STUDYING PRONUNCIATION VARIANTS IN FRENCH BY USING ALIGNMENT TECHNIQUES

Philippe Boula de Mareüil & Martine Adda-Decker

LIMSI-CNRS
BP 133 F-91403 Orsay CEDEX
Tel.: ++33/0 1 69 85 81 19 — Fax: ++33/0 1 69 85 80 88

ABSTRACT
In this contribution, we address the realization of well-known pronunciation variants in French: the so-called mute e (or schwa), liaisons (sandhi phenomena) and mid vowels in non-word-final syllables, whose behavior with regards to vowel harmony is investigated. We compare the occurrences of such pronunciation variants in large corpora of read and spontaneous speech. Their phonetic transcriptions are automatically obtained by aligning the acoustic data with a pronunciation graph, derived from orthographic transcriptions and pronunciation dictionaries which include relevant pronunciation variants. The frequency of occurrence of schwas and liaisons, together with the pronunciation of mid vowels according to their underlying representation are then computed and analysed linguistically.

1. INTRODUCTION
A number of factors, such as speaking style and speech rate, individual speaker habits and dialectal region, come into consideration to account for speech variability [1]. A speech recognition system can be used to automatically label very large speech corpora in order to carry out further linguistic analyses. By aligning the data with acoustic models which allow for pronunciation variation, the observed alignments provide frequencies for the variants involved in the corpus. Explanations for the observed variants can be proposed at a linguistic level, by the speech data characteristics, or at a speech engineering level, by the properties of the acoustic models [2]. As opposed to previous experiments [3], pronunciation variants are here phonologically motivated.

In this study, we use a speech recognition system [4] (the accuracy of which was demonstrated by a series of evaluations) to examine the use of three types of pronunciation variants in French: the optional /œ/, liaisons and vowel harmony. The schwa, which may or may not be spoken, influences the number of syllables, and hence prosody. Schwa and liaison (i.e. the realization of a normally mute final consonant in the context of a following word which begins with a vowel) are two intrinsic aspects of the French phonology. As for vowel harmony, its rules stipulate that the vowels of a word or a phrase are influenced by a vowel appearing in the same domain, so that all the vowels in this domain turn out to share certain features. This general definition is restricted in French to the influence of a stressed vowel on mid vowels which precede (in case of final stress) or which follow (in case of initial stress) [5].

After presenting these three families of pronunciation variants, we describe the methodology applied for the analysis of variants. We then focus on transcription guidelines for the development of the pronunciation lexicon. Using the alignments obtained, we quantitatively investigate the realization of our pronunciation variants.

2. VARIANT TYPES: LINGUISTIC DESCRIPTION
The schwa vowel and liaison consonants have been referred to as sequential variants [3], because they allow for a variable number of phonemes (by deletion/insertion). In comparison, the third type of variation addressed here, vowel harmony, only induces changes on the paradigmatic axis, amongst mid vowels in French. It therefore belongs to what can be called “parallel variants” (enabling the replacement of a canonical phoneme by another phoneme).

2.1. The schwa vowel
The orthographic e, which is called mute (but also decaying, unstable, feminine, dull, obscure, middle or schwa because it is often omitted in conversational or colloquial speech), when maintained, is somewhere, according to opinions, between the open /œ/ and the closed /ø/ — see references in [6]. But even if these phonemes are its closest neighbors, and even if the pronunciation /œ/ appears to be preferred nowadays, the realization of schwa does not merge exactly into the archiphoneme /Œ/, owing to the absence of lips rounding in the case of /œ/ [7]. The multiplicity of denominations, as well as the doubts concerning its color support the shady nature of this e, which is defined more by its phonological behavior than by its timbre.

What interests us here is: when is the schwa deleted? Now we are confronted in this field (even more than in the others) to the speech rate, to the situation, to the cultural and linguistic environments (the speaker’s region of origin). Confusion reigns especially concerning deletion in word initial syllable: e.g. la semaine (“the week”), la cerise (“the cherry”) [8]. Studies in descriptive phonetics have shown that some people are more inclined to pronounce de venir (“to become”) as dev’nir than d’venin or on se demande (“we wonder”) as on se d’mande rather than as on s’demande.

The pronunciation of the final orthographic e (this is also valid for -es and -ent) is optional. As for the non-final e, it is assumed to be obligatory when preceded by two or more consonants, or when followed by a liquid+/j/ — e.g. chancellor (“chancellor”) —, but is often optional elsewhere: e.g. charretier (“cartier”).

