Investigating syllabic structures and their variation in spontaneous French

Martine Adda-Decker, Philippe Boula de Mareüil, Gilles Adda & Lori Lamel

Spoken Language Processing Group & Situated Perception Group
LIMSI-CNRS, BP 133
91403 Orsay Cedex, FRANCE

Abstract

The paper presents a study of syllabic structures and their variation in a large corpus of French radio interview speech. A further aim is to show how automatic speech recognition (ASR) systems can serve as a linguistic tool to consistently explore virtually unlimited speech corpora. Automatically selected subsets can be manually checked to accumulate knowledge on pronunciation variants. Our belief is that better formalised knowledge of variant mechanisms will ultimately contribute to improve pronunciation modelling and ASR systems. This study is meant to be a step in this direction. The linguistic phenomena we are particularly interested in, are sequential variants (i.e. variants with different numbers of phonemes) which may or not entail syllabic restructuring. These variants, frequent in spontaneous speech, are known to be particularly difficult for speech recognizers. To focus on sequential variants, a methodology has been set up using descriptions at the phonemic, syllabic and lexical levels.

This study reports on a radio corpus composed of thirty 1-hour shows of interviews. Spontaneous speech is found to have a larger proportion of closed syllables than found in the canonical syllables derived from orthographic transcriptions. As expected, the optional schwa contributes to a large amount of variation in syllabic structure. Less well described phenomena are also observed, such as other vowels (/u/, /ɛ/, /i/ and /a/) being deleted in a non-final (unstressed) position. Unstressed CV syllables, when preceded by an open syllable, are likely to undergo syllabic restructuring: vowel deletion together with backward onset-coda transfer. Complex syllables tend to be simplified: liquid consonants are often deleted, more often in coda than onset position. /v/ is the most deletion-prone consonant in both onset and coda positions. Finally, a substantial percentage of occurrences of word-final schwa syllables may completely disappear.

Résumé

Dans ce papier, nous traitons des structures syllabiques et de leur variation dans un corpus de parole en français issu d'entrevues radio-diffusées. Un des buts est de montrer comment des systèmes de reconnaissance automatique de la parole (RAP) peuvent servir d'outils linguistiques pour explorer de façon cohérente des corpus virtuellement illimités. Des sous-ensembles automatiquement sélectionnés peuvent être vérifiés manuellement pour accroître notre connaissance des variantes de prononciation. Notre conviction

est qu'une meilleure formalisation des mécanismes à l'oeuvre dans la parole contribuera en définitive à améliorer la modélisation des prononciations et les systèmes de RAP: cette étude se veut une étape dans cette direction. Les phénomènes linguistiques auxquels nous nous intéressons en particulier sont les variantes séquentielles (i.e. celles qui induisent un nombre variable de phonèmes), qui peuvent selon les cas conduire à une restructuration syllabique: ces variantes, fréquentes en parole spontanée, sont connues pour poser problème à la reconnaissance. Pour se focaliser sur elles, une méthodologie a été mise au point, utilisant des descriptions aux niveaux phonématique, syllabique et lexical

Cette étude repose sur un corpus de parole de radio constitué de trente émissions d'une heure. La parole spontanée révèle une plus grande proportion de syllabes fermées que dans les syllabes canoniques dérivées des transcriptions orthographiques: comme on peut s'y attendre, le schwa optionnel contribue pour une grande part à la variation de structure syllabique. Des phénomènes, moins bien décrits ont également été observés : d'autres voyelles telles que /u/, /ɛ, /i/ et /a/ peuvent tomber en position inaccentuée (non finale). Les syllabes CV non accentuées, précédées d'une syllabe ouverte, sont enclines à la restructuration : effacement de la voyelle et transfert attaque-coda. Les syllabes complexes tendent à être simplifiées : les consonnes liquides tombent souvent, plus en position de coda qu'en position d'attaque. Le /v/ est la consonne la plus facilement élidée indépendamment de sa position dans la syllabe. Enfin un pourcentage substantiel de syllabes faibles de fin de mot, ayant un schwa comme noyau, peuvent disparaître complètement.

Key words: Pronunciation dictionaries, Pronunciation variation, Spontaneous Speech Recognition, Reduction Phenomena, Syllabic Restructuring, Syllable Deletion

---

1 Introduction

Speech recognition has made tremendous progress this past decade, with a significant decrease in recognition word error rates. Present challenges concern improved language modelling and pronunciation modelling. The problem of pronunciation variant modelling appears to be crucial especially for spontaneous speech. Even though language complexity (in terms of vocabulary size and perplexity) is smaller than it is for broadcast news transcription tasks, decoding accuracy on conversational speech is significantly worse, indicating that the word models are insufficient. In addition to the problem of what is commonly addressed as disfluencies (Shriberg 1994), spontaneous speech decoding seems to suffer from inappropriate modelling of reduced pronunciations. Many efforts have been spent these last years on pronunciation variants by the ASR community. Some recent workshops have addressed pronunciation modeling for ASR (the ESCA Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition, Rolduc, May 1998; the ISCA

Email address: {madda,mareuil,gadda,lamel}@limsi.fr (Martine Adda-Decker, Philippe Boula de Mareüil, Gilles Adda & Lori Lamel).

Pronunciation Modeling and Lexicon Adaptation for Spoken Language, Estes Park, September 2002) and an interesting overview can be found in (Strik and Cucchiarini 1999).

Reductions produce either different (centralised) phonemes (van Son and Pols 2003), fewer phonemes, or even fewer syllables. Reductions seem to affect the least informative speech portions: function words that are predictible from the context, idioms (e.g. *c'est-à-dire*, "that is"), morphological items (in particular endings), dates, discourse markers in spontaneous speech, etc. As the acoustic word models are obtained by phone model concatenation according to the pronunciation dictionary, appropriate variant descriptions are required.

Whereas the commonly adopted acoustic HMM (Hidden Markov Model) structure can implicitly account for speech lengthening, especially stemming from hesitation phenomena, and for parallel variants, pronunciations with a number of phonemes differing from the one specified in the pronunciation dictionary are poorly dealt with. Phoneme insertions may occur: the schwa in French is a well-known example. But the most problematic situation corresponds to missing phonemes. Acoustic phone models may implicitly capture limited reductions. For example, schwa models may simply represent the surrounding consonants, thus modelling schwa deletion. The drawback is then the loss of phonemic genericity: reductions beyond the context-dependent phone model scope (generally triphones) necessarily need to be explicitly represented in pronunciation dictionaries.

Sequential reductions are more or less described: we all know about phenomena such as *isn't it* or *it's* in English. In German, *ins* for *in das* ("in the") and *glaub's* for *glaube es* ("believe it") are also lexicalised forms which are respectively compulsory and optional. In French, similar phenomena occur, which are less lexicalised. If some determiners and pronouns have an obligatory reduced form before a word starting with a vowel (*l', d', s', m', qu'*. . . instead of *le/la, de, se, me, que*. . .), other reduced pronunciations are not reflected in writing: *il y a* ("there is") is most often uttered as *y a*, and *je ne sais pas* ("I don't know", /ʒənəsɛpa/) may have an acoustic realisation close to /ʃɛpa/, where the *ne* of the French *ne ... pas* negation is omitted and /ʒ/ and /s/ are combined to form /ʃ/. The scope of sequential reductions often goes beyond word boundaries: typically one or more short function words are involved. It is common practice to address this problem pragmatically by adding "glued words" in the pronunciation dictionary (e.g. *did you, want to*). Such word sequences are often improperly called compounds by the pronunciation modelling community as they are represented by single lexical entries with appropriate reduced pronunciations. The limits of the "phonemic-sequence" model for spontaneous speech are highlighted in (Greenberg and Chang 2000, Greenberg 2002), and variants are discussed with respect to syllabic structure and stress accent.

Given the large degree of variation in spontaneous speech (sociolinguistic, stylistic, situational and speaker-dependent), studies on very large corpora are needed to

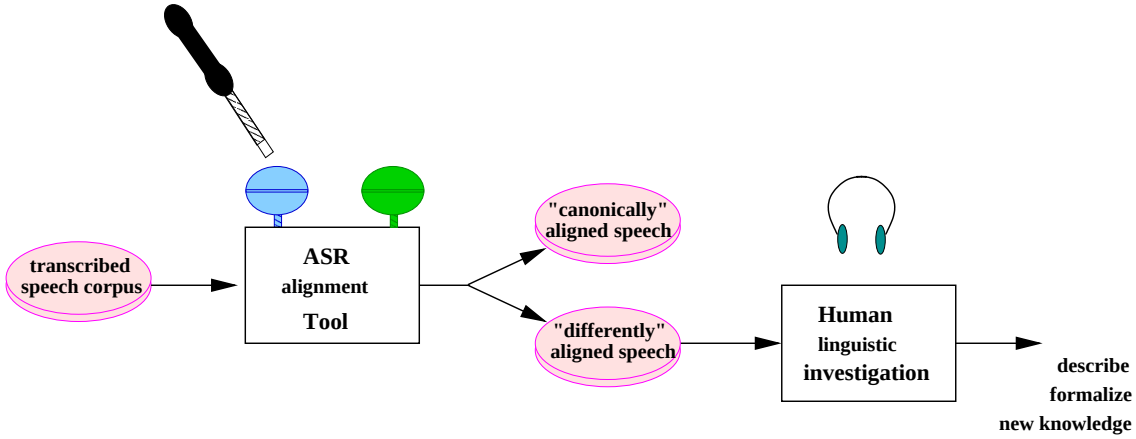

Fig. 1. Automatic speech recognizer regarded as a tool to automatically select parts of audio corpus, which deviate from an expected representation. These parts are of potential interest for more in-depth linguistic investigation.

provide a more extensive description of pronunciation variants. Recent progress in speech modelling provides the opportunity of using a speech recognizer to help analyse large acoustic corpora, and this is an important aspect of this contribution. Facing alternative pronunciations, it is not certain that the recognizer will prefer the same option as a human. Nonetheless, a given ASR system will consistently make the same decisions over the entire corpus, and can be parameterized to best fit the investigator's needs. In addition, the obtained outcomes may be better suited for ASR than the observations of an expert: whether for instance a schwa is deleted [1] is often unclear, since our perception is biased by our understanding. Figure 1 gives an overview of how the ASR tool can be used for linguistic purposes: depending on the configuration, unexpected pronunciations may be located. By exploring virtually unlimited speech corpora, more precise or even new knowledge may be produced. Our belief is that better formalised knowledge of pronunciation variant mechanisms will ultimately be helpful for pronunciation modelling and ASR systems.

