LEARNING OF A PHONOLOGICAL COMPONENT FROM BREF CORPUS

A. Mailland, M. de Calmès, G. Pérennou
Institut de Recherche en Informatique, Toulouse, France

ABSTRACT

We introduce a phonological component model based on contextual phonological groups (cpg's) and multi-pronunciation groups (mpg's). The first ones are word substrings having several pronunciations depending on the context. The second ones are HMM like model of pronunciation in a given context.

The purpose of this paper is both to describe the current version of this phonological component and to present the results obtained from BREF Corpus.

INTRODUCTION

Recent developments of speech recognition have proved that taking in account phonological information improve speech recognizers - see for exemple [1], [2].

This is a point that the authors of BDLEX project had in mind [3]. Then, we have developed a phonological model compatible with HMM modeling [4].

We present here the method used in the model for learning, the parameters of this model from corpora and we give the results obtained from BREF corpus.

THE MHAT PHONOLOGICAL MODEL

The general model – the MHAT model (Markovian Harmonic Adaptation and Transduction) is described in [4]. We use here a particular model where the lexicon contains the phonological representation of inflected words.

The figure 1 shows the structure of the model.

Syntactic level S

The representations are the surface forms which are generated by the grammar. They consist of word-class strings. (S,S) transformations insert word boundaries : # for required liaison, ≠ for optional liaison and | for prohibited liaison. These boundaries depend on the syntactic stucture of the sentence.

These boudaries play an important role in French phonology. In [5], we have proved that a markovia biclass model is a good approximation of the (S,S) transformation (a ideal model must include prosodical features).

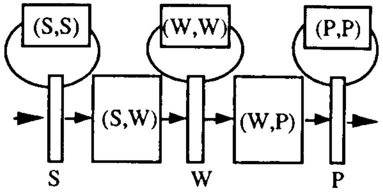

Figure 1 - MHAT model.

W-level

The representations are strings of both subword units and word boundaries #, ≠ or | . In our model, the subword units are contextual phonological groups (the cpg's) defined as strings of interdependent phonemes having context-dependent realisations. For example the word «grandes» (big) has, in our model, the W-representation gRɑ̃<˜də> instead of the standard representation /gRɑ̃də/. The first three units are the same in the two representations (they are called trivial units in our model) but the last one in the W-representation covers two phonemes which have interdependent and context-dependent realisations.

An harmonic transformation (in the sense given in Golsmith [6] , see also [4]) adapts W-representations by rewriting all cpg's into context-independent phonological units and deleting all boundaries. The result is a new W-representation, called here phonotypical representation, which consists of a string of multi-pronunciation groups (the mpg's units). For example, the W-representation of «grande fenêtre» is gRɑ̃<˜də> # <fə>nɛ<trə> which becomes the W'-representation gRɑ̃(˜də) fŒnɛ(trə>) where three units have multiple pronunciations depending on the speaker,

the style ... but not of the context. The first one is (˜də) which can be pronounced [dŒ] or [n]. The second one is (trə>) which can be pronounced [tRŒ] or [tR] or [tR]. The third one is Œ which have three possible realisations : [ø], [ə̲], [œ]. The other units have only one pronunciation and are considered as trivial mpg's. In order to represent cpg's and mpg's, we adopte the following conventions :

- if the phonetic substring x has one and only one pronunciation, their phonological and phonotypical representations are x (x is a trivial cpg and a trivial mpg).

- if x is constituted by interdependent phonemes (that is the phonetic realization of one among them which is dependent on the phonetic realization of the others) then :

- <x> is a cpg
- (x) is a mpg

If <x> depends only on its context within the word, it's a trivial cpg which will not be affected when it will be inserted in a sentence. Then, the notation (x) is used instead of <x>.

More details are given in previous papers [7], [8], [9]. Probabilities can be assigned to phonetical rules. Thus, the phonetical rule for (˜də) which associates two realisations : dŒ and n is represented by :

(˜də) —> dŒ (0.4) | n (0.6)

In this way, mpg's can be seen as hidden Markov model subword units.

Phonetic level

At this level the P-representations consist of strings of phonetic units in a given alphabet (here the IPA of the standard transcription of French). Thus, the last example can have several P-representations : [gRɑ̃dœ fŒnɛtR] [gRɑ̃n fŒnɛtRœ] , [gRɑ̃n fŒnɛt] ...

Internal adaptations occur also at this level for taking into account coarticulation effects. This will not be discussed here.

Here, the model used is simplified. It takes into account only end word mpg's describing phonological phenomena such as liaison. Few not trivial internal mpg's are conserved for foreign origin words.

