PRONUNCIATION VARIANTS IN FRENCH: SCHWA & LIAISON

Martine Adda-Decker*, Philippe Boula de Mareüil*† & Lori Lamel*
* *Spoken Language Processing Group, LIMSI-CNRS, Orsay, France*
† *now with Élan Informatique, Toulouse*

ABSTRACT

In this communication, we focus on well-known pronunciation variants in French, which are style-related and also partially depend on the speakers' socio-cultural background: the so-called mute *e* (or schwa) and liaisons (sandhi phenomena). We compare the occurrence of such pronunciation variants in large corpora of orthographically transcribed speech: the BREF corpus comprised of 120 hours of read newspaper speech and a portion of the MASK corpus containing 35 hours of spontaneous speech collected via a travel information dialog system. Phonemic transcriptions of the speech corpora were automatically obtained by aligning the acoustic data with a pronunciation graph, derived by expanding the orthographic transcription using a pronunciation dictionary which includes relevant variants. The frequency of occurrence of schwas and liaisons were computed from the aligned phonemic transcriptions, and the effect of lexical frequency is investigated.

1. INTRODUCTION

As reported in the literature [7] pronunciation variants can be related to a variety of factors such as the speaking style, speaking rate, individual speaker habits and dialectal region.

Speech recognition systems can be used to automatically label large speech corpora in order to carry out further linguistic analyses. By aligning the data with acoustic word models which allow for pronunciation variation, the observed alignments provide frequencies for the main variants in the corpus. The alignment results do of course depend on the permitted pronunciation variants, the acoustic models, and more generally on the parameters of the speech recognizer. 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.

In this contribution, we use a speech recognition system to examine the use of two types of pronunciation variants in French: the optional /ə/ and *liaisons*. The schwa, which may or may not be spoken (in word-internal position or at the junction of successive words) influences the number of syllables, and as a consequence the prosody. The liaison (i.e. the realization of a normally mute final consonant in the context of a following word which begins with a vowel or glide, or mute *h*) are two intriguing specificities of the French language.

We start with a linguistic presentation of these two families of pronunciation variants, and address their importance for automatic speech processing, and in particular for automatic speech recognition. We then describe the methodology applied for the analysis of variants, focusing on the development of the pronunciation lexicon. Using the alignments obtained, we quantitatively investigate the link between word frequency and pronunciation variants: frequent words are likely to drop schwas and to generate liaisons, and vice-versa for rare items.

2. VARIANT TYPES: SCHWA & LIAISON

The schwa-vowel and liaison-consonants are two variant types which allow for a variable number of phonemes in French, what we refer to as sequential variation.

2.1. Linguistic description

2.1.1. The schwa /ə/ vowel The orthographic *e*, which is called *mute* (but also *decaying, unstable, feminine, dull, obscure, middle, neutral* or *schwa* because it is more often than not omitted in conversational speech and, when maintained is somewhere (according to opinions), between the open /œ/ and the closed /ø/. But even if these phonemes are its closest neighbors, and even if the pronunciation /œ/ appears to be preferred, the realization of schwa does not merge exactly into the archiphoneme /Œ/, probably owing to the absence of lips arounding in the case of /ə/ [9]. The multiplicity of denominations, as well as the doubts concerning its timbre support the shifty nature of this *e*.

What interests us here is: when is the schwa deleted? Confusion reigns especially concerning deletion in word initial syllable: e.g. *la semaine ("the week"), la cerise ("the cherry")* [5]. When *e*'s are separated by a single consonant, the French elide about every other one. These are usually of even rank, with *j'te*, but *je n'* (in spontaneous speech the schwa in the negation is deleted more often than not), and *c'que*: *de c'que je n'te l'red'mandais pas*. The pronunciation of the /ə/ is directly linked to syllabification. Studies in descriptive phonetics have shown that some people are more inclined to pronounce *devenir ("to become")* as *dev'nir* than *d'venir* or *je le sais ("I know it")* as *je l'sais* rather than as *j'le sais*.

