ESCA Workshop on Phonetics
and Phonology of Speaking Styles
Barcelona, Spain
September 30 - October 2, 1991

A Formal Tool for Modelling "Standard" Phonetic Variations

Eric Laporte
Centre d'études et de recherches en informatique linguistique (CERIL)
25, cours Blaise-Pascal
F-91000 Evry
France

Abstract

Speech processing requires that some more phonetic variability than in current systems should be taken into account. CERIL has undertaken the description of a category of variability. In this paper, we discuss the limits of this description in terms of its extent (which styles are retained and which are discarded?) and of its precision (which minimal difference between two variants can be taken into account?). Then we present the formal and technical means used to model phonetic variations: a phonemic dictionary and local finite-state automata.

***

Utterances recorded for research in the field of speech processing are generally read in an elaborate, careful, slow style, which is used only, at least in the case of French, for reading poetry aloud or for making public announcements in noisy places. However, the conditions of use of some speech-processing systems would require other styles of speech to be taken into account. Therefore, the interest taken in careful reading style is now becoming less exclusive.

In connection with this evolution, there is a traditional view of a "standard" pronunciation among those who handle phonetic strings: each word would have a standard phonetic form, and could have many non-standard or reduced forms. This standard pronunciation is generally that of careful reading style, but sometimes they differ. For example, in words like *louer,* traditional transcription of French, as it is supported notably by dictionaries, systematically opts for synaeresis (`[lwe]`) and discards diaeresis (`[lue]`): the more reduced form is thus considered standard, though the longer form is the one preferred in careful reading style... The notion of non-standard forms is therefore close, but not equivalent, to the more objective notion of reduced forms. Actually, isolating one of the elements out of a set of variants is rather artificial and is often the result of a normative tradition: we doubt that such a simplified notion of standard is of any theoretical or practical relevance to phonetic variability.

1. The extent of the description

As soon as one finds it necessary to deal with some more phonetic variability than in careful reading style, an important and interesting issue is: where to stop? Marginal styles unavoidably appear in spontaneous speech: very reduced or very hurried speech, stammering, speaking while laughing, speaking a foreign language, speaking with food in one's mouth, interruptions in words, etc. Because of marginal styles and situations, the methods used with reading-style recordings cannot be re-used as they are with spontaneous-speech recordings.

Our hypothesis is that some variability can be described systematically, provided such marginal styles are discarded. In other words, we think it possible to select a "standard" level of variability, namely that of spontaneous speech on spoken media or in good conditions. This set of styles includes speech variants that do not belong to any "standard" pronunciation in the usual sense (e.g. J. Carson-Berndsen, 1990). In French, it is characterized by a number of types of frequent phonetic variations (cf. E. Laporte, 1989):

- deletion vs. non-deletion of final [ʁ] after a stop or fricative and before a consonant, as in *paraître tard,* transcribed [parɛttar] or [parɛtʁøtar];
- nasalization vs. non-nasalization of a stop after a nasal vowel and before an obstruent, as in *lampe torche,* transcribed [lɑ̃ptɔrʃ] or [lɑ̃mtɔrʃ];
- synaeresis vs. diaeresis of [i], [u], [y] before a vowel, as in *louer,* transcribed [lwe] or [lue];
- a significant variation in the aperture of unstressed [e], [ø], [o], as in *fêler,* transcribed [fele] or [fɛle];
- intervocalic consonant vs. geminate, as in *illicite,* transcribed [ilisit] or [illisit];
- etc.

We have also included in our description variations which do not belong to such a general type but are restricted to a word or a small list of words, e.g. [s] vs. [z] in *abasourdir,* transcribed [abasurdir] or [abazurdir]. We have included some geographical or social variants, but discarded those that seem too specifically regional, obsolete, or marginal in a way or another: e.g. southern French and Québec French have not been taken into account, because we do not want to describe a heterogeneous mixture of dialects, which would be better described separately. We discarded even major reductions when they are felt as abnormal: e.g. *immédiatement* [imedjatmɑ̃] may happen to be reduced to [imedjamɑ̃], but such a reduction is not ordinary and has no conventional or idiomatic status in any style.

This level of variability is not typical to a particular spoken or written corpus or to the speech habits of a particular speaker or small group of speakers, but of a linguistic community (E. Laporte, 1990). It is accessible to phoneticians through direct listening. The selection and delimitation of such a level of

variability cannot be based on entirely objective grounds, it can only be (sometimes arbitrarily) estimated by native linguists from the community.

