic structures are not totally isomorphic to the articulatory phonetic structures, no more than they are to the various acoustic phonetic structures used by automatic speech processing systems (synthesis or recognition).

In this communication, we will discuss the relation between the surface utterance level and the more abstract level of phonotypical transcriptions as it has been defined in the ESPRIT project SAM (13). More precisely, we will use this term as meaning prediction of pronunciation expressed in broad notation. We will not discuss the problem of precise phonetic description which uses, for example, diacritical signs.

In order to discuss this problem, the generative approach is sufficient. The general architecture of the phonological component is given in Fig. 1.

### 2.1. Lexical representations provided by BDLEX

The goal of BDLEX is to provide different linguistic tools for teams working on text and speech processing. BDLEX contains 25,000 lexical items, from which 350,000 inflectional forms can be derived by means of a morphological inflexion component.

![Fig. 1. General organization of the phonological component](image)

The information associated with a lexical item concern:
- spelling,
- phonology, including an indication of syllabization,
- inflectional morphology (conjugation of verbs, inflexion of nouns and adjectives),
- derivational morphology,
- morpho-syntax,
- frequency of appearance in the texts.

Only information pertaining to the phonological component is described here. Complementary information is given in (7) and (8).

The field PHON_SYLL provides the phonological representation in syllables and in feet of the lexical item (ex: \s\sa;\ma\di/). Variants of double timbre vowels are distinguished only in cases of /s/, /a/, /e/ and /a/ in the context of a closed, accented syllable.

The status of consonantic segments is determined by a phonological diacritical sign, which can be one of the following:
- fixity : this is the default status; ex: the consonants in bec represented by be/c/
- latency : X is used if X is only pronounced when followed by a vowel, separated at most by o or # (ex: the final t in petit /pati/ t').
- mixity : X if X is stable before # or $ and otherwise latent (ex: huit */ui t'/ where * is used for the aspirate h, that is to say an h preventing the liaison).
- non-fixity : X if X' or X, that is if X is not stable.

To these four cases (absence of diacritics, , and $) we add three others to indicate interindividual variants in lexical representation:
- X' : never pronounced by some speakers, always pronounced by others (ex: ananas -> /ananas/).
- X' : stable or mixed according to the speaker (ex: cinq -> /s/ /k'/).
- X* : latent or mixed according to the speaker (ex: suspect -> /suspekt/ here it is the final consonant cluster kt which is latent or mixed).

These diacritics will be changed during the lexical selection by one of the three following rules, depending upon the dialect:
X' -> X or A (this last symbol signifying the null string)
X* -> X or X'
X+ -> X' or X

Idiolectal particularities will also be processed by an appropriate choice of the lexical representation.

Terminal phonemes are the ones subject to these variations. For this reason, in order to facilitate the representation, a field for the phonological final FPH was included in BDLEX. This field contains the final phonemes having a status different than that of a stable phoneme; moreover, in the preceding examples, it also contains the final schwa.

BDLEX accounts for the fact that items have one or more variants. The VAR field indicates the type of variant, which can be:
- orthographic (ex: acupuncture /acupuncture/)
- phonological (ex: ananas, six, ...)
- morphological (ex: scenarios / scenarii/ and others (ex: oukase /oukase/).

<table>
<thead>
<tr>
<th>Variants</th>
<th>G</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage sur 25 000 items</td>
<td>1.25</td>
<td>0.5</td>
<td>5.5</td>
</tr>
</tbody>
</table>

### 2.2 Phonological representations

Only part of the syntactic information is pertinent for the application of phonological rules. In particular, the simplified boundary marks generally suffice. We adopt those of F. Dell presented in (2):

- § boundary of a phonological phrase, as defined by Chomsky-Halle:
- # strong boundary between words where the liaison is not possible
- **# weak boundary between words, making the liaison obligatory
- + internal boundary in a word, separating two morphemes

If a liaison is optional, the boundary # or the boundary ## will be chosen according to whether it is pronounced or not; the symbol $ can be used to indicate the possibility of either.
According to these conventions, the following is the phonological representation of the beginning of the first sentence of the EUROBOM French corpus:

```
§ votre  * nymc/ţ* de*apel* $ sk8lasl*t*  swo8ši8 siţ * ketravE siţ * sk8šia sek $ ...
```

2.3. The phonological expert system GEPH

With these phonological rules in its knowledge base, the expert system GEPH (for Générateur Phonologique) converts the utterances to the underlying form in order to determine the surface phonetic representations of pronunciation variants.

A set of phonological rules describing "standard French" was adapted by F. Dell and M. Lézat (2) (9) (10). M. Lambert and M. Rossi (4) (11) provided another set of rules describing Marseille French. These rules were integrated in the first version of GEPH on a DPS 8 (5) (3). A new version has been implemented on an IBM PC clone. This set must be completed by rules giving a finer description of the pronunciation.

2.4. Rule-conversion in GEPH

The phonological representations passed on to the phonological component are converted in GEPH as a list of vectors U(n), 1 ≤ n ≤ N, each one representing a unit of rank n in the string. The components of the vectors are as follows:
- the identifier ph(n)
- the features ti(n), 1 ≤ T,
- the class,
- the status,
- the position in the syllable
- the position of the syllable in the word.

In order to come as close as possible to the classic phonological formulation, we used relative time marks instead of absolute ones: thus U(1) is the unit under the read head, U(-1) the preceding unit, U(1) the following unit, etc.