2.2. Liaison consonants
Mute consonants in French, in a weak position or in isolated words, may be pronounced when the word which follows immediately begins with a graphemic vowel, a mute h or
sometime a glide: les yeux (“the eyes” /lezjø/). Stated as such, this rule is too general, and must be refined: the liaison phenomenon should be applied only within phrases — for example, in the configuration determinant + determined, where there is a close grammatical link between the terms.

3.2 Pronunciation lexica and representation of variants

The pronunciations and their variants were generated by a grapheme-to-phoneme converter, into which optional schwas and liaisons as well as vowel harmony rules were introduced. The GRAPHTON+ software was used, whose word error rate on several 30,000 word running texts is less than 1% [11].

For the purpose of this study, the description level is “phonetically”. And rules are determined phonologically, rather than regionally or sociolinguistically. The transcription system is very close to the International Phonetic Alphabet (IPA), completed by a set of diacritics and meta-symbols to represent the pronunciation variants. Within the framework of a project reported in [12], diacritics have been introduced so as to distinguish optionally realized phonemes which correspond to floating or latent segments, and neutralization phenomena (particularly of mid vowels, in unstressed position).

For distinguishing between optional and obligatory realizations of schwas (/a/- and /æ/ respectively), a kind of “three consonant law” is applied (a mechanical, dynamic well-formedness constraint which in a modern Optimality Theory framework could simply be rewritten as *CCC [6]): final schwas and non-final schwas, in the context (VH#CeCV) are marked as optional, as in samédi (“Saturday”), whereas the schwa is marked as obligatory in a word like vendredi (“Friday”). The lexicon also includes liaison phonemes (/l/, /pl/, /yl/, /yl/). Finally, /e/ and /e/, specifying the preceding phoneme, indicate opening and closing respectively. In order to choose, for instance, between /ɔ/ (an open /ɔ/ which tends to be closed) and /o</ (a closed /o/ which tends to be open), from the underlying transcription, two rules (and only two) are specified for pairs /o/ (IPA) and /o/.

These rules are applied sequentially:

1. An open (resp. closed) vowel tends to be closed (resp. open) in open (resp. closed) syllable, except in “full” word-final syllable (when a schwa is optional, as in alcalose, the mid vowel keeps its quality).

2. In a sequence of mid-closed mid-open (resp. mid-open mid-closed) vowels, the first one gets open (resp. closed). This rule applies in the sequences /e...e, e..o, o...e, o...a/; whatever the complexity of the consonant cluster /.../. In particular, glides are transparent: they do not stop the propagation.

This leads to: fête /fête/ (“party”); testé /testé/ (“tested”); microphone /mikrofône/ (rule 2, which opens the closed /o/ of the prefix), jeunesse /ʒœnzœ/ ("youth": rule 1 gives /ʃœnœsœ/, then rule 2 opens the /œsœ/). Cascade phenomena such as resyllabification resulting from the drop of a schwa are ticklish (see above). In the well-known case of médecin (“physician”) for example, we have /me<œsœ/, which can be expanded into four possible pronunciations: /meœsœ/, /meœsœ/, /meœsœ/, /meœsœ/.

By “underlying form” we mean the vowel quality defined in the lexical representation of the word, which is used here as a starting point for the derivation of harmony and neutralization...
rules. The underlying form of the graphic ‘o’, elsewhere than in word-final syllable, is /ɔ/ except before lengthening consonant (/l/), and except in the prefixes psycho-, micro-, auto-, etc., where the target phoneme is /ɔ/ [13]. For the ‘e’ before a double consonant, the default rule is the following: /ɛ/ if the ‘e’ is initial and if the consonant is not ‘t’ — e.g. ecchymose (“bruise”) /ɛkmɔz/ or effort /ɛfɔr/ /ɛ/ in other instances, more often than not. It is actually more complicated.

### 3.3. Speech corpora

This study makes use of the BREF read speech corpus [14] and part of the MASK spontaneous speech task-oriented corpus [15]. The BREF corpus contains 66,500 sentences extracted from the Le Monde newspaper and read by 120 speakers (100 hours of speech). In the corresponding 26,000 word lexicon, 37% of the words contain optional schwas and over 25% have liaisons. The MASK data consist of 38,000 sentences of train information, with a total of 409 speakers and 35 hours of speech. The corresponding word list only contains 2,000 entries, 35% of which have schwas, 30% have liaisons, and about 10% include at least one /ɛ/ or /ɛ/.