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 obligatory when preceded by 2 or more consonants *brebis ("sheep")*, when followed by a liquid+/j/ *chancelier ("chancelior")*. It is usually present when it is in phonological opposition with another word *pelage ("coat")* vs *plage ("beach")*, but is often optional elsewhere: *samedi ("Saturday"), charretier ("carter"), retour ("return")*. The /ə/ is also optional within compound words formed with *garde-* and *porte-* when the 2nd element is at least disyllabic: *garde-manger ("pantry"), porte-bonheur ("charm")*.

2.1.2. Liaison consonants Mute consonants in French, in a weak position or in isolated words, may be pronounced when the word that follows immediately begins with a graphemic vowel or a mute *h*, thus avoiding a hiatus: *les yeux ("the eyes")* /le**z**jø/.

Stated as such, this rule is too general. The liaison phenomenon should be applied only within and not across phrases. Liaison is more frequent between determiners and nouns than between nouns and adjectives. Liaison is rarely made with adverbs, but can be found on adverbs of quantity before adjectives as in the word sequence *plus ouvert ("more open")*. While not prohibited, successive liaisons are generally avoided by speakers. The so-called liaison phenomenon is a survival of a time when all final consonants were pronounced, and today its realization depends on the speaking style and on the speakers' socio-cultural background. A limited number of consonants are used for *liaison*: /z/, /t/, /n/, /r/, /p/ corresponding to the written forms {*-s,-z,-x*}, {*-d,-t*}, *-n, -r, -p* respectively. The most common liaisons are /z/ and /t/ arise from plural marks on determiners, adjectives, nouns and verbs.

How and when do is liaison made? In liaisons, some modifications may occur. The final consonants *-z, -p* and *-t* are articulated according to their orthography. The voicing characteristic of some other consonants can change: *-s* & *-x* are realized as a /z/ when they serve for liaison, and conversely, ' *-d* → /t/. It should be mentioned that the plural forms of idioms are often uttered as in the the singular: *guets-apens ("ambushes")* /gɛ**t**ɑpɑ̃/. In BDLEX [10], we only noticed 6 exceptions to this rule *petits-enfants ("grand-children")* /p(ə)ti**z**ɑ̃fɑ̃/ whereas there were 47 forms containing *-s* -vowel in this dictionary.

In general, liaisons have no effect on the previous vowel, except for *-n* which may engender the denasalization of vowels such as the /ɔ̃/ in *bon ("good")* or *non-intervention*), and /ɛ̃/ in adjectives like *ancien*. There are 84 in BDLEX but only a few are anteposable. As for oral vowels, in singular forms finishing with *-er*, the /e/ becomes /ɛ/ due to the *-r* liaison: *dernier homme ("last man")* /dɛRnjɛ**r**ɔm/.

2.2. Influence on automatic speech recognition

If liaison and mute-e are not properly accounted for in the recognition system, then recognition errors are likely to occur. It is thus important to represent them, either directly in the lexicon, as phonological rules or implicitly in the acoustic models. The most common liaisons are made by inserting the phoneme /z/ after words ending with *-s -z*, or *-x* precede words starting with a vowel. To give an idea of the magnitude of this phenomenon, over 25% of the words of the lexicon used in this study have a /z/ liaison. In earlier work, we experimented with a straightforward solution of adding optional liaison phonemes to all applicable words. Unfortunately, this exhaustive approach did not reduce the word error rate, as the large number of variants introduced additional homophone sequences and introduced different errors.

Some example errors involving /z/ liaisons made by our French AUPELF'97 system (Adda et al., 1997a) are shown in Table 1. The lower part of Table 1 gives the phonemic transcriptions in the recognition lexicon. These examples illustrate that missing phonemes in an acoustic word model (formed by concatentating phone models according to the pronunciation in the lexicon) may introduce errors. The system may choose a solution which respects the observed consonant-vowel structure of the data, by exchanging vowels (/e/ and /i/ of *écrites* are replaced by /ə/ and /ɛ/ of *secrètes*) or consonants (/z/ and /g/ of *-s anglais* are replaced by /s/ and /b/ of *semblait*).

A similar problem arises with the optional word-final schwa. When a schwa is present in the acoustic observation and missing in the lexicon, a small function word (conjunction, pronoun, etc.) is often inserted. An example error of this type is given Table 2.

