

## *French liaison in the light of corpus data*

JACQUES DURAND

Université de Toulouse-Le Mirail & CNRS CLLE-ERSS

CHANTAL LYCHE

Universitet i Oslo

This is the prefinal version of J. Durand & C. Lyche (2008) French liaison in the light of corpus data. *Journal of French Language Studies*. 18/1: 33-66. Please refer to the published article for citations.

### **Abstract**

*French liaison has long been a favourite testing ground for phonological theories, a situation which can undoubtedly be attributed to the complexity of the phenomenon, involving in particular phonology/syntax, phonology/morphology, phonology/lexicon interfaces. Dealing with liaison requires stepping into all the components of the grammar, while tackling at the same time the quicksands of variation. The data on which a number of formal analyses are based have often been a source of concern since liaison, in part because of its intrinsic variable character, requires extensive and robust data. In the wake of the results from the study of other corpora , we present here extensive results based on the PFC database (Phonologie du français contemporain: usages, variétés et structures) and point to their implications for models of linguistic structure. While we do not believe that a motivated theoretical account can be mechanically extracted from the data, we conclude that future analyses will have to take explicitly into account the results of extensive corpus work as well as sociolinguistic surveys, acquisition studies, experimental phonetics and psycho-(neuro-)linguistic investigations, including the relationship between speech and writing. As stressed in Chevrot, Fayol and Laks (2005), these analyses will have to acknowledge that French liaison is not a homogeneous locus but a multi-faceted phenomenon requiring us to accept, without demur, the crossing of disciplinary boundaries.*

## *Introduction*

French liaison has long been a favourite testing ground for phonological theories, a situation which can undoubtedly be attributed to the complexity of the phenomenon, involving phonology/syntax, phonology/morphology, phonology/lexicon interfaces.<sup>1</sup> Dealing with liaison requires interaction with all the components of the grammar, while tackling at the same time the quicksands of variation. The data on which a number of formal analyses are based have been a source of concern for various phonologists (Morin, 1987; Kaisse, 1985 to cite but two) as liaison, in part because of its intrinsic variable character, requires that the data be extensive and robust. Comprehensive corpus based descriptions have, in fact, been published and, in the wake of a long structural tradition, Ågren (1973), Encrevé (1988) and De Jong (1994) who respectively analysed a radio corpus, the speech of politicians, and the Orléans corpus, have contributed significantly to a better description and understanding of the phenomenon.<sup>2</sup> On a smaller scale, Green and Hintze (1990, 2001) working within close family networks have added a number of pertinent observations. The empirical base has been augmented with experimental phonetics (Fougeron, Goldman and Frauenfelder, 2001; Nguyen, Wauquier-Gravelines, Lancia and Tuller, to appear; Spinelli and Meunier, 2005) and acquisition studies (e.g. Chevrot, Dugua and Fayol, 2005; Chevrot, Chabanal and Dugua,

<sup>1</sup> A number of colleagues have helped us in the preparation of this article. We particularly wish to thank Jean-Pierre Chevrot, Marianne Durand, Pierre Encrevé, Julien Eychenne, Bernard Laks, Noël Nguyen, Sylvain Detey, Gabor Turcsan, Doug Walker and two anonymous JFLS reviewers for their remarks or observations. All remaining errors are ours.

<sup>2</sup> See too Smith (1998, 1999: 130-184), which offers a sociolinguistic study of 20 hours of radio speech and a comparison with Ågren's results. As in other areas, the sociolinguistic work initiated in Canada, in particular by Henrietta Cedergren and David Sankoff, provides a strong model for corpus based work which French specialists in Europe have been slow to follow (but see Armstrong 2001). For liaison, see e.g. Tousignant and Sankoff (1979).

2007; Wauquier-Gravelines and Baud, 2005), and it might seem as if the factual aspect of the phenomenon has been satisfactorily circumscribed. We claim however that this is not the case, that the existing corpora do not offer a balanced picture of a highly varied phenomenon, and that each corpus provides a description of only a limited aspect of the French language, thus restricting the scope of its use. In this situation, thorough normative descriptions such as Fouché (1959) still constitute implicitly the basis for formal and/or pedagogical studies. We want to stress again, as Morin (1987) did so eloquently, that sound theories need sound data. We contend that the PFC (Phonologie du français contemporain) project<sup>3</sup> (Durand, Laks and Lyche 2003, 2005) with its protocol and coding system (described in §2) constitutes a step towards gathering a rich base providing the robust data necessary for adequate descriptions and sound analyses. Based on extensive data drawn from a minimum of ten investigation points and one hundred informants, we will argue that liaison cannot be seen as a single phonological process, but that it is partly morphosyntactic, partly phonological, partly phonetic and partly the result of the speaker's knowledge of the orthographic system, particularly in the areas most sensitive to sociostylistic variation.