For MASK, the 10 most frequent words account for 30% of the corpus, whereas for read newspaper speech, they cover about 20% of the data. The 100 most frequent words cover 80% of MASK, but slightly less than 50% of BREF. Due to the domain-specificity of the MASK corpus, 1,000 words cover essentially the entire corpus. Even if thousands of distinct items occur, only few of them have many repetitions (cf. Zipf’s law).

### 4. RESULTS

#### 4.1. Schwa

Following Dell’s rhythmic schwa processing [8] Table 1 summarizes results concerning the occurrence of final and non-final schwas in the BREF and MASK corpora. “Final schwa’s” refer to words terminated by an orthographic ‘e’, preceded by a consonant or by ‘qu’. And “monosyllable” here refers to a word whose unique vowel is a schwa: ce (“this”), de (“of”), je (“I”), le (“the”), me (“me”), ne (“not”), que (“that”), se (“oneself”), te (“you”), covering 9% of BREF and 16% of MASK.

<table>
<thead>
<tr>
<th></th>
<th>BREF</th>
<th>MASK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>word</td>
<td>%observed (#)</td>
</tr>
<tr>
<td>final</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monosyllable</td>
<td>97 (93,488)</td>
<td>65 (27,477)</td>
</tr>
<tr>
<td>polysyllable</td>
<td>19 (32,493)</td>
<td>10 (4,087)</td>
</tr>
<tr>
<td>non-final</td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial syllable</td>
<td>98 (23,595)</td>
<td>77 (2,793)</td>
</tr>
<tr>
<td>internal</td>
<td>20 (6,837)</td>
<td>14 (392)</td>
</tr>
</tbody>
</table>

*Table 1: Word level analysis of final and non-final schwas which are maintained or dropped in BREF and MASK.*

In the BREF read speech corpus, almost all the initial syllable schwas are pronounced, even in monosyllables. Some exceptions are the pronouns -je (“I”) and -ce (“this”), which are enclitics in interrogative forms. Despite their tokenization as monosyllables, their behavior is closer to that of polysyllables.

For the MASK corpus, about 2/3 of the schwas are maintained in monosyllables and in initial syllable position. In contrast, the majority of schwas are dropped in other positions.

While on average 70% of non-final schwas are pronounced, the word position is seen to be an important factor. We examined the behavior of word internal schwas, where the immediate phonetic context is known easily: a schwa is most frequently recognized by the system when it is between two sonorant consonants, for instance between [l] and [m] — e.g. également (“also”) — or between [m] and [g], as in the frequent word aimerais (“would like”). This does not necessarily imply that the speaker actually uttered a schwa, which is a limit of the acoustic but objective analysis. Nonetheless, it is interesting to compare this with the following.

For the schwa in monosyllables, we examined a phonetic context which goes beyond the word level, in which case the three consonant law may be violated. In sequences of monosyllables, the schwa falls more often than not after fricatives, and more after fricatives than after plosives: hence the pronunciations such as je n’ (the schwa in the negation is deleted more often than not) or c’que, rather than j’ne or ce qui’, in an average of 90% of instances on BREF. This result is consistent with the phonological Sonority Sequencing Principle, which seems to play a greater role than other criteria such as the voicedness/voicelessness opposition in this respect. This tendency remains true (also on MASK), although less clear if we consider monosyllables followed by a word whose first syllable contains an optional schwa (e.g. je r’cherche): proportions are only 65-80%. As far as the three consonant law is concerned, it applies across words in less than 20% of the occurrences on MASK, facing more than 70% on BREF. Nevertheless, false starts, filled pauses and hesitation phenomena make it more difficult to interpret the MASK word sequences.

#### 4.2. Liaisons

Realization of liaison, more so than schwa, is a word-dependent issue: this explains why we refer to the “lexical entry” category. Liaison is analyzed here as a function of word frequency. In Table 2, we distinguish subsets of % of most and least frequent word tokens in the whole corpus; the 100% line corresponds to the subset of lexical entries for which more than 100 liaison contexts are observed in the corpus. Table 2 then gives the following information: the number of entries in the respective lexica having a possible liaison (i.e. the consonant is optional in the lexicon and the word appears in the corpus before a word starting with a vowel); the percentage of entries for which at least one liaison is observed (i.e. we can consider that the liaison is not forbidden in the particular context); the number of word occurrences in the corpus where liaisons could occur; and the percentage of realized liaisons.

About half the possible liaisons are realized in BREF (55.3%) and in MASK (42.8%) as shown by the 100% entry in Table 2. The BREF data show that liaison is more frequent in common words (75.5%) than in rare words (18.1%). Essentially all word tokens (3 exceptions) of the 100% set belong to the 1% set. For the 52 lexical entries in common, only 9 words have liaison realization rates under 50%: e.g. the coordination conjunction et (“and”), which is known to forbid liaison.