In this paper, we focus on syllabic structures and their variation in a large corpus of French radio interview speech. The aim of this study is to detect sequential pronunciation variants (i.e. variants with different numbers of phonemes), and to relate them to syllabic restructuring. These types of variants are considered the most problematic in the present state-of-the-art of ASR systems since improper alignment reduces the acoustic model accuracy. Instead of limiting the linguistic representations to word and phoneme levels, as often is the case, a syllable level is introduced to describe sequential variants at an intermediary level between words and phonemes typically used by ASR systems. There are many reasons to consider syllables: syllables can be seen as basic speech production units and they play an important role in speech perception. Syllables also allow for a better overall temporal location, whereas phonemes may be hard to locate precisely in time, both for

[1] In this paper, *phoneme deletion* means *no distinct temporal segment* can be isolated for a given phoneme.

humans and for machines. Moreover phonotactic constraints condition the occurrences of shorter pronunciations, and the intermediate syllable level allows us to examine observed variants with respect to expected syllabic structures.

In the next section, syllabification rules of spoken French are described: we emphasise the effects of schwa and liaison on the syllabic structure. Section 3 briefly describes the speech corpus and outlines the general methodology of aligning speech transcripts to spoken syllables via written syllables. Section 4 explains the link between the word level and an intermediary written language-based syllable level. Section 5 focuses on the spoken syllable level, and Section 6 presents results on syllable restructuring using the W-syllable alignments. To compare with what happens in other languages, we will refer to recent studies by (Corbin 2003) and (Su and Basset 1998), which compare British English with Taiwanese Mandarin and French, by applying a common methodology based on a manually segmented corpus of 15 minutes (each language) of spontaneous speech.

2 Syllabification rules in French

In psycholinguistics, syllables are often considered as the information processing units of perceptual mechanisms, for acoustic-phonetic decoding (Pallier 1994). Various experiments (using a fragment detection task or other techniques) demonstrated that it is hard to focus one's attention on precise phonemes independently from syllables, suggesting that the latter are identified first. Evidence is also provided by the history of writing (syllabaries are older than alphabets), word games, children's speech and production errors such as slips of the tongue, which exhibit numerous constraints of syllable structure (Fowler et al. 1993). Finally, this unit is necessary for putting forward rules governing stress patterns.

Yet, syllabification, that is the segmentation of the spoken string into syllables, differs from one language to another: it depends on the linguistic communities' conventions, and a universal phonological theory does not exist (Vogel 1982). In English (the rhythm of which is not syllable-timed but stress-timed), researchers do not even agree on the number of syllables in words such as *communism, hour, real* (Ladefoged 1975); and ambisyllabic consonants (which could belong to either syllable simultaneously) as in *salad* are common (Cutler et al. 1986). In French, a consonant cannot constitute a syllable, and each syllable contains one and only one vowel.

Since Saussure (Saussure 1915), a hundred years ago, various theories have been proposed to account for the tendency of some consonant sequences to be split. According to the so-called Sonority Sequencing Principle, phonemes may be arrayed along a sonority scale according to their vowel-likeness, roughly corresponding to their aperture or degree of loudness (perceived intensity). Vowels are the most

sonorous type of phoneme, followed in turn by glides, liquids, nasals, fricatives and plosives. The Sonority Principle stipulates that phonemes located in the beginning of a syllable must have increasing sonorities and that syllabic edges are placed just before the minimum of sonority. But this contradicts another principle, that of maximum onset, in cases such as *costume*. The Maximum Onset Principle (MOP) stipulates that the syllabic boundary between two vowels separated by consonants is placed so as to maximise the number of consonants in the onset of the second syllable. These consonants, though, must constitute "legal" clusters, i.e. clusters which may appear at the beginning of a word in the language (Kahn 1976).

According to the Sonority Sequencing Principle, consonant clusters containing an /s/ followed by two or more consonants undergo a syllabic break after the latter (e.g. *obstruer* /ɔbs.tʁy.e/ in French). The tautosyllabicity of this /s/ with regards to the following consonant is controversial, since a French word may begin with what Italian grammar calls "impure *s*" (e.g. *sport*), without being disyllabic.

In French, it is traditionally assumed that, irrespective of the grammatical and orthographical word tokenization, each consonant belongs to the same syllable as the vowel immediately following. In particular, a syllabic break falls before an intervocalic consonant, even though "resyllabification" is not complete in some cases (Fougeron et al. 2002): in V#CV and VC#V contexts (where # denotes a word boundary), acoustic cues may enable distinctions such as *cas légal* ("legal case") vs *cale égale* ("equal hold"), both pronounced /kalegal/, since in the latter case the schwa is generally dropped in standard French.

The schwa, which may or may not be spoken (thus influencing the number of syllables), is one of the most intricate aspects of French phonology (Verney Pleasants 1956, Martinet 1971, Dauses 1973, Dell 1973, Walter 1976, Lacheret-Dujour and Péan 1994, Durand and Laks 2000). Consider the word *amener* ("to bring"): it has two or three spoken syllables (/am.ne/ or /a.mə.ne/) depending on whether the /ə/ is realised or not. Even if it enables a phonological opposition between words such as *pelage* ("coat" /pəlaʒ/) vs *plage* ("beach" /plaʒ/), the schwa vowel is generally optional.

Liaison is another complicated phenomenon which is directly linked to the syllabification process. Liaison consists in the realisation of a normally mute final consonant in the context of a following word which begins with a vowel. For example, the word sequence *les îles* ("the islands") pronounced /le/ and /il/ in isolation are pronounced /lezil/ in connected speech, and liaison results in a cross-word syllabification /le.zil/. Only a limited number of consonants are used for liaison: /z/, /t/, /n/, /ʁ/, /p/ – in order of frequency of occurrences. Cross-word syllabification makes word boundary recognition and thus lexical access perceptually more difficult.

How and when is liaison made? We are here in a ticklish field (Delattre 1966, Fouché 1969, Lucci 1983, Encrevé 1988, Eggs and Mordellet 1990, Léon 1993,

Fougeron et al. 2001), and there is no consensus to answer this question, which goes beyond the scope of this paper. Rules for mandatory, optional and forbidden liaison in French can be found in the literature: these rules mainly rely on morpho-syntactic information (e.g. liaison is mandatory within determiner-noun sequences). An ASR tool and a morpho-syntactically annotated speech corpus were recently used to automatically quantify the occurrences of liaison with respect to 20 such rules (Boula de Mareüil et al. 2003). Results are highly consistent with a priori explicited rules, and contribute to strengthen our belief that ASR systems are powerfull tools for large corpus analyses.

2.1 Syllabification procedure

Several syllabification algorithms for the French language exist. The average agreement of these algorithms on 38,549 tokens found in the phonetic transcriptions of the BDLEX (Pérennou and Calmès 1987) French lexicon is close to 96% (Goslin et al. 1999). Therefore, they may be viewed as relatively accurate predictors of human syllabification.

The syllabification procedure used in this study is part of the LIMSI grapheme-to-phoneme (G2P) converter GRAPHON+ (Boula de Mareüil 1997), whose pronunciation word error rate on several 30,000 word running texts is less than 1% (Boula de Mareüil et al. 1998). Syllabification can be optionally carried out after the G2P conversion proper. Silent pauses indicated by punctuation marks are considered as syllable breaks. Other syllabic breaks are obtained by using the first applicable rule among the list of rules given in Table 1. The same syllabification procedure can be applied with phonemic input (see Section 5).

2.2 Syllabic structures of standard French

Languages exhibit different syllabic structures. A study by Delattre (Delattre 1965) found the following proportions of consonants to vowels per syllable: 1.6 in French and Spanish, 2.1 in English and 2.5 in German. The syllabic structures of standard French, resulting from a manual syllabification on a corpus of spoken utterances (Wioland 1985), are reported in Table 2. The French language appears to prefer free (or "open") syllables, which account for roughly 80% of all syllables. With about 55% of occurrences, the CV type, which is the least marked syllable, is the most frequent. Liaisons as well as phenomena such as the use of *cet* for *ce* ("this"), *mon* for *ma* ("my") before a vowel contribute to this trend – increasing the number of free syllables and decreasing the number of syllables with an empty onset.

Nevertheless, colloquial French forms such as *d'jà* for *déjà* ("already"), *déj'ner* for

Table 1

Syllabification rules for French (left column), used by the LIMSI G2P converter (GRAPHON+, which can also syllabify a phonemic string given in input). The syllabic break is noted by a dot, ə stands for a maintained schwa; V={vowels}, L={liquids}, G={glides} O={obstruents: plosives, fricatives or nasals} and C={any consonant}. C{0;4} means 0 to 4 consonants. For each rule, the right columns show a word example and the effect of the syllabification rule on the example. The word parts for which the rule applies are underlined.