2. The precision of the description

Another limitation of the description of variability comes from the fact that major phonetic variations cannot be put on the same plane as the slightest acoustic details, and cannot be represented with the same tools. To clarify this limit, we use strings of phonetic symbols, which are more readable than strings of feature bundles, and we take the International Phonetic Alphabet as a standard of precision in phonetic transcriptions. We have therefore included only phonetic variations which are important enough to involve differences in (precise) phonetic transcriptions of the variants, like in the examples above and like e.g. A. Lacheret-Dujour (1990). We have discarded prosodic variations, including stress, and minor acoustic variations: taking them into account with a satisfactory reproducibility would have required using a corpus of recordings, thus limiting the scope of the description. For example, acoustic differences between epenthetic [t] and non-epenthetic [t] in English (J. Kelly and J.K. Local, 1989; J. Coleman, 1990) are too slight to be represented in IPA. In French words like *signer,* intervocalic [nj] is traditionnally considered as having a variant [ɲ], but even if such a variation is still in use, the phonetic difference is very little (J.J. Spa, 1984), and we consider them phonetically equivalent. Direct auditive perception is too imprecise for such minor variations. Besides, the description of major phonetic variations takes priority.

In the case of phonetic variations, each of the forms has its own phonetic transcription and is referred to as a phonetic or allophonic variant. In addition to precise phonetic transcriptions, we use phonemic representations in order to model phonetic variability, e.g. /lu+e/ for [lue] and [lwe] (*louer*). This phonological treatment is purely linear and phonemic: representations are strings of phonemes; for readability, phonemes are considered as atomic, and not as feature bundles. We have no claims as regards the psychological status of phonemic representations: we consider them as abstract strings built up in order to represent and generate observable phonetic strings. A set of phonetic variants can often be represented by a unique phonemic form. This form can be:

(i) either different from each of the phonetic variants, like /lu+e/ for [lue] and [lwe] (*louer*),

(ii) or identical to one of the variants, like /fele/ for [fele] and [fɛle] (*fêler*).

In the latter case, one of the phonetic variants, here [fele], is selected as canonical. This selection is more or less arbitrary. In order to be able to generate all the variants from the phonemic form, phonologists usually choose the most informative variant, which is more often a careful form rather than a reduced form. However, this choice does not necessarily mean it is a standard form.

This kind of formal system with several levels of abstraction is common practice for phonologists, but it is less used in speech processing. In the speech-processing systems where no such distinction between levels of abstraction is actually implemented, the status of transcriptions is sometimes puzzling. They can be compared to the "phonetic" transcriptions of conventional dictionaries of French: these are expressed in IPA but they usually give only one form in the case of phonetic variations, they implicitly include the result of some phonological analysis and in fact they are rather abstract symbolic representations. When such transcriptions are matched directly with the acoustic signal in a speech-recognition system, results are uncertain. On the other hand, the distinction between two levels of abstraction: precise phonetic transcriptions and phonemic representations, is quite operational and efficient. Industrial experiments in speech recognition (e.g. C. Waast, 1990) obtained interesting results on major phonetic variations by having different phonetic variants represented by different transcriptions: matching precise phonetic transcriptions with the acoustic signal remains a challenging problem, but is easier, since precise phonetic transcriptions are closer to the acoustic signal. The implementation of this principle on a large scale requires a distinction between phonetic transcriptions and a more abstract level.

3. Modelling the conditions of use with local finite-state transducers

Not all variants of a given element are acceptable in all conditions. A number of factors determine the variants that can be uttered:

- lexical factors, e.g. the variation between intervocalic consonant and geminate is observed in *illicite* but not in *village*;
- phonological factors, e.g. *louer* has two phonetic variants `[lue]` and `[lwe]`, and the reduced variant exists for a subset of the conjugated forms of the verb, but not for *loue* `[lu]` or *louera* `[lura]`, for syllabic reasons;
- grammatical factors and others more difficult to cope with.

We neglect stylistic factors, i.e. when phonetic variants have stylistic differences, we consider them equivalent in their usage, though phonetically not equivalent. Lexical factors can be formally described in electronic dictionaries (E. Laporte, 1990). Phonological factors are taken into account by the device that converts phonemic strings into phonetic strings. The most frequently used formal tool for this conversion takes the form of a rewriting system, i.e. a formalism as powerful as Turing machines. Paradoxically, in most cases, the rewriting rules apply in such a way or have such properties that the power of a finite-state transducer suffices to express the conversion, although finite-state transducers are much less powerful than Turing machines. Moreover, any transduction to convert phonemic strings into phonetic strings is likely to be a local map (M.P. Béal, 1987), i.e. its output at a given spot depends only on a bounded part of its input string around this spot and not on the whole input string. The class of local maps is equivalent to that of local finite-state transducers, a strict subclass of finite-state transducers. This simple, readable formal tool is easily handled by quick algorithms. This is why we

implemented phonemic-to-phonetic conversion with local finite-state transducers expressed as such.