The most variable cases (i.e. optional or potential liaisons) above the 50% threshold are the conjunction mais (“but”) and (semi-)auxiliary verbs. The parts-of-speech determiner, pronoun and (anteponable) adjective all have over 78% realized liaisons: for the 223 occurrences of the possessive determiner/ adjective nos (“our”) all possible liaisons were made.

#### 4.3. Mid vowels

The alignment output provides a series of lines such as:

```
```

which means that out of 1,577 occurrences of the word réservation in MASK, the first ‘é’ is 933 times against 644 closer
to the /ɛ/ model than to the /e/ model. On 12,678 (resp. 71,410) occurrences including at least one /ɛ/ or /e/ in BREF (resp. MASK), mid vowels with diacritics are aligned in 73% (resp. 60%) with their underlying quality: /ɛ/ with [ɛ], /e/ with [e]...

Breakdowns according to the vowel type display quite different results between the read and spontaneous speech corpora. In MASK, for example, only half of the /ɛ/ls are aligned with [e], because of the large population of words in réserve- (rather open). This may be interpreted in terms of articulation, which would be more careful in read speech than in spontaneous speech. But these differences are rather due to the fact that we have too scarce data in MASK. Anyway, the [ɛ] of a word like mauvais(e) (“bad”) and the [e] for the digram ‘ai’ in a word like aimé (“liked”) reflect the tendency to harmony affected by the open (resp. closed) phoneme of the following syllable. Interestingly in return, harmony closing the [ɛ] is somewhat blocked by the presence of a subsequent /ɛ/ — e.g. libéré (“freedom”). As for the front series, labial vowels in deuxième (“second”) and déjeuner (“to have lunch”) are, in their realizations, consistent with their underlying vowel quality, despite some uncertainty in the latter case.

In spite of orthography, the vowel /ɛ/, in fact in période (“period”) and words in réserve-, takes the mid-open vowel quality of the stressed vowel which follows. This has been confirmed by acoustic measurements [16], and contradicts [10] who observes, from the analysis of a TV corpus, that the presence of an explicit mark of mid-closed pronunciation (here the acute accent) prevents the vowel from being open.

5. CONCLUSION

Using speech recognition and more precisely automatic alignment, we observed the realization of schwas, liaisons and unstressed mid vowels in French read and spontaneous speech. We noticed that at least 2/3 of optional schwas are maintained in monosyllables and in initial syllable, whereas a vast majority of optional schwas are elided elsewhere. As for liaisons, considering the sole latent phonemes in a right vowel context, we observed a liaison realization in half of the cases. Because of homonymy phenomena, we only considered lexical frequency criteria: the more frequent words are, the more they generate liaisons and vice versa. While frequency-related, liaisons are also strongly linked to lexical, syntactic and idiomatic levels: they should be linked with part-of-speech tags in a more precise study. Moreover, they should be related to prosody, which also enables us to evaluate the cohesion between the words of the speech flow.

For mid vowels, the transcription system we adopted, although more precise than the strict IPA, does not enable one to distinguish easily whether a mid-open vowel is closed by vowel harmony or not. Nor does it enable us to assess the influence of derivational morphology. Nevertheless, we think it is worth extending it to speakers of non-standard dialects.

6. REFERENCES


<table>
<thead>
<tr>
<th></th>
<th>BREF</th>
<th>Mask</th>
</tr>
</thead>
<tbody>
<tr>
<td>freq.</td>
<td>lexicon</td>
<td>speech</td>
</tr>
<tr>
<td></td>
<td>#entries with £</td>
<td>% entries with observed £</td>
</tr>
<tr>
<td>1% most</td>
<td>114/83</td>
<td>80/71</td>
</tr>
<tr>
<td>n% most</td>
<td>879/50</td>
<td>64,293</td>
</tr>
<tr>
<td>100% most</td>
<td>4,084/25</td>
<td>79,936</td>
</tr>
<tr>
<td>n% least</td>
<td>221/22</td>
<td>343</td>
</tr>
<tr>
<td>100£</td>
<td>55/50</td>
<td>45,086</td>
</tr>
</tbody>
</table>

Table 2: Analysis of possible and observed liaisons (£) as a function of word frequency in the BREF and Mask corpora. Subsets correspond to the 1, n, 100% most frequent words and to the n% least frequent words: n=10 for BREF and n=50 for Mask, given the very different lexical distributions in the two corpora. 100£ corresponds to the subset of lexical entries for which over 100 liaison contexts are observed.