(A)	Reference transcription	System hypothesis
	plaintes écrites	*plaintes secrètes*
	industriels anglais	*industriels semblait*
(B)	**Orthographic form**	**Phonemic form**
	plaintes	plɛ̃t
	écrites	ekRit
	secrètes	səkRɛt
	industriels	ɛ̃dystRijɛl
	anglais	ɑ̃glɛ ɑ̃glɛz
	semblait	sɑ̃blɛ sɑ̃blɛt

Table 1: Examples of recognition errors due to missing *liaison*. Reference transcriptions and system hypotheses (**A**); corresponding lexical entries in the pronunciation lexicon (**B**).

(A)	Reference transcription	System hypothesis
	publique	*public que*
(B)	**Orthographic form**	**Phonemic form**
	publique	pyblik
	public	pyblik
	que	kə

Table 2: Example recognition errors due to missing *word-final schwa*. Reference transcription and system hypothesis (**A**); corresponding lexical entries in the pronunciation lexicon (**B**).

3. METHODOLOGY & EXPERIMENTAL CONDITIONS

In this section we describe the methodology, lexica and corpora used in this study.

3.1. Alignement using phone-based acoustic models

The acoustic phone models are sets of continuous density hidden Markov models (HMMs) with Gaussian mixture. Context-dependent phone models can be used to account for allophonic variation observed in different contextual environments. In order to determine the sequence of realized phones in a given utterance, a Markov chain is formed by concatenating the phone pronunciations associated with the words in the corresponding orthographic transcription. This is then used to constrain the search space for the decoder, aligning the phones with the speech signal. If pronunciation variants are represented in the lexicon or added by phonological rules, a phone graph is constructed and aligned with the signal. In this case the decoder will produce the most likely sequence of phones along with the time alignment.

3.2. Pronunciation lexica

The pronunciations and their variants were generated by a grapheme-to-phoneme converter, into which optional schwas and liaisons were introduced. The program GRAPHON+ [3] was used, whose phoneme error rate on several 30,000 word running texts is less than 1% [8].

For the purpose of this study, where we are concerned only with the pronunciations of words, the description level is "roughly

phonetic", with no supra-segmental markers. The description used in the transcription system is very close to IPA, completed by a set of diacritics and meta-symbols to specify the pronunciation variants. In particular:

- the lexicon includes liaison phonemes (/n/, /p/, /r/, /t/, /z/);
- the lexicon distinguishes optional and obligatory realizations of schwas. A "3-consonant rule" is applied: final schwas and non-final schwas, in the context (V#)C*e*CV), are marked as optional as in *petit ("small")*, whereas the schwa is marked as obligatory, in words like *vendredi ("Friday")*.

The transcription of *petit* is /p(ə)ti p(ə)tit_/, which can be expanded into /pəti/, /pti/, and, into /pətit/, /ptit/ in a liaison context. In return, the place of possible epentheses is not noted (e.g. the possible appearing of a schwa within *ours blanc ("white bear")*.

3.3. Speech corpora

This study makes use of the BREF corpus [1] of read speech and a part of the MASK [2] corpus of spontaneous speech. For both corpora, the speakers have no marked accent. The BREF data contains 66,500 sentences from 120 speakers. 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 from 409 speakers. The word list comprises 2,000 entries, 35% of which contain schwas and 30% have liaisons.

Figure 1 shows the cumulative lexical coverage of the speech corpora as a function of the word frequency rank. For MASK (spontaneous task-oriented speech), 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 the MASK data, but slightly less than 50% of BREF. The coverage of the BREF data is close to linear on the logarithmic scale (Zipf's law). A much stronger slope is observed for the MASK data, between ranks 10 and 200, due to the domain-specificity of the corpus. 1000 words are seen to cover essentially the entire corpus.