GENPHON system

The purpose of GENPHON is to transform an orthographic form into a phonotypical one. This single transcription takes all the various possible pronunciations into account. It consists of three units :

- a lexical access module which permits to generate the phonological form with syntactic categories from orthographic form. A phonological form is composed of cpg's and mpg's.
- a module for positioning the phonological boundaries.
- a phonological module which yields the phonotypical transcription by using phonological rules.

For each word of a sentence, the phonological representation and syntactic class are retrieved from a BDLEX-derived lexicon [10].

Boundaries are positioned according to the bigram model introduced in [5] based on the work of P. Delattre [11].

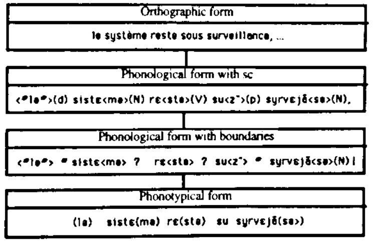
Orthographic form
le système reste sous surveillance, ...
Phonological form with sc
<#le#>(d) sistɛ<mə>(N) rɛ<stə>(V) su<z˜>(p) syrvɛjɑ̃<sə>(N),
Phonological form with boundaries
<#le#> # sistɛ<mə> ? rɛ<stə> ? su<z˜> # syrvɛjɑ̃<sə>(N) |
Phonotypical form
(le) sistɛ(mə) rɛ(stə) su syrvɛjɑ̃(sə>)

Figure 2 : Example of phonotypical generation.

In order to generate the phonotypical form of the utterance from the phonological form with cpg's, we use a set of phonological rules [5]. For every cpg's class, this rulebase provides a mpg for a given context.

Figure 2 shows an example of phonotypical form generation.

THE LEARNING SYSTEM

The proposed learning system (cf. fig.3) uses a variant of the alignment tool VERIPHON [12], developed at IRIT. It affords both advantages of being specially well adapted to align mpg like groups and of supplying statistics about the pronunciation variants observed within the corpus. As input, it takes a phonotypical transcription stemmed from the phonological component and its corresponding phonetic transcription proceeded from speech corpora. The system yields the aligned utterance represented by a string of mpg/phonemes couples.

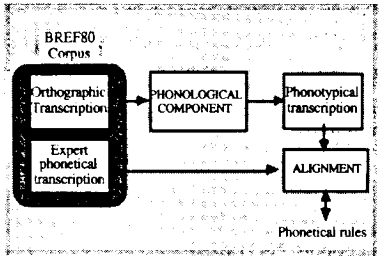
BREF80
Corpus
Orthographic Transcription
Expert phonetical transcription
PHONOLOGICAL COMPONENT
Phonotypical transcription
ALIGNMENT
Phonetical rules

Figure 3 : Learning system.

RESULTS

GENPHON and the learning system have been experienced on the BREF80 Corpus [13]. The results have some phonetic implication and allow us to specify the impact of the phonological alterations on the automatic speech recognition.

BREF80 has been transcribed in two steps :
1) a phonetic transcription is yielded by GRAPHON, the grapheme-to-phoneme conversion system of LIMSI,
2) this transcription has been rectified by experts.

It is composed of 5323 sentences pronounced by 80 speakers in a dictation style.

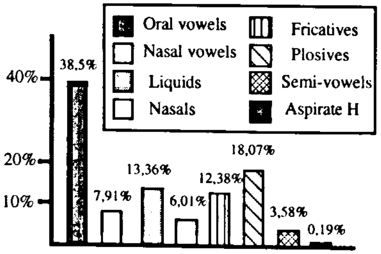

Figure 4 : Trivial mpg distribution.

BREF80 includes 345 000 occurences of mpg. This means 387 distinct mpg's including 40 trivial mpg's. These trivial mpg's account for over 90 % of the occurences.

Most of trivial mpg's is represented by standard phonemes. Their distribution is given figure 4. Nearly 50% are vowels.

For our purpose, the most important phenomena occur in the non-trivial ending mpg's. Figure 5 shows their distribution by phonetic class.

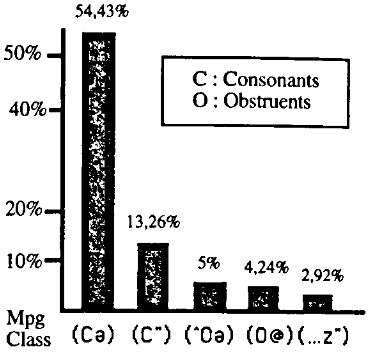

Figure 5 : Non-trivial mpg distribution.

The maximum for the (Cə) class is certainly due to numerous monosyllabic pronouns, articles, ...

(C") represents a latent consonant, for example in the word *encombrant* /ãkɔ̃brã(t")/.

(ˆOə) and (O@) are respectively in the words *poursuite* /pursɥi(ˆtə)/ and *tarif* /tari(f@)/.

(...z") represents the end of a plural word, for example in *nouvelles* /nuvɛ(ləz")/.