Syllabification rules			Word	Pronouniation	→	Syllabification
ə C{0;4}V	→	ə . C{0;4}V	*refroidi*	ʁəfrwadi	→	ʁə . fʁwadi
VV	→	V . V	*réalise*	ʁealiz	→	ʁe . aliz
VCV	→	V . CV	*image*	imaʒ	→	i . maʒ
VCGV	→	V . CGV	*studio*	stydjo	→	sty . djo
VOLV	→	V . OLV	*public*	pyblik	→	py . blik
VCCV	→	VC . CV	*objet*	ɔbʒɛ	→	ɔb . ʒɛ
VOLGV	→	V . OLGV	*emploi*	ɑ̃plwa	→	ɑ̃ . plwa
VCCGV	→	VC . CGV	*victoire*	viktwaʁ	→	vik . twaʁ
VCOLV	→	VC . OLV	*esprit*	ɛspʁi	→	ɛs . pʁi
VCCCV	→	VCC . CV	*expert*	ɛkspɛʁ	→	ɛks . pɛʁ
VCOLGV	→	VC . OLGV	*altruiste*	altʁɥist	→	al . tʁɥist
VCCCCV	→	VCC . CCV	*expier*	ɛkspje	→	ɛks . pje
VCCCCGV	→	VCC . CCGV	*exploit*	ɛksplwa	→	ɛks . plwa

déjeuner (“lunch”) and *m’sieur* for *monsieur* (“Sir”) may be observed, where the drop of an unstressed vowel leads to a resyllabification, transforming simple CV structures into more complex syllabic units.

While these phenomena are well known to linguists and speakers of French (e.g. (Léon 1993), their prevalence in spontaneous speech and consequences for ASR are not clearly established. We study these effects with the aid of a speech recognition system that is used to automatically label a large speech corpus, in order to carry out further linguistic analyses. By aligning the data with acoustic word models which allow for pronunciation variation (e.g. optional schwas and liaisons), the

Table 2
Syllable types of standard French using C (consonants) and V (vowels) classes (after (Wioland 1985)). The last column shows resyllabification if a schwa is produced. The total number of syllable types is thus reduced from 14 to 8 (in bold).

Syllables	Example	Pronunciation	Syllables with schwa
CV	*veau*	vo	**CV**
CCV	*gré*	gʁe	**CCV**
CVC	*masse*	mas{ə}	CV - CV
V	*eau*	o	**V**
CCVC	*grade*	gʁad{ə}	CCV - CV
CVCC	*test*	tɛst{ə}	**CVC** - CV
VC	*hâte*	ɑt{ə}	V - CV
CCCV	*strie*	stʁi	**CCCV**
CCVCC	*Brest*	bʁɛst{ə}	**CCVC** - CV
CCCVC	*strate*	stʁat{ə}	CCCV - CV
VCC	*ogre*	ɔgʁ{ə}	V - CCV
CVCCC	*filtre*	filtʁ{ə}	CVC - CCV
CCCVCC	*strict*	stʁikt{ə}	**CCCVC** - CV
CVCCCC	*dextre*	dɛkstʁ{ə}	**CVCC** - CCV

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).

In the following sections, the speech corpora and methodology used in this study are described, syllables are more extensively introduced and results are presented.

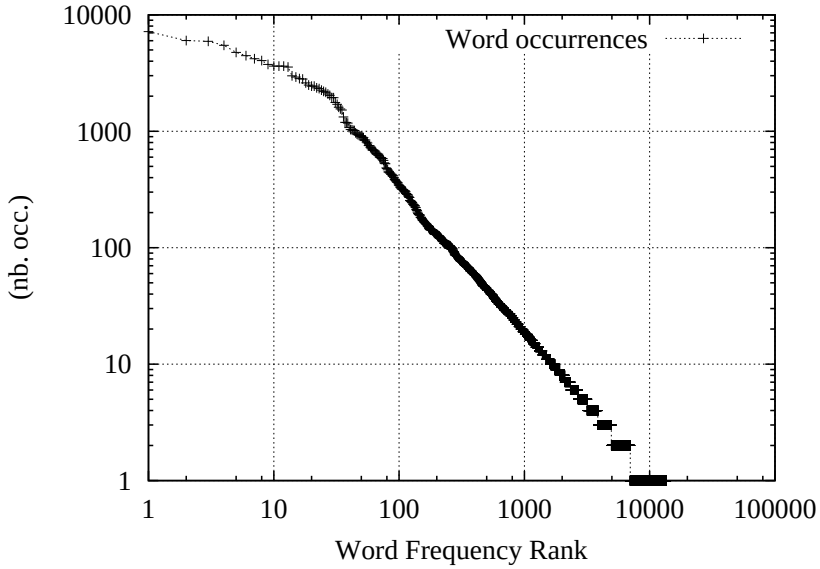

Fig. 2. Number of word occurrences as a function of word frequency rank in the reference transcriptions of the 30-hour radio interviews corpus (Zipf distribution).

3 Speech Corpus and Approach

The speech corpus used in this study contains thirty 1-hour shows of interviews involving most often one professional anchor speaker and one artist or politician, but some shows include more speakers. The speech is of studio quality and most of the speakers are native. On the whole, the speech style can be described as fluent, spontaneous and only partially prepared. Reference orthographic transcripts have been produced semi-automatically. In order to reduce the transcription cost, automatic transcripts were generated by the LIMSI system for French broadcast news (Gauvain et al. 1999) and then manually checked and corrected including hesitations and word fragments.

The corpus contains a total of approximately 245k word occurrences, with 13.5k distinct lexical entries. Figure 2 shows the number of word occurrences as a function of frequency rank – axes are in log scale. The evolution of the curve is roughly linear (Zipf's law): the Zipf distribution of natural language corpora shows that there are only few very frequent words and that a very large set of the observed words occur only a few times. Hence statistics about pronunciation variants cannot be drawn for these items. Analysing pronunciation variants with respect to syllable structures simplifies generalisations and descriptions of cross-word phenomena. In this corpus, only 12% of the lexicon (1600 words) have more than 10 occurrences. However the 150 most frequent monosyllabic words account for more than 60% of the corpus. These words are highly predictable and they may be recovered perceptively with only limited acoustic information.

The phonetic transcriptions were determined using the speech recognizer, tuned for

the alignment and sequential variant detection task. The orthographic transcription is used during alignment (instead of a language model during recognition). Since the pronunciation dictionary can contain multiple entries per word, the decoding space during alignment corresponds to a phone graph including all allowable pronunciations. For a given word, all pronunciations are equiprobable, independently of the pronunciation length (no insertion/deletion penalties). More details about the alignment procedure and its reliability can be found in (Adda-Decker and Lamel 1999), where we have shown that the number of detected variants decreases with the number of contexts in the acoustic models. In order to reduce the possibility that the models already incorporate some of the reductions of interest here, we have chosen to use a small set of 130 context-dependent continuous density hidden Markov models (HMMs) with Gaussian mixtures.

Our configuration allows a very large number of sequential variants: all sequential variants are included sytematically without concern as to whether or not they are linguistically relevant. This is part of the methodology: since only variants which are explicited before can be observed we need to overgenerate. In practice, only a small number of the numerous variants are observed, thus validating the approach. The most frequent variants were manually checked to verify whether or not they are consistent with what we know/expect: checking is done by listening to part of the aligned segments and looking at spectrograms.

With the help of this methodology, we aim to identify syllabic restructuring due to well-known phenomena such as schwa deletion and *liaison* in French. This also helps to partially validate the approach: if already known and described phenomena are automatically detected, other unexpected items should be considered with care. A further aim is then to identify less described deletion phenomena concerning vowels (i.e. syllable nuclei), consonants or even whole syllables.

Figure 3 shows a generic syllable representation composed of an onset and a rhyme. The onset is optional and, if present, may contain a single consonant or a consonant cluster. The rhyme has a mandatory nucleus which corresponds to a unique vowel in French. The coda, like the onset, is optional and may be composed of one or more consonants. The right part gives the structure of the most frequent syllable: the CV-type syllable with a single consonant onset and a rhyme limited to the vowel nucleus.

In the following, we describe common syllable restructuring phenomena, due to schwa elision and to the French liaison phenomenon. Figure 4 illustrates how the presence or absence of a schwa changes the syllable structure within the word *amener* (“to bring”): the left pronunciation has three open syllables V.Cə.CV; on the right, the schwa syllable (/mə/) is deleted, resulting in a two syllable pronunciation, with a closed first syllable VC.CV. A similar restructuring of open syllables can also occur across word boundaries: for example, the sequence *près de Paris* /pʁɛdəpaʁi/ (“near Paris”) corresponds to the syllable strucure CCV.Cə.CV.CV. Schwa dele-

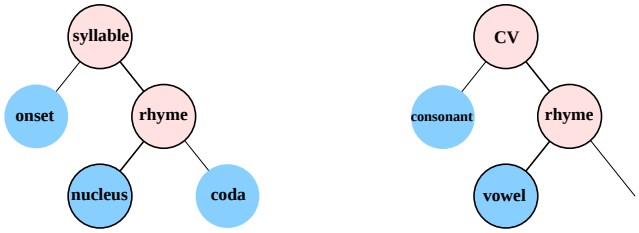

Fig. 3. **Left**: General syllable structure with optional consonant onset, syllable nucleus (which contains a unique vowel segment) and optional consonant coda. **Right**: example of CV syllable, which is the most frequent in French.

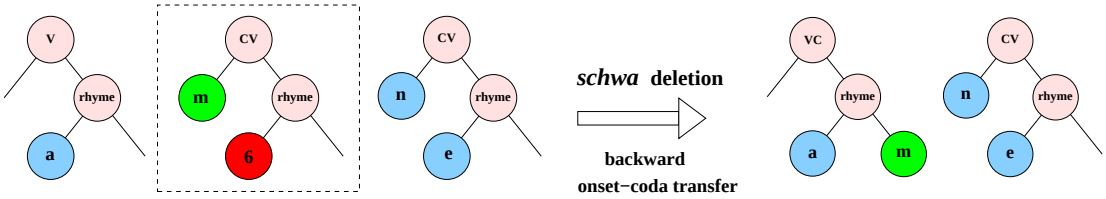

Fig. 4. Role of the schwa in syllable restructuring. The example shows the word *amener* (“to bring”) with 3 syllables on the left and 2 syllables on the right. The deletion-prone schwa syllable is surrounded by a dotted frame. When the schwa is not pronounced, the onset /m/ moves to the coda of the previous syllable.