Individual phenomena of phonetic variability were expressed in elementary transducers. For example, two transducers contribute to converting /lu+e/, the phonemic string given by the dictionary, into [lue] and [lwe] (*louer*). They convert /u+/ into the phonetic variants [u] and [w], provided that the left context does not belong to a set of sequences of consonants, and that the right context is a vowel different from /u/. These transducers are expressed in a readable form, so as to be conveniently written and modified:

```
                , !nlmrjɲŋ,     u,     +,     aeøoiy,
                                u,
                                w.
!ptkbdgfsʃvzʒ,  nlmrjɲŋ,        u,     +,     aeøoiy,
                                u,
                                w.
```

This readable form is compiled into transducers for processing strings.

Elementary transducers are combined together. The simplest combination occurs between transducers expressing independent phenomena, i.e. when possibilities of variation are independent: for example, the transducers for /i+/, /y+/, /u+/ are independent and can apply at the same time to a phonemic string. When transducers have to be applied in a definite order, the output of one being the input of another, the combination takes the form of mathematical composition. (A composition of local maps is a local map.) For example, the transducers that nasalize a stop after a nasal vowel and before an obstruent (*lampe torche* /lãp tɔrʃ/ [lãmtɔrʃ]) must apply to the output of those that delete final /r/, since this deletion can bring about the contact between the stop and the obstruent, e.g. in *chambre criminelle* /ʃãbr kriminel/ /ʃãbkriminɛl/ [ʃãmkriminɛl].

Transducers are simple but powerful enough to model all the phonetic variations we have examined. In the examples above, individual phenomena are of a narrow scope and can be represented separately because possibilities of variation are independent. The corresponding elementary transducers convert phonemes one by one. In phenomena of a broader scope, one can observe longer-distance interdependencies between possibilities of variation. For example, the possibilities of combining elisions in a sequence of *e*-muets, as in *je ne te le*, obey constraints: the phrase

Si je ne te le dis pas

can be pronounced [siʒøntøldipa] but normally not [siʒnøtlødipa]. These interdependencies between elisions are bounded by the maximum length of sequences of *e*-muets in French, which is of about 10 *e*-muets. These dependencies can be represented in a local transducer that recognizes the whole sequence, then emits the phonetic strings for it.

Another example is the phonetic variants of the pre-verbal pronoun *il*. When *il* occurs immediately before a word beginning with a consonant, it has two phonetic variants [il] and [i] with a stylistic difference:

Il travaille	[iltravaj]	[itravaj]
Il lit	[illi]	[ili]
Il lui plaît	[illɥiple]	[ilɥiple]

However, when *il* occurs immediately before the pre-verbal pronoun *le, la, les, l'*, only the longer variant [il] is used[1]:

Il le sait	[illøse]	*[iløse]	*[ilse]
Il la voit	[illavwa]	*[ilavwa]	
Il l'utilise	[illytiliz]	*[ilytiliz]	

The same is true for the plural form, *ils*. This phenomenon can be represented with a local transducer which recognizes *il(s)* and the beginning of the word immediately after, and then emits the phonetic form(s) for *il(s)*.

Conclusion

Local transducers seem to us a rich formal tool, able to model various phenomena of a more or less broad scope. And yet it is a simple, convenient tool: transducers combine in different ways and are implemented in quick algorithms. We hope they will help with the huge descriptive work required to elaborate a systematic treatment of phonetic variability.

References

BEAL, Marie-Pierre, 1987. *Codage, automates locaux et entropie,* Thèse de doctorat, Université Paris 7, 101 p.

CARSON-BERNDSEN, Julie, 1990. "Phonological Processing of Speech Variants", *Proceedings of the 13th International Conference on Computational Linguistics,* Helsinki, pp. 21-24.

COLEMAN, John, 1990. "Unification Phonology: Another look at "synthesis-by-rule", *Proceedings of the 13th International Conference on Computational Linguistics,* Helsinki, pp. 79-84.

KELLY, J., and J.K. LOCAL, 1989. *Doing Phonology,* Manchester University Press.

LACHERET-DUJOUR, Anne, 1990. *Contribution à l'analyse de la variabilité phonologique pour le traitement automatique de la parole continue multilocuteur,* Thèse de doctorat, Université Paris 7, 303 + 46 p.

[1] A star * marks unacceptable utterances.

LAPORTE, Eric, 1989. "Applications of Phonetic Description", in *Lecture Notes in Computer Science* vol. 377, *Electronic Dictionaries and Automata in Computational Linguistics,* M. Gross and D. Perrin eds., Berlin-New York: Springer-Verlag, pp. 66-78.

LAPORTE, Eric, 1990. "Le dictionnaire phonémique DELAP", in *Langue française* n° 87, *Dictionnaires électroniques du français,* B. Courtois et M. Silberztein éds., Paris : Larousse, pp. 59-70.

SPA, J.J., 1984. "Le [ɲ] est mort. Vive le /ɲ/ !", in *Lingvisticae Investigationes* VIII:1, pp. 151-159, Amsterdam-Philadelphie : Benjamins.