Here, we can only present two results about important phonological phenomena of french : the liaison and the schwa elision. More exhaustive results are given in [9].

Frequency of liaison realization depends on numerous factors [11]. Using this model, the study of the liaison processing has been made in [4].

We give, figure 6, liaison frequencies for some final mpg's in a plural word.

Mpg Class	Liaison	Non-liaison
(Wəz")	23,91%	76,09%
(Bəz")	12%	88%
(ˆQəz")	12,04%	87,96%
(ˆQLəz")	24,32%	75,68%

Figure 6 : Liaison after a plural word where W : set of liquids and nasals, B : set of voiced obstruents, Q : set of unvoiced obstruents and L : set of liquids.

Non-liaison is always more frequent. The more frequent liaisons produce in the (ˆQLəz") class (in this case schwa is pronounced). The average of liaison frequency is 18,06% for these classes.

We give, figure 7, schwa elision frequencies on Consonant-Schwa mpg's. These results shows that the kind of consonants plays a part in the elision phenomenon.

Mpg	Realizations			
(Lə)	LŒ	42,69	L	57,31
(Nə)	NŒ	24,36	N	75,64
(Fə)	FŒ	38,28	F	61,72
(Pə)	PŒ	76,23	P	23,77

Figure 7 : Frequencies per cent of schwa elision where L : set of liquids, N : set of nasals, F : set of fricatives and P : set of plosives.

These results illustrate the fact that frequencies of phonological phenomena are very variable according to mpg's in which they appear. This implies that a such component must be train from large corpora.

CONCLUSION

Using the phonological model based on contextual phonological groups (cpg's) and multi-pronunciation groups (mpg's), we have showed that it was possible to learn automatically pronunciation likelihoods associed to mpg's. This learning has been made thanks to a phonological lexicon where words are represented in cpg's, and two bases of rules. The first one defines cpg's. The second one describes mpg's and is learned from a transcribed corpus.

Achieved results concern a speaker population in a given communication situation : text reading.. Such results show a learning ability for a task of oral man-machine communication. They allows to show the impact of phonological variations in oral production and better to place the role of the phonological component in speech recognition systems.

Such a phonological component is compatible with speech recognition systems based on HMM.

REFERENCES

[1] J.-L. Gauvain, L.F. Lamel, G. Adda, M. Adda-Decker, "Speaker-Independent Continuous Speech Dictation", Eurospeech93, pp. 125-128.

[2] E.P. Giachin, A.E. Rosenberg & C.-H. Lee, "Word Juncture Modeling Using Phonological Rules for HMM-Based Continuous Speech Recognition", Computer Speech and Language,pp.155-168 (1991).

[3] G. Pérennou, "BDLEX : A Data And Cognition Base Of Spoken French", Proceedings of ICASSP, Tokyo, pp. 325-328 (1986).

[4] G. Pérennou, "Phonological Component in Automatic Speech Recognition. The case of Liaison Processing", Levels in Speech Communication - Relations & Interactions, pp. 211-223, 1995.

[5] G. Pérennou, "Introduction aux groupes à prononciations multiples suivi d'un sous-ensemble phonologique du français", IRIT report 94-07-R, 1994.

[6] J. Goldsmith, "Autosegmental & Metrical Phonology", Basil Blackwell, 1990.

[7] A. de Ginestel-Mailland, G. Pérennou, M. de Calmès, "Une approche de la phonologie en reconnaissance de la parole", Interface des mondes réels et virtuels, Montpellier, 22-26 Mars 1993, pp. 345-354.

[8] A. de Ginestel-Mailland, M. de Calmès, G. Pérennou, "Multi-Level Transcription of Speech Corpora from Orthographic Forms", Eurospeech93, pp. 1441-1444.

[9] A. Mailland, M. de Calmès, G. Pérennou, "Transcription multi-niveau d'un corpus de parole", JEP94, pp. 309-313.

[10] G. Pérennou, D. Cotto, M. de Calmès, I. Ferrané, J.-M. Pecatte, "Le projet BDLEX de base de données lexicales du français ecrit et parlé", Séminaire Lexique, Toulouse, 21-22 Janvier 1992, pp. 153-171.

[11] P. Delattre, "Studies in French and comparative phonetics", Mouton & Co. The Hague, 1966.

[12] G. Pérennou, H. Kabré, M. de Calmès, J.M. Pécatte, N.Vigouroux, "Une approche de l'étiquetage automatique indépendant du locuteur", Séminaire variabilité du locuteur, Avignon, 20-21 Juin 1989, pp. 61-67.

[13] L Lamel, JL Gauvain, M Eskénazi, "BREF, a Large Vocabulary Spoken Corpus for French", Eurospeech91, Genova, 24-26 Sep 1991. pp. 505-508.