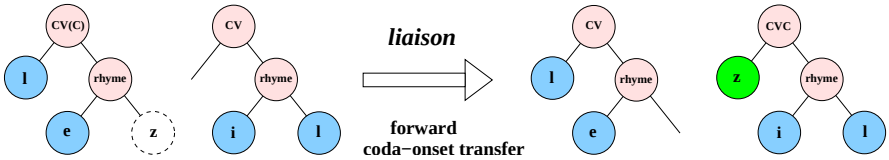

Fig. 5. Role of the liaison (/z/ here) in syllable restructuring. The example shows the word sequence *les îles* (“the islands”). The left part shows syllables of the 2 isolated words, and the potential liaison consonant is shown in a dotted circle. The right part shows the 2 words uttered together with the liaison consonant having moved to the onset of the following vowel – French tends to avoid empty onsets.

tion results in a CCVC.CV.CV structure (/pʁɛd-pa-ʁi/), where the onset /d/ of the weak syllable moves to the free coda position of the previous open syllable. The backward onset-coda transfer, respecting the Sonority Sequencing Principle, gives preference to closed syllables rather than to complex onsets.

Figure 5 shows an example of the French liaison phenomenon for the word sequence *les îles* (“the islands”), where the normally mute word-final consonant is pronounced in the context of a successor word starting with a vowel. In the left part, syllables are organised respecting the word boundaries; the right side corresponds to spoken syllables. Liaison reduces the proportion of empty onsets. Whereas liaison always entails a cross-word reorganisation, schwa deletion may occur within words, but statistically cross-word effects are the most important. The schwa-nucleus function words occur at the beginning of the Zipf distribution (see Table 3). These frequent cross-word effects motivated the distinction between written language-based syllables and spoken syllables.

Table 3
Schwa-nucleus function words with their number of occurrences and their corresponding rank in the Zipf distribution.

word	rank	#occur
de	1	7169
je	3	5917
que	4	4750
le	9	3748
ne	21	2438
ce	25	2201

To study French syllable structure variation, syllables are first defined at two levels: one corresponding to a written language word level and the other to a spoken language sentence level. Figure 6 gives an overview of these W-syllables (also standing for "word syllables") and S-syllables (also standing for "sentence syllables"). Starting with the orthographic transcription of each word, a canonical pronunciation is produced. Each phonemic word is then split into syllables using the GRAPHON+ syllabification rules (see Table 1). This results in a canonical W-syllable transcription, where all word boundaries match syllable boundaries. A syllable formalism is defined using full and partial syllables (zero vowel). As already seen, the most frequent words are short function words which are in general monosyllabic: *de* ("of"), *est* ("is"), *je* ("I"), *que* ("that"), *et* ("and"), *vous* ("you"), *la* ("the")... The need for partial syllables arises from the presence of short function words in written French, which are reduced to a single consonant: *c'* ("this"), *l'* ("the"), *j'* ("I"), *n'* ("not"), *d'* ("of"), *qu'* ("that"), *m'* ("me"), *s'* ("oneself"). For example, the written determiners *le* or *la* ("the") reduce to *l'* before a vowel-initial word (*l'artiste* "the artist") [2] . There are roughly ten of these entries which are reduced to one consonant: they have to be combined to the vowel of the following word to form a full (admissible) speech syllable. Albeit limited in number, these entries are frequent in the language, accounting for about 5% of the words in the corpus. Spontaneous speech can also give rise to partial or degenerate syllables (without nuclei) in word fragments limited to consonant speech segments.

A W-syllable dictionary covering the whole corpus is built, and pronunciation variants added to the canonical phonemic form. W-syllables allow a straightforward link with the lexical level, which is the modelling unit in state-of-the art speech recognition systems. In spoken language however, syllable boundaries often occur at different locations from word boundaries, as we have just seen. The second syl-

[2] These reduced items which are glued to the successor word in written text are considered as separate lexical items. Text normalisation inserts a blank after the apostrophe, thus avoiding vocabulary explosion and keeping a consistent lexical unit definition: the sequences *le politicien* ("the politician") and *l'artiste* ("the artist") both comprise 2 words in French.

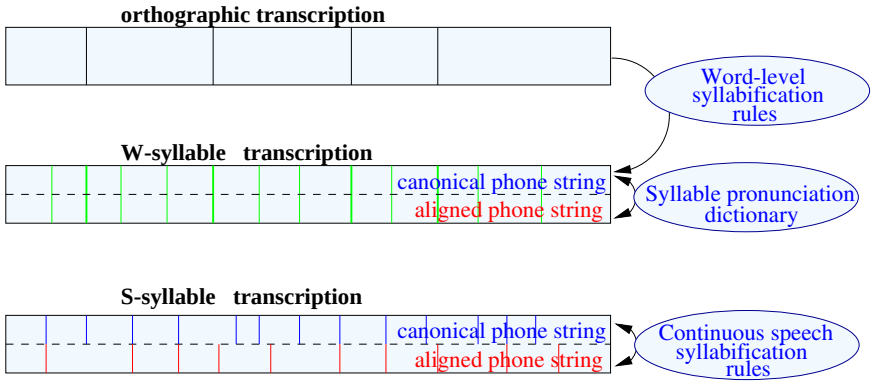

Fig. 6. Overview of the method with different levels of representation.

Table 4
Examples of lexical entries transcribed into maximal-length canonical phonemic strings, which are then split into W-syllables according to the syllabification algorithm.

lexical entry	MLC string	W-syllables
une	ynə	y nə
développer	devəlɔpe	de və lɔ pe

lable level aims at representing the spoken syllables (S-syllables), ignoring word boundaries. S-syllables are generated by applying the GRAPHON+ syllabification rules to the complete phonemic chains, without knowledge of word boundaries. For both W-syllables and S-syllables, the canonical forms are compared to the aligned forms and a description of syllable structure variation is proposed. The aligned W-syllables are then analysed to measure vowel deletions; consonant deletions with respect to their positions in the syllable; and syllable deletions with respect to their positions in the word. Whereas the absolute figures depend on the system's accuracy, comparisons may reveal valuable information about general syllabic restructuring mechanisms.

4 Written language syllables

4.1 Word-level syllables

Each lexical entry is phonemically transcribed into a maximal-length canonical (MLC) phonemic string. By maximal-length, we mean that all possible phonemes are supposed to be pronounced, in particular schwas. For example, the word *amener* ("to bring") has the pronunciation /amәne/. This MLC string is then split into **written language syllables** or W-syllables, using the syllabification algorithm described in Section 2 (see Table 4).

There are about 370k W-syllables in the corpus, with 1570 distinct W-syllables. It should be noted that the most common 1050 syllables account for 99.8% of the corpus. The full W-syllable list includes a large number of rare events arising from foreign proper names, word fragments (truncated words of spontaneous speech) and from errors (mainly transcription spelling errors and subsequent grapheme-to-phoneme conversion problems). Figures concerning the occurrences in the corpus of the different syllable types are given in Table 5: the observed syllable structures are identical to those presented in Section 2, with an additional C class corresponding to the partial syllables (see the *W-syll isol* column). The partial syllable C is mainly due to syllabification performed on isolated word syllables. If these partial C syllables are merged with the onset of the following syllables, figures slightly change, with a reduction of the V structure and an increase of the CV structure (see the *W-syll cont* column). In both the *W-syll isol* and *W-syll cont* columns, CV syllables represent roughly 60% of the corpus, the V syllable around 13%, with CCV and CVC occurring about 10% each. There are 46k occurrences of W-syllables which can produce a liaison consonant: only 11k of them are in a right vowel context. Column *W-syll+liaison* displays percentages when liaison consonants are shifted to the onset of the following vowel-initial syllable. The six CV, V, CCV, CVC, VC and CCVC syllable types account for 99% of the corpus in the last two columns.

4.2 W-syllable pronunciation dictionary

A canonical W-syllable transcription has been derived from the lexical transcription (see first two lines of example in Table 7). In order to align the W-syllables with the acoustic signal, a pronunciation dictionary with variants is introduced. As we are mainly interested in reduction phenomena (inducing a smaller number of phonemes than theoretically expected), any shorter phone sequence included in the MLC form is allowed (see Table 6). In addition to these variants an optional schwa is added to each pronunciation, and each syllable may be reduced to a simple schwa.

4.3 Optional W-syllables

In the W-syllable pronunciation dictionary, each entry can be reduced to one phoneme. Beyond these reductions, we want W-syllables to be optional: if a W-syllable has not been uttered, it should be possible to skip it. For example the word-final /tʁə/ syllable of the word *orchestre* (/ɔʁ.kɛs.tʁə/) may completely disappear in a sequence such as *orchestre de chambre* (“chamber orchestra”): Alignments are carried out using a W-syllable graph corresponding to the W-syllable transcription, where every other syllable may become optional as shown in Figure 7.

Table 5
Different W-syllable types observed in the corpus with their percentage of occurrence. The partial syllable C is mainly due to syllabification carried out on isolated words (*W-syll isol* column). The *W-syll cont* column gives corrected full syllable percentages, where partial syllables are glued to the following syllables. Finally, the *W-syll+liaison* column corrects liaison syllabification. ϵ means that the percentage is less than 0.05.

syllable type	*W-syll isol*	*W-syll cont*	*W-syll+liaison*
CV	57.6	63.2	68.2
V	14.6	12.0	9.7
CCV	9.8	10.5	10.7
CVC	9.2	10.3	7.9
C	4.3	-	-
VC	2.6	2.0	1.1
CCVC	1.0	1.1	0.8
CCCV	0.5	0.5	0.5
CVCC	0.3	0.3	0.3
VCC	0.2	0.1	0.1
CCVCC	ϵ	ϵ	ϵ
CVCCC	ϵ	ϵ	ϵ
CCCVC	ϵ	ϵ	ϵ
VCCC	ϵ	ϵ	ϵ
CCCVCC	ϵ	ϵ	ϵ

Table 6
Excerpt of the W-syllable pronunciation dictionary (the left side corresponds to the W-syllable and the right part to the optional smaller length pronunciations).