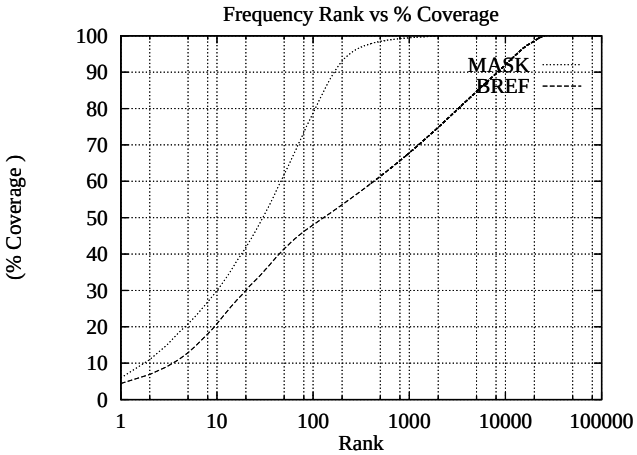

Figure 1: Lexical coverage of spontaneous speech (MASK) and read speech (BREF) corpora as a function of word frequency rank.

		BREF read	MASK spontaneous
schwa	word	%observed (#)	%observed (#)
final	monosyll.	97 (93488)	65 (27477)
	polysyll.	19 (32493)	10 (4087)
non-final	initial syll.	98 (23595)	77 (2793)
	internal	20 (6837)	14 (392)

Table 3: Analysis of final and non-final **schwas** that are maintained or dropped, in the BREF and MASK corpora. "Monosyllabic word" here refers to a word whose unique vowel is a schwa.

4. EXPERIMENTAL RESULTS

In this section we present experimental results on the occurrence of schwas and liaisons in read and spontaneous speech. The data analysis accounts for word frequency.

4.1. Schwa

Table 3 summarizes results concerning the occurrence of final and non-final schwas in the read speech BREF corpus and spontaneous speech MASK corpus. By "final schwas", we mean words terminated by an orthographical 'e', preceded by a consonant or by 'qu'. In the BREF corpus, almost all the initial syllable schwas are pronounced, even in monosyllabic words. Some exceptions are the pronouns *-je ("I")* and *-ce ("this")*, which are enclitics in interrogative forms. In polysyllabic words, final schwas are produced twice as often as in spontaneous speech.

For the spontaneous speech (MASK corpus) about 2/3 of the schwas are maintained in monosyllabic words and in syllable initial position. In contrast, the majority of schwas are dropped in other positions. Schwas are seen to be less often present than in the BREF corpus. 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 since the immediate phone context is easily known. A schwa is most frequently recognized by the system when it is between two sonorants, for instance, between /l/ and /m/ (e.g. *également ("also")*) or between /m/ and /r/), as in the frequent word *aimerais ("would like")*. This does not imply that the speakers actually uttered a schwa, a limit of the acoustic but objective analysis.

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, which will need to be foreseen in a speech recognition system. We distinguish for BREF/MASK respectively:

- the 1% most frequent word tokens;
- the 10%/50% most frequent word tokens;
- the 10%/50% least frequent word tokens.
- the lexical entries ending with a liaison consonant that occur at least 100 times in a right vocalic context;

Tables 4 and 5) give the number of entries in the respective lexica having a possible liaison (ie., 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 (ie., we can consider that the liaison is not forbidden in the particular context); the number of word occurrences

BREF - read speech				
lexicon (25600 entries)			speech corpus	
subset N	#entries with $\mathcal{L}$	%entries with observed $\mathcal{L}$	#$\mathcal{L}$ contexts	% $\mathcal{L}$ obs.
1%	114	83	48771	75.7
10%	879	50	64293	64.2
100%	4084	25	79936	55.3
10% least	221	22	343	18.1
100$\mathcal{L}$	55	50	45086	79.0

Table 4: Analysis of possible and observed liaisons ($\mathcal{L}$), as a function of word frequency in the BREF corpus. Subsets correspond to N most frequent words and **10% least** frequent words. 100$\mathcal{L}$ corresponds to the subset of lexical entries for which more than 100 $\mathcal{L}$ contexts are observed in the corpus.

MASK - spontaneous speech				
lexicon (2260 entries)			speech corpus	
subset N	#entries with $\mathcal{L}$	%entries with observed $\mathcal{L}$	#$\mathcal{L}$ context	% $\mathcal{L}$ obs.
50%	156	42	11435	42.9
100%	200	35	11480	42.8
50% least	44	11	45	11

Table 5: Analysis of possible and observed liaisons ($\mathcal{L}$), as a function of word frequency in the MASK corpus.

in the corpus where liaisons could occur; and the percentage of realized liaisons. 100$\mathcal{L}$ corresponds to the subset of lexical entries for which more than 100 $\mathcal{L}$ contexts are observed in the corpus.