2.5. Language for expressing the rules LER

The rules are expressed in the form

```
if condition [likelihood] then ACTION
```

**Condition** — This condition is expressed by means of a predicate or a conjunction (the logical operator AND) of predicates, placed between parentheses and separated by commas. A compound predicate is presented as a disjunction of conditions, separated by commas and placed between curly braces — i.e. { }. For example: {A,(B,{C,D,E,F}),G}). Through this recursive definition, we can express any formula of elementary predicate calculus.

Most of the primitive predicates can be obtained in the form c(n)x, where c is a component of U and x is a constant. It is equally possible to compare two positions in the string by means of the formulae c(n)=c(m) or c(n)≠c(m).

Besides this general formulation, a set of primitive predicates allows an abbreviated form for the conditions which brings them as close as possible to those used by phonologists:
- predicates (...) positive features: ti(n)=+ will simply be written t(n), ex: cons(n).
- predicates for negative features: ti(n)=− will be written -t(n), ex: -cons(n).
- predicates for classes: if X is a class, X(n) and -X(n) are abbreviations for class(n)=X and class(n)≠X respectively.

**Likelihood** — When the condition for the application of a rule is satisfied (when it is true) the application of an action can be obligatory or optional. In this last case, the likelihood must indicate the frequency of application of the rule in the idiolect or in the dialect.

**Action** — The basic formulation is the assignment c(n)x, meaning that the new value of the component c in position n is x. GEPH also proposes redefined actions for greater readability:
- statut(n) = fix is equivalent to FIX(n)
- statut(n) = mix is equivalent to MIX(n)
- U(n) = Λ is equivalent to TRONC(n)

In this last case, it should be remarked that U can be replaced by an identifier which must determine it entirely, except for the status. The default status is fixed, which is the status of the epenthetic segments that are known to us (ex: epenthesis of the yod in «plié» /pl unnatural/). A conjunct of actions is expressed as a list of actions separated by commas. Here is an example of a rule: stabilizing a latent consonant before an affix whose initial is not a consonant. This rule is formulated first by using the usual phonological conventions, then with the LER (a more precise formulation is given in §3).

```
FIX(obl)
C* —> C / _____ + -C
if lat(0), cons(0), unite(1)=+, -cons(2) [1]
then FIX(0).
```

One can notice that the read head is positioned at C*. If the condition is satisfied, it must change the status of the unit that is under the read head: the consonant becomes stable. This result is obtained by means of the action FIX(0).

3. RULE BASE STRUCTURE

In GEPH the rules are grouped in blocks of rules. These blocks are strictly ordered. Each block takes as input utterance the output utterance of the preceding block. The input of the first block is the phonological representation. The output of the last block is the phonetic representation(s).

In a block, the rules are mutually exclusive in the sense that as soon as a rule is applicable to a position of the input utterance, the following rules in the block can no longer apply to that position. Fig. 2 shows the structure retained. The groups of rules that vary for an idiolect or a dialect are marked with an asterisk.

Tables 1 and 2 show with the generic example of the adjective "bon" how the idiolect comes into play in nasalization phenomena in French. The rules used are the following:

```
(PIX) [syll,fix] —> [fix]
(NAS-1)* V —> NAS(V) / ___,+ t-cons
(LIA) [syll,fix] —> [fix] / ___,# t-cons
(NAS-2) V —> NAS(V) / ___,N*
(id. NAS-1)
(TRUNC) [syll,fix] —> Λ
```

Indeed, there are essentially two kinds of speakers. Only the pronunciation of the third utterance opposes them. We can equally see that in the rule base proposed here, it suffices to make NAS-1 unavailable in order to obtain the pronunciation realized by the first type of speakers and on the contrary to make it mandatory in order to obtain the one pronounced by the second kind. In order to take into account random behavior in a given population, one needs only to assign the appropriate likelihood to NAS-1.
4. CONCLUSION

BDLEX and GEPH are tools allowing the expert phonologist to model a dialect or an idiolect. Using a rule base developed in collaboration with F. Dell and M. Plénat, we were able to effectuate experiences on a corpus of utterances from phonological documents.

The ESPRIT project SAM offered another experimental ground: automatic phonotypical transcriptions and their comparison with the actual pronunciation of a group of four speakers. Phonotypical transcription is considered here in the sense of a prediction of the pronunciation (in broad transcription) conforming to a given (dialectal) norm.

GEPH allowed to obtain automatically at the end of the syllable organization block five phonotypical transcriptions:

- standard transcription, adopted by SAM for French
- a transcription characterizing each speaker (predictions that come closer to the actual pronunciation).

In collaboration with the Institut de phonétique d’Aix-en-Provence, we have also begun an experiment aiming at the construction of a rule base describing the pronunciation of Marseille French. The major question here is whether these diverse French dialects can be described with a common lexicon and rule base, the differences simply being expressed as a change in the likelihood of the rules.

The phonological model underlying the current rule base is generative. GEPH is designed to accommodate other types of models, in particular multilinear ones where the syllabification plays a more important part. Other models of rules bases can be used as soon as we dispose of systematic descriptions of French using these new theories.

**BIBLIOGRAPHY**


(9) M. Plénat, F. Dell, *Consomes terminales en français*, doc. de travail GRECO (43 p. dact.)