W-syll.	pronunciations
sa	sa s a saə sə aə ə
tʁi	tʁi tʁ ti ʁi t ʁ i tʁiə tʁə tiə ʁiə tə ʁə iə ə

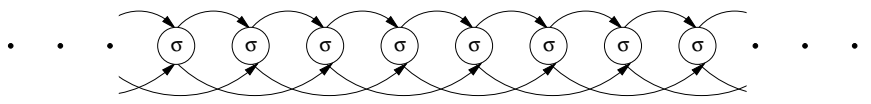

Fig. 7. Alignment graph built from reference transcriptions. Each node corresponds to a W-syllable and every other syllable is optional.

Table 7
Excerpt of a sentence with its lexical transcription, the word syllable transcriptions (canonical and aligned W-syll), and sentence syllables – canonical S-syll from the MLC phonemic sequence and aligned S-syll from the aligned sequence. The last column gives the number of syllables for the different representations. In the W-syllable /ɛ(t)/, corresponding to the word *est* (“is”) (t) indicates a /t/ liaison if the right context is favourable.

Lexical	*je*	*pense*		*que*	*c'*	*est*	*de*	*la* ...	#syll
canonical W-syll	ʒə	pɑ̃	sə	kə	s	ɛ(t)	də	la ...	8
aligned W-syll	ʒə	pɑ̃	s	kə	s	ɛ	d	la ...	8
canonical S-syll	ʒə	pɑ̃	sə	kə	sɛ		də	la ...	7
aligned S-syll	ʒə	pɑ̃s		kə	sɛ		dla	...	5

5 Spoken language syllables

This section is devoted to the description on a spoken syllable basis. Here, we do not consider word boundaries during the syllabification process. The MLC phonemic strings and the corresponding aligned strings are syllabified using the GRAPHON+ rules described in Section 2, thus producing S-syllables. Table 7 shows the beginning of an example sentence with its corresponding W- and S-syllables. Whereas in the W-syllable representation, each canonical W-syllable may be either skipped (W-syllable deletion) or aligned to one of the possible pronunciations of the dictionary (which may be reduced to a partial syllable), the S-syllable represents full syllables irrespective of word boundaries. In this example, no W-syllable deletion occurred (8 W-syllables), but for the S-syllable level, there are 7 canonical syllables and only 5 aligned syllables due to schwa deletions.

The aligned S-syllable /dla/ does not occur in canonical W-syllable transcriptions and more generally in syllabified lexica. But /dla/ corresponds to a common OLV (obstruent-liquid-vowel) syllable structure and can hence be considered as a spoken language specific syllable.

Syllabification of the MLC phonemic strings produces about 350k S-syllables, which is roughly 5% less than the W-syllable syllabification and can be explained by the absence of partial syllables. Using the aligned phonemic string, a total of 300k S-syllables is measured, corresponding to a 15% deletion rate. The number of distinct syllables increases from 1560 (canonical W-syllables) to about 1700 sentence level syllables (canonical S-syll). The additional syllables are due to cross-word syllabification, and consist to a large amount in distinct CVC syllables. Cross-word syllabification on the aligned sequences (aligned S-syll), introduces more than 6,000 distinct syllables. Figure 8 shows syllable occurrence counts for both the canonical W-syll transcriptions as well as the aligned S-syllables. Whereas the aligned S-syllable curve has a central part which is almost linear, Zipf’s law does

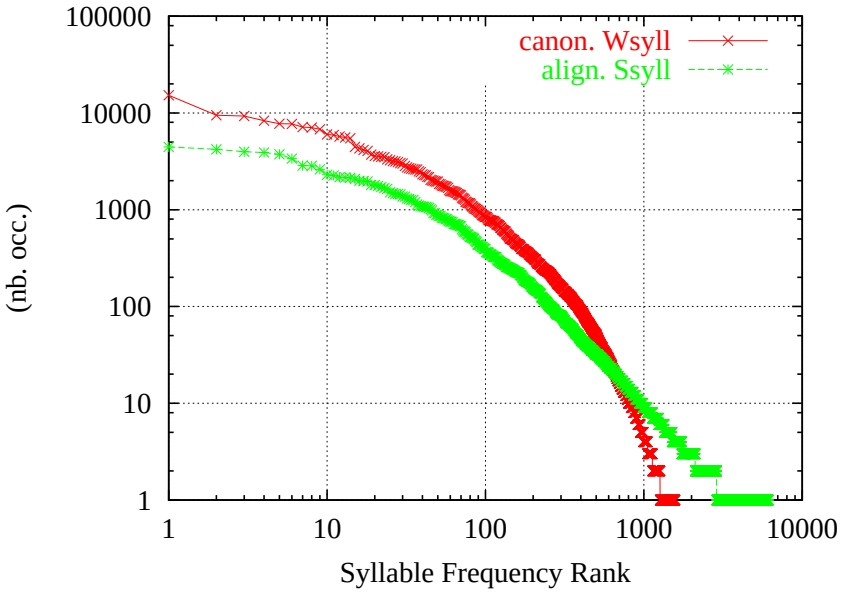

Fig. 8. Number of occurrences of the canonical W-syllables and the aligned S-syllables as a function of syllable frequency rank.

not apply for the canonical syllables. Pronunciation variation statistics can be collected for almost any canonical W-syllable. The alignment of non-standard phone sequences (as compared to standard MLC sequences) is the explanation for the larger number of distinct S-syllables. 1,100 aligned S-syllables occur more than 10 times (with a coverage of 94.9%). In comparison, for W-syllables, we have only 800 sylables with a frequency of occurrence greater than 10 (but a coverage of 99.3%). Many of the observed S-syllables are in common with the W-syllables. But cross-word resyllabification allows the creation of "new" syllables, i.e. those that do not occur in isolated French words. For instance, there are 27 syllables starting with /ʒl/, which corresponds to a resyllabification of word sequences like *je le* (“I ... so”). One of these syllables is /ʒlɛs/, which may arise from the word sequences *je l’espère* (“I hope so”), *je laisse* (“I let”). . . New syllables are also due to cross-word syllabification involving word fragments.

Table 8 shows syllable structures of S-syllables (both S-canon and S-align). We can notice that the number of closed syllables increases from roughly 10% for the canonical S-syllables to more than 16% for the aligned spoken language syllables. The most frequent closed syllable structure is CVC (11.6%). The more complex syllables (CCVC, CVCC) are significantly more frequent for the aligned S-syllables than for the standard canonical S-syllables. The most common open CV syllable structure represents 60% of the aligned S-syllables. The overall percentages measured for the main syllable types remain nonetheless similar to the percentages measured for the W-syllables (see Table 5). The relatively high V syllable rate (14.6%) obtained for isolated W-syllables is reduced here to about 12%. A smaller rate of V syllables could have been expected, given the cross-word context and measured vowel deletions. Investigating the automatic alignment and syllabifi-

Table 8
Most frequent spoken syllable types in French for the canonical S-syll column (from the syllabified MLC sequence)and the aligned S-syll column (from the syllabified aligned phone sequence), as observed in the corpus. Closed syllables are more frequent in the aligned S-syllables.

syllable type	%canonical S-syll	%aligned S-syll
CV	67.3	60.4
V	11.8	12.5
CCV	10.5	9.2
CVC	**7.6**	**11.6**
VC	**1.1**	**1.6**
CCVC	**0.6**	**1.4**
CVCC	**0.4**	**1.4**
CCCV	0.4	0.4

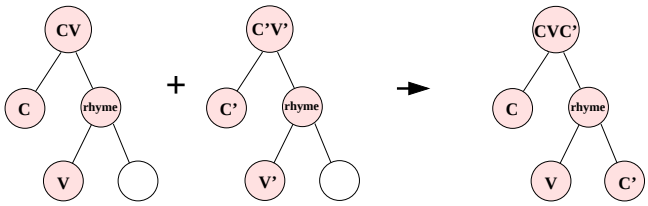

Fig. 9. Resyllabification producing closed CVC' syllables due to V' vowel deletion.

cation results, we could observe that simple vowels (often the schwa) are sometimes aligned with unclearly uttered syllables (e.g. repetitions of word fragments): such alignments produce V syllables.

6 Analysis of alignment results

Syllabic restructuring can result from vowel and/or consonant deletions. Figure 9 illustrates a typical cross-word resyllabification of two consecutive CV syllables into a CVC syllable (e.g. **vous re***cherchez* ("you are looking for") /**vu.ʁə.**ʃɛʁ.ʃe/ → [**vuʁ**.ʃɛʁ.ʃe]) due to schwa deletion. This kind of resyllabification is very common, as indicated by the CVC figures in Table 8, which show a significant increase of 4% (absolute) from canonical to aligned S-syllables.