For the BREF corpus, about half the possible liaisons are realized (see the 100% entry in Table4. Liaison is more frequent in common words (75% realization) than rare words (under 20% realization). Essentially all word tokens (3 exceptions) of the 1%-line belong to the set of $\mathcal{L}$. For the 52 lexical entries in common, only 9 words have liaison realization rates under 50%:

- two nouns: *temps (“time”), ans (“years”)*, with 789/123 occurrences in a liaison context and 20/0% liaisons realized respectively. For *temps*, liaisons occur systematically in expressions such as *de temps en temps (“from time to time”).*
- two adverbs ending in *-rs*: *alors (“then”), toujours (“always”)* 119/194 occurrences in a liaison context, 0/4% liaisons realized;
- the words *après (“after”)* and *depuis (“since”)*, which may be either prepositions or adverbs;
- the negation adverb *pas (“not”)*, and the semi-auxiliary/noun *fait (“does”/“fact”)* where liaisons are realized almost 50% of the time;
- the conjunction *et (“and”)* which is known to forbid liaisons.

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

The MASK data are difficult to compare to BREF, primarily because the vocabulary is much more specialized, and the study was carried out on less data. On average the liaison realization rate is lower: 42.8% vs 55.3%. This is largely due to the high frequency of words like *voudrais (“would like”)* in the MASK corpus, which do not generate as many liaisons as function words do. But even these (e.g. *un “a”, en (“in”)* have lower rates of liaison realization than are observed in BREF. As for the read speech data, the rate of of observed liaisons is higher for frequent words than for rare ones.

5. DISCUSSION

The liaison, while frequency-related, is strongly linked to syntactic and idiomatic levels. As mentioned above, the word *temps, (“time”)* usually has no liaison, except for idioms like *de temps en temps*, *en temps et lieu* , *en temps ordinaire*. The word *un* commonly occurs as a determiner and as such has almost systematic liaison. When used as a pronoun, the liaison is avoided.

For schwa insertion and deletion we investigated subsets of the data, focusing on the most frequent words. The schwa occurrences on frequent words are similar to the figures presented on the complete corpora. Comparing read and spontaneous speech, schwas are more often pronounced in the first case. For example, the schwa is present in 96% of the instances of the word *sera (“will be”)* which occurs 680 times in the BREF corpus. The same word occurs about 250 times in the spontaneous speech data, where the schwa is dropped in 70% of the occurrences. This type of analysis is particularly important in understanding the behavior of speech recognition systems and to account for its accuracy when processing the acoustic signal, since variants often increase the homophone rate and decrease the decoding performance in terms of computational requirements.

REFERENCES

[1] Lamel L.F., Gauvain J.-L., Eskénazi M., BREF, a Large Vocabulary Spoken Corpus for French, *Eurospeech*, 1991.

[2] Lamel L., Rosset S., Bennacef S., Bonneau-Maynard H., Devillers L., Gauvain J.L., Development of Spoken Language Corpora for Travel Information, *Eurospeech*, 1995.

[3] Boula de Mareüil P., *Étude linguistique appliquée à la synthèse de la parole à partir du texte*, PhD Thesis, Université de Paris XI, 1997.

[4] Béchade H.-D., *Syntaxe du français moderne et contemporain*, Presses Universitaires de France, Paris, 1992.

[5] Dell F., *Les règles et les sons*, Hermann, Paris, 1973.

[6] Encrevé P., 1988, *La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français*, Travaux linguistiques, Éditions du Seuil, Paris.

[7] Fouché P., *Traité de prononciation française*, Éditions Klincksieck, Paris, 1969.

[8] Boula de Mareüil P. et al., Evaluation of grapheme-to-phoneme conversion for text-to-speech synthesis in French, *LREC*, 1998.

[9] Martinet, *La prononciation du français contemporain* Librairie Droz, Genève/Paris, 1971.

[10] Pérennou G., de Calmès M., BDLEX, base de données lexicales du français écrit et parlé, *CERFIA*, Toulouse, 1988.