Different questions may arise during alignment analysis: Do consonants disappear more than vowels? Are some types of phonemes more deletion prone? Since the schwa vowel is known to be highly instable and schwa syllables can be considered as weak syllables, do they disappear significantly more than other syllables? Do

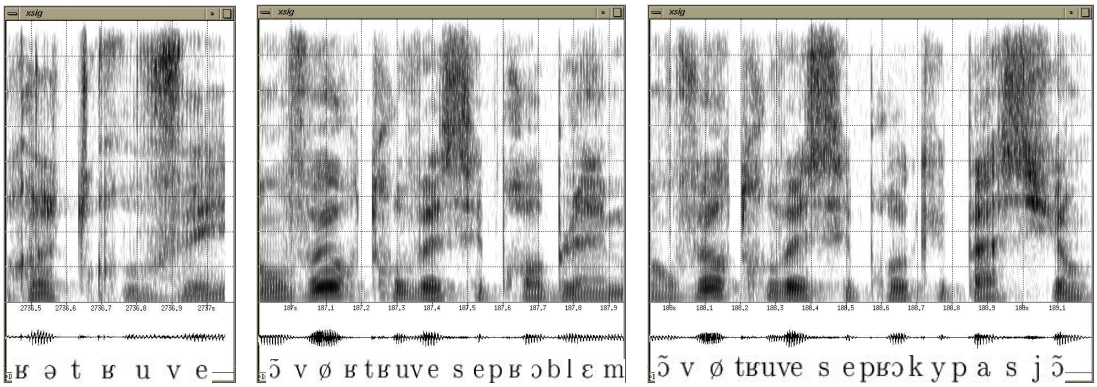

Fig. 10. Spectrograms illustrating /ə/ and /e/ vowels and /ʁə/ syllable deletions.
Left: full pronunciation of the word **retrou***ver* /**ʁə.tʁu**.ve/
Middle: schwa deletion entailing resyllabification:
on **veut retrou***ver ses problèmes* /õ.**vø.ʁə.tʁu**.ve.se.prɔ.blɛm/ → [**vøʁ.tʁu**]
Right: /ʁə/ syllable deletion and /e/ vowel deletion
on **veut retrou***ver ses* **préo***ccupations* /õ.**vø.ʁə.tʁu**.ve.se.**pre.ɔ**.ky.pa.sjõ/ → [**vø.tʁu**] [**prɔ**]

the monosyllabic function words with a schwa behave as other schwa syllables (in particular word-final schwa syllables)? Do the disappearing syllables more often correspond to function words than parts of polysyllabic words? What is the most deletion-prone position of the syllable: word-initial, word-internal or word-final?

In the following, we investigate vowel, consonant and syllable deletions with respect to the W-syllable representation. In contrast to S-syllables, W-syllables guarantee a straightforward link with the lexical level, potentially providing insight about word modelling problems due to the cross-word syllabification.

6.1 Deleted vowels

The global deletion rate measured for vowels is 15%. This rate drops to 6% if schwas are excluded. High deletion rates are observed for /ɔ/ (10%), /u/ (8%), /ɛ̃/ (7%), /ɛ/ (7%), /i/ (5%), /a/ (5%).

While vowel deletion may occur at **VV sequences** within words, it is more typical at word boundaries. Vowels are also prone to deletion in **N (nasals)** and **L (liquids) contexts**. Another consonant context favouring vowel deletions corresponds to **C_C**, where left and right C phonemes are **equal or close**: *six cents* (“six hundred”) /sisã/ may be reduced to [ssã], and *si c’était* (“if it were”). /sisetɛ/ to [ssetɛ]. Even if the underlying vowel can be identified, there is no distinct segment in the acoustic signal. In Table 9, some examples of observed vowel deletions are reported. Likewise, the spectrograms in Figures 10 and 11 illustrate that the vowels are indeed missing in the acoustic signal, and that this is not an artifact due to the automatic alignment procedure.

Table 9
Examples of vowel deletion within words and across words in different contexts.

word	**VV**	→	**V**
extraordinaire	/ɛkstʁ**a**ɔʁdinɛʁ/	→	[ɛkstʁ**a**ʁdinɛʁ]
mais enfin	/m**ɛ**ɑ̃fɛ̃/	→	[m**ɑ̃**fɛ̃]
j'ai été	/ʒ**ɛe**te/	→	[ʒ**e**te]
word	**CVC**	→	**CC**
left or right C nasal or liquid			
cinéma	/si**nem**a/	→	[si**nm**a]
comment	/**kɔ**mɑ̃/	→	[**km**ɑ̃]
personnel	/pɛʁ**sɔn**ɛl/	→	[pɛʁ**sn**ɛl]
voulait	/**vul**ɛ/	→	[**vl**ɛ]
left and right C equal or close			
il allait	/i**lal**E/	→	[i**ll**E]
vous voulez	/**vuv**ule/	→	[**vv**ule]
carrière artistique	/kaʁjɛ**ʁaʁ**tistik/	→	[kaʁjɛ**ʁʁ**tistik]
musicien	/my**zis**iɛ̃/	→	[my**zs**iɛ̃]
miscellaneous - frequent word sequences			
c'est à	/**sɛt**a/	→	[**st**a]
c'est pas	/**sɛp**a/	→	[**sp**a]
je sais pas	/**ʒəsɛp**a/	→	[**ʃp**a]

Vowel deletion seems to be more common in French than in British English, where stressed vowel deletion is avoided, and in Taiwanese Mandarin, where the vowel bears the tone ((Corbin 2003) and (Su and Basset 1998)). This is even more obvious in the most frequent (function) words, whereas the tendency in British English is to drop the initial or final consonant in monosyllabic words such as *him, but*. If the central schwa-vowel deletion seems to be shared by the three languages, all vowels can undergo deletion in French. This is not the case in English and Chinese, where low vowels appear to be more resistant, more preserved than the other vowels.

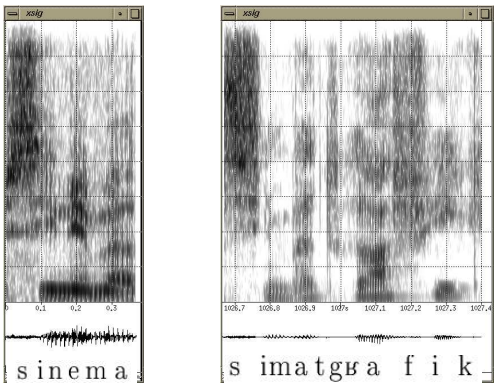

Fig. 11. Spectrograms illustrating /ne/ syllable and /o/ vowel deletions.
Left: full pronunciation of the word *cinéma* /si.ne.ma/
Right: /ne/ syllable and /o/ vowel deletions.
cinématographique /**si.ne.ma.to**.gʁa.fik/ → [**si.mat**.gʁa.fik]

6.2 Deleted consonants

The average consonant deletion rate is 13%. The deletion rates in onset and coda position differ significantly: in the onset position the consonant deletion rate is 11%, whereas the coda consonant deletion rate is close to 30%. Our results are in agreement with previous work by Duez (Duez 2003) and comparable to studies of English (Greenberg and Chang 2000). Mandarin of course is different since consonants can only be syllable-initial (if final nasals are considered as part of the vowels). In syllable-initial position, the most deleted consonant is the voiced fricative /v/ (20%), occuring in frequent words such as *vous* (“you”), *avec* (“with”), *avez*, *avait* (“have, had”). Liquids and glides are also often deleted in this position (12% to 17%). Deletion rates are lowest for unvoiced fricatives.

Liquids account for more than 35% of deletions, whereas they represent 1/4 of consonants. As examples of liquid deletions, the word *film* is often pronounced as [fim], and the syllable /pli/ as in *compliqué* (“complicated”) is aligned with [pi] in close to 25% of occurrences. The truncation of words such as *montre* (“watch” or “show”) and *prendre* (“to take”) resulting from the drop of the final liquid is a well-known phenomenon in spoken French. The analysis of our data confirms that, for words in *-tre* and *-dre* preceding a consonant, the pronunciations [t] and [d] (rather than [tʁ] and [dʁ] respectively) are preferred. After the schwa elision in this context, the liquid falls in 240 occurrences, and is maintained together with the plosive in 170 occurrences. This way, too massive a violation of the three consonant law is avoided – see (Durand and Laks 2000). The drop of the liquid also occurs in *il/ils* (“he/they”) before a consonant. In this context, roughly 30% of these tokens are aligned with the pronunciation [i] (300 occurrences) rather than [il] (700 occurrences). In the XVIIth century, the pronunciation [ifo] for *il faut* (“it is necessary

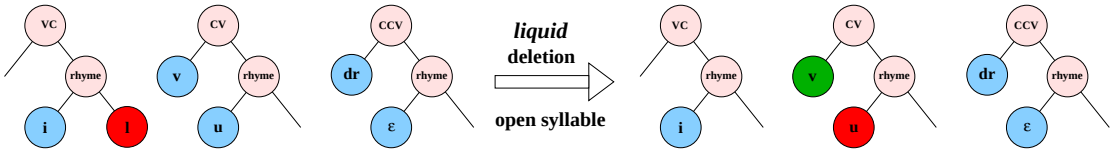

Fig. 12. Example *il voudrait* (“he’d like”). **Step 1: Liquid deletion** in coda position resulting in *i’ voudrait* and an open syllable which favours Step 2.

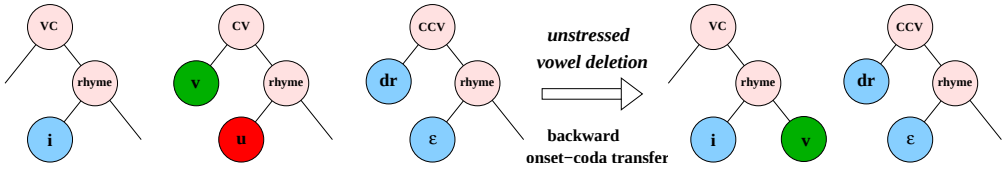

Fig. 13. **Step 2: unstressed /u/ vowel deletion** and syllable restructuring. The mechanism is identical to the schwa restucturing mechanism.

to”) was even considered as the norm, whereas [ilfo] was stigmatised as pedant (Walter 1988). Things have changed since then, but the current tendency could lead back to the XVIIth century pronunciation.

In British English, alveolar consonants represent a majority of the deleted consonants, due to the loss of the [t] in auxiliaries such as *don’t*, and the loss of the [d] in *and*, especially before a homorganic consonant. In Taiwanese Mandarin, the tendency is more scattered, but voiced fricatives are more likely to be deleted than unvoiced consonants. This is well documented (Su and Basset 1998) and holds true for French, in our corpus too (19% vs 12% deletion rates for voiced/unvoiced plosives and fricatives).

Figures 12 and 13 show how liquid and unstressed vowel deletions are able to transform the sequence /ilvudʁɛ/ (*il voudrait* “he would like”) which corresponds to a VC.CV.CCV sequence into a pronunciation such as [ivdʁɛ] resulting in a VC.CCV sequence. In English, the reduction of “he would” (CV.CVC) is even lexicalised: “he’d” (CV.CVC → CVC). As this lexicalisation example witnesses, English allows for similar syllable restructuring.

The spectrograms in Figure 14 illustrate deletion of /v/ in syllable onset position. Both are examples of the word *avec* (“with”) preceded by a vowel (V), a context in which vowel deletion is likely to occur. In these examples, the VV sequence is reduced to a single vowel, with the /a/ of *avec* being more or less deleted. In the right spectrogram, *avec* is reduced to a minimum perceptive cue of a velar plosive.

6.3 Deleted syllables

Using the W-syllable alignment with the optional syllable graphs, 6% of W-syllables are skipped (deletions). A small part of these missing syllables can be attributed to

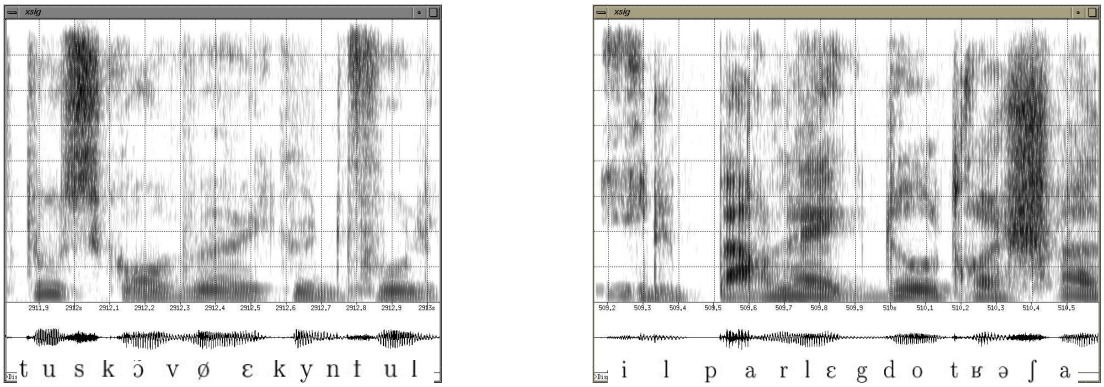

Fig. 14. Spectrograms illustrating cross-word reductions involving more than a syllable. Examples show the word *avec* (“with”) /avɛk/ in a left vowel context.
Left: *tout ce qu’on veut avec une foule* /tus.kɔ̃.**vø.a.vɛk**.yn.ful/ → [**vø.ɛk**]
Right: *il parlait avec d’autres chats* /il.paʁ.**lɛ.a.vɛk**.do.trə.ʃa/ → [**lɛg**]

Table 10
Examples of W-syllable deletions. The left part gives examples of deleted CV syllables corresponding to monosyllabic function words with a ə vowel. The right part shows non-schwa V syllable deletions from short function words. #del gives the number of deleted syllables. For each syllable, the deletion rate is given (between 8.5% and 18.3%).

word	rank	W-syll.	#del	%	word	rank	W-syll.	#del	%
ne	21	nə	569	23.5	*ai*	30	ɛ	484	22.3
de	1	də	1054	14.9	*à, a*	10,17	a	1137	16.8
le	9	lə	437	11.8	*y*	33	i	231	14.4
je	3	ʒə	485	8.2	*une*	22	**y** nə	317	10.5
que	4	kə	432	8.0	*est*	2	ɛ	598	10.1

transcription errors: for spontaneous speech, there are segments which are difficult to transcribe at a word level because of unclear articulation, hesitation and repetition of function words or word fragments. However, an important part of these deletions corresponds to well-known linguistic phenomena. In particular, it is widely acknowledged that phoneme and syllable deletions affect frequent (function) words more often than rare (lexical) items (Malmberg 1985). Among the observed deletions, 40% (9k occurrences out of 24k) correspond to schwa syllables. Of the remaining deleted W-syllables 30% (7k) are V syllables, mainly /a, ɛ, e, y, ɛ̃, i, ɔ, o, ɑ̃, u/. For the most part these syllables correspond to function words (2.6k) or word-initial syllables (3.3k). Table 10 shows that deletion rates for function words with a schwa are similar to those measured for other function words: function words are prone to deletion whatever the vowel identity. Partial syllable deletions (which are simple consonant deletions) also come from function words: the most often omitted partial syllables are *n’, l’, d’, qu’, j’* (Table 11 left). Deletions of more complex syllables usually correspond either to frequent words *il (9%), non*

Table 11
Examples of **partial** W-syllable deletion (left) and of final **CCV** W-syllable deletion. The deletion rate is particularly high for the negation *n'* (24.8%). The word-final **CCV** syllables (in bold) are prone to deletion or at least to reduction.

word	rank	W-syll.	#del.	%	carrier word	W-syll.	aligned	#del	%
n'	26	n	534	24.9	*exemple*	ɛg zɑ̃ **plə**	ɛg zɑ̃ -	68	18.8
l'	14	l	250	8.5	*capable*	ka pa **blə**	ka pa -	94	12.6
qu'	29	k	159	8.3	*être*	ɛ **trə**	ɛ-	210	6.4
d'	28	d	159	7.9					
c', s'	12, 66	s	306	7.2					

(7%), vous (6%), oui (6%), mais (5%) ("he, no, you, yes, but") of VC and CV types, or to word endings in CCV (Table 11 right).

To measure the link between syllable deletion and syllable position within the word, the corpus has been partitioned into 4 subsets of words with a given number of syllables: the monosyllabic, disyllabic, trisyllabic and polysyllabic (> 3 syllables) word sets. Table 12 shows W-syllable deletion rates for the 4 subsets. In French, the word-final syllable, if not a schwa syllable, bears lexical stress, at least in a prosodic phrase final position (Delattre 1965). In the adopted MLC pronunciation formalism, many words have a final schwa resulting in a weak final syllable. In each of the 4 subsets, words have been separated depending on the last vowel being a schwa or not. Monosyllabic words roughly correspond to the most frequent function words, at least for the schwa set. We can observe that the deletion rates are highest for these monosyllabic function words, with the deletion rate of schwa function words nearly twice the rate of other monosyllabic words. Whereas in the latter subset the deletion rates are above 10% for V syllables (cf. Table 10), more complex syllable structures are less deletion prone. The deletion rates of final schwa syllables (11.7-14.3%) are very close to those of the monosyllabic function words (11.3%). This suggests that the monosyllabic schwa function words behave as other schwa syllables (in particular word-final schwa syllables). Even if part of the automatically found deletions may be be related to modelling problems, others are clearly due to syllables missing in the acoustic signal. As expected, final syllables (respectively "penultimate" syllables for schwa-final words) are the most resistant, as shown in bold in Table 12. The lowest deletion rates are for the "penultimate" syllables in schwa-final words: this position is less affected by cross-word coarticulation than the non-schwa final syllables. Measured deletion rates are somewhat higher for word-initial syllables (4.4-6.6%) than for word-internal syllables (3.7-4.2%). This result deserves some further investigation. The V syllable structure which is frequent in word-initial position, but only seldom observed word-internally, is the main explanation here. For example consider the trisyllabic word set with the 5.5% deletion rate in initial position. Removing V syllable initial words from the trisyllabic word set, the deletion rate drops to 2.9% (233 syllable deletions in word initial position out of 8013). In

Table 12
Percentages of measured Wsyll deletions in mono-, di-, tri- and poly-syllabic words. For each N-syllabic word set, schwa final and non-schwa final subsets are considered separately. Deletion rates are measured with respect to the total number of syllables in the specified position. High word initial deletion rates are mainly due to V syllables. More complex syllable structure in initial position are less deletion prone than word-internal syllables.

monosyllabic word set						
	non schwa			schwa		
	#occur	#del	%del	#occur	#del	%del
	122,895	7,771	6.3	28,373	3,206	11.3

disyllabic word set						
	non-schwa final			schwa final		
position	#occur	#del	%del	#occur	#del	%del
initial	32,232	2,122	6.6	24,063	1,063	**4.4**
final	32,232	816	**2.5**	24,063	2,820	11.7

trisyllabic word set						
	non-schwa final			schwa final		
position	#occur	#del	%del	#occur	#del	%del
initial	11,853	652	5.5	10,303	531	5.2
penultimate	11,853	497	4.2	10,303	157	**1.5**
final	11,853	273	**2.3**	10,303	1,478	14.3

polysyllabic word set						
	non-schwa final			schwa final		
position	#occur	#del	%del	#occur	#del	%del
initial	5,026	286	5.7	4,926	221	4.9
internal	6,010	225	3.7	5,987	223	3.7
penultimate	5,026	224	4.6	4,926	78	**1.6**
final	5,026	128	**2.5**	4,926	652	13.2

contrast the deletion rate for the V syllables is 11% (419 out of 3840). For the disyllabic set the 6.6% word-initial deletion rate corresponds to 3% for non-V syllables and 13.5% for V syllables. Typical examples, where initial vowel deletion occurs in disyllabic words are *avait, avez* (“had”, “have”), *enfin* (“at last”), *avec* (“with”), *alors* (“then”), and for trisyllabic words *aujourd’hui* (“today”), *écoutez* (“listen”) in sequences such as *vingt ans aujourd’hui, au fond aujourd’hui, je leur dis écoutez, je l’ai rencontré écoutez j’avais dix ans*.

Results tend to show that phone deletions depend on the position in the syllable and that the acoustic realisation is correlated with word-position. In future modelling of context-dependent phones for ASR, more elaborated contexts can be conditioned, not only on neighbouring phones, but also on the position of the phone in the syllable and the position of the syllable in the word.

7 Conclusions and perspectives

Whereas we all can cite examples of more or less severe reduction phenomena in spontaneous speech, the pronunciation variants are only partially known and they need more extensive description. An ASR system has been used as a linguistic tool to investigate large speech corpora of tens of hours of speech, and to quantify pronunciation trends. In this contribution, we described a new methodology for carrying out corpus analysis on a syllable basis with W-syllables (obtained by syllabifying maximum length canonical pronunciations of isolated words) and S-syllables (where syllabification is carried out on the phoneme string without considering word boundaries). The number of canonical W-syllables is limited to about 1500, whereas the number of effectively aligned S-syllables is significantly larger (several thousands). The use of W-syllables allows us to relate the word level syllables to spoken ones, although French (unlike English) is supposed to ignore word boundaries when syllabifying an utterance. The limited number of W-syllables simplifies the description of the observed variation and facilitates generalisation.

For the different W- and S-syllables used, we found relatively stable syllable structure distributions, with the CV structure accounting for more than half of the data. Whereas French theoretically admits 14 different syllable structures (using C and V classes), the 6 structures CV, V, CCV, CVC, VC and CCVC syllables account for 99% of the corpus. Open syllables (CV, V, CCV, CCCV) account for about 90% of the W-syllables in the corpus. Closed syllables are more frequent in the aligned S-syllables (16%), which best correspond to what was actually said (the aligned and syllabified phone sequence). In speech, the increase of closed syllables is due to vowel deletions and syllabic restructuring. Whereas syllable deletions are relatively frequent for S-syllables (15%), which always measure full syllables (the vowel nucleus is mandatory in both canonical and aligned S-syllables), W-syllables have a much lower deletion rate of 6% (the vowel is mandatory only in the canonical form). Deletions mainly occur in cross-boundary positions. Concerning W-syllable deletions, most of them are due to highly predictable function words and word endings or to V syllable word beginnings. Investigating syllable-position dependent phone deletions, we could measure vowel deletion rates of 15/6% when including/excluding the schwa. Wheres a global consonant deletion rate of 13% could be measured independently of the consonant position within the syllable, coda deletions are three times as frequent as onset deletions. These different deletion options allow the transmitted word rate to increase without physically increasing the speech

rate. Or equivalently, it allows us to utter more words with fewer phonemes.

This study suggests that phone-in-syllable and syllable-in-word contexts might be of interest for acoustic phone modelling. This is an important direction for future developments. The perspectives of this study are diverse: the developed framework helps describe and quantify more or less well known linguistic phenomena on a syllable basis. Generic rules can then be formulated to generate pronunciation variants, even for rarely observed or unobserved words, for which variants cannot be estimated statistically. Plausible rules can address word-initial vowel deletion, backward onset-coda transfer and forward onset-onset transfer if the resulting onset is permissible (e.g. *c'est impossible* /**sɛ.tɛ̃**.pɔ.si.blə/ → [**stɛ̃**.pɔ.si.blə]. In future work, we intend to refine the present approach and extend the analysis of the alignment results. The syllable-based framework can also serve as a tool for manual transcription checking: omitted syllables point to either linguistic phenomena or simply to transcription errors. Finally, this research may contribute to syllable modelling for word fragments and out-of-vocabulary words.

8 Acknowledgement

The audio corpus used in this study was provided by the Radio Archives of INA (Institut National de l'Audiovisuel) in the context of a collaboration with their research and experimentation department.

References

Adda-Decker, M., Lamel, L., 1999. Pronunciation variants across system configuration, language and speaking style. *Speech Communication* **29** 83-98.

Boula de Mareüil, P., 1997. Étude linguistique appliquée à la synthèse de la parole à partir du texte. PhD thesis, University of Paris XI, Orsay.

Boula de Mareüil, P., et al., 1998. Evaluation of grapheme-to-phoneme conversion for text-to-speech synthesis in French.In: Proc. LREC 1998, Grenada, pp. 641-646.

Boula de Mareüil, P., Adda-Decker, M., Gendner, V., 2003. Liaisons in French: a corpus-based study using morpho-syntactic information. In: Proc. ICPhS 2003, Barcelona, pp. 1329-1322.

Corbin, O., 2003. Phoneme deletion in spontaneous British English. In: Proc.ICPhS 2003, Barcelona, pp. 2813-2816.

Cutler, A., Mehler, J., Norris, D. Segui, J., 1986. The Syllable's Differing Role in the Segmentation of French and English. Journal of Memory and Language. **25**, 385-400.

Dauses, A., 1973. Études sur l'e instable dans le français familier. Niemeyer Verlag. Tübingen.

Delattre, P., 1965. Comparing the phonetic features of English, Spanish, German and French. Julius Gross Verlag. Heidelberg.

Delattre, P., 1966. Studies in French and comparative phonetics. Mouton & Co., Paris/London/The Hague.

Dell, F., 1973. Les règles et les sons. Hermann. Paris.

Duez, D., 2003. Modelling Aspects of Reduction and Assimilation in Spontaneous French Speech, In Proc. IEEE-ISCA Workshop on Sponrtaeous Speech Processing and Recognition, 2003. Tokyo.

Durand, J., Laks, B., 2000. Relire les phonologues du français : Maurice Grammont et la loi des trois consonnes. Langue française. **126** 29-38.

Eggs, E., Mordellet, I., 1990. Phonétique et phonologie du français. Théorie et pratique. Niemeyer Verlag. Tübingen.

Encrevé, P., 1988. La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français. Éditions du Seuil. Paris.

Fouché, P., 1969. Traité de pronunciation française. Klincksieck. Paris.

Fougeron, C., Goldman, J.-P., Frauenfelder, U.H., 2001. Liaison and schwa deletion in French: an effect of lexical frequency and competition. In: Proc. Eurospeech 2001. Aalborg. pp. 639-642.

Fougeron, C., Bagou, 0., Stefanuto, M., Frauenfelder, U.H., 2002. Á la recherche d'indices de frontière lexicale dans la resyllabation. In: Proc. JEP, Nancy 2002. pp. 125-128.

Fowler, C.A., Treiman, R., Gross, J., 1993. The structure of English syllables and polysyllables. Journal of Memory and Language. **32** 115-140.

Gauvain, J.-L., Lamel, L., Adda, G., 1999. The LIMSI 1998 HUB-4e transcription system. In: Proc. DARPA Broadcast News Workshop. Herndon 1999, pp. 99-104.

Goslin, J., Content, A., Goldman, J.-P., Frauenfelder, U.H., 1999. Human and machine syllabification in French: A comparison. In: Proc. 2^{nd} Journées de Linguistique. Nantes 1999, pp. 75-80.

Greenberg, S., Chang, S., 2000. Linguistic dissection of Switchboard-Corpus Automatic Speech Recognition Systems. In: Proc. ISCA-ITRW Workshop on ASR. Paris 2000, pp. 195-202.

Greenberg, S., Carvey, H., Hitchcock, L., Chang S., 2002. Beyond the Phoneme: A Juncture-Accent Model of Spoken Language. In: Proc. Human Language Technology Conference (HLT), San Diego 2002.

Kahn, D., 1976. Syllable-based generalisations in English phonology. PhD thesis, MIT (published by Garland, New York, 1980).

Lacheret-Dujour, A., Péan, V., 1994. Towards a Prosodic Cues-Based Modelling of Phonological Variability for Text–to-Speech Synthesis. In: Proc. ICSLP, Yokohama 1994 pp. 1763-1766.

Ladefoged, P., 1975. A Course in Phonetics. Harcourt Brace Jovanovich Inc., New York/Chicago/San Francisco/Atlanta.

Léon, P.R., 1993. Précis de phonostylistique. Parole et expressivité. Fernand Nathan. Paris.

Lucci, V., 1983. Étude phonétique du français contemporain à travers la variation situationnelle. Publications de l'Université des Langues et Lettres de Grenoble, Grenoble.

Malmberg, B., 1975. La phonétique. Presses Universitaires de France. Paris.

Martinet, A., 1971. La prononciation du français contemporain. Librairie Droz, Genève/Paris.

Pallier, C., 1994. Rôle de la syllabe dans la perception de la parole : études attentionnelles. PhD thesis, EHESS, Paris.

Pérennou, G., Calmès, M. de, 1987. BDLEX, base de données lexicales du français écrit et parlé. Travaux du CERFIA, Toulouse.

Saussure, F. de, 1915. Cours de linguistique générale. Payot, Paris.

Shriberg, E., Preliminaries to a Theory of Speech Disfluencies. PhD thesis, University of California, Berkeley.

Strik, H., Cucchiarini, C., 1999. Modeling pronunciation variation for ASR: A survey of the litterature. Speech Communication. **29** 225-246.

Su, T.-T., Basset, P., 1998. Language dependent and independent spontaneous speech phenomena. In: Proc. SPoSS, La Baume-les-Aix 1998, pp. 55-58.

Van Son, R.J.J.H., Pols, L.C.W., 2003. An Acoustic Model of Communicative Efficiency in Consonants and Vowels taking into Account Context Distinctiveness. In: Proc. ICPhS, Barcelona 2003, pp. 2141-2143.

Verney Pleasants, J., 1956. Études sur l'*e* muet, timbre, durée, intensité, hauteur musicale. Klincksieck, Paris.

Vogel, I., 1982. La sillaba come unità fonologica. Fenomeni Linguistici **2**. Zanichelli, Bologna.

Walter, H., 1976. La dynamique des phonèmes dans le lexique français contemporain. France-Expansion, Paris.

Walter, H., 1988. Le français dans tous les sens. Éditions Robert Laffont, Paris.

Wioland, F., 1985. Les structures syllabiques du français. Slatkine-Champion, Genève/Paris.