### **1. Liaison behaviour in French**

It is well known that, historically, liaison phenomena are part of a general process of linking which allowed consonants to survive final consonant deletion. In generative phonology, the pioneering work of Schane (1968) broke with the American structural tradition and its strict separation of levels by assuming that liaison was a phonological phenomenon, but sensitive to morphological and syntactic information. A single explanation was given for liaison behaviour: the postulation of latent final consonants which, if not deleted, would be automatically resyllabified forward by an unspecified rule of *enchaînement*. Selkirk (1972)

<sup>3</sup> Cf. the PFC website: <http://www.pfc-projet.net>

followed suit and laid the background to a large number of modern studies of liaison from a theoretical perspective. Her work associated the following features with liaison:

(1) Liaison

- (a) Enchaînement (systematic forward linking *petit ami* = [pəti-tami])
- (b) Strong and regular link with phrasal syntax (e.g. X-bar theory)
- (c) The liaison consonant is underlying and belongs to the linking word

Each of these claims depends crucially on the data it takes into account. Selkirk's main source was Fouché (1959), a normative reference book aimed primarily at foreign students and teachers of French, listing over 30 pages of impossible instances of liaison, and claiming to describe "la prononciation soignée [...] des Parisiens cultivés nés vers la fin du XIXe siècle ou plus tard." The empirical basis of Fouché's observations are, however, open to serious questioning as underlined by Morin (2001) and Laks (2002) who demonstrate that so-called standard French is a hydra invested with multiple definitions. Even if many specialists agree on restricting SF to a geographical area (Paris) and to a social norm, that of educated speakers, the circumscription of French remains hard. This circumscription has been made worse by the fact that descriptions of French have regularly been extended in a variety of often conflicting directions.

Thus according to Martinon (1913 : vii) : 'Pour que la prononciation de Paris soit tenue pour bonne, il faut qu'elle soit adoptée au moins par une grande partie de la France du Nord'. Bruneau (1931 : xx-xxi), claims roughly the same thing: the prototypical speaker of standard French is 'le bourgeois parisien cultivé, plus largement le Français cultivé de toutes les grandes villes du nord de la France'. Durand (1936) restricts the norm to 'la petite bourgeoisie parisienne'. Pichon (1938) goes even further and narrows it down to 'les plus

vieilles familles parisiennes dont sont issus les officiers généraux et les évêques'. Fouché (1959) widens it to 'la prononciation en usage dans les conversations soignées chez les Parisiens cultivés'. For Martinet and Walter (1973), the group to be chosen as a norm are 'des personnes cultivées, de résidence normale parisienne, mais d'une assez grande mobilité géographique'. Malécot (1977) introduces the notion of relaxation : 'la conversation sérieuse mais détendue de la classe dirigeante de la capitale'. In recent textbooks, this conundrum is by no means solved. Coveney (2001) for example, who is particularly sensitive to sociolinguistic parameters, offers a description of 'Supralocal French' (following Wioland 1987: 69) defined as 'the neutral form of pronunciation' which is 'characteristic of the well-educated classes of the northern two-thirds of France' (2001: 4).<sup>4</sup> In this context, it does come as no surprise that the range of assumptions for liaison behaviour is far from homogeneous. Making too many or too few liaisons can be seen as socially positive or negative according to the source. Hermant, quoted by Milner and Regnault (1987), asserts : 'Je vous accuserai d'aller à la bourgeoisie si vous faites trop de liaisons. Nos pères en faisaient fort peu.' Delattre (1968 : 58), on the other hand, declares : 'A mesure que l'on s'éloigne de cette classe [i.e. la plus cultivée, JD/CL], le nombre de liaisons diminue; certaines liaisons qui sont à la frontière des obligatoires et des facultatives, sont presque toujours observées par les uns et presque jamais par les autres.' As for geographical variation, one suspects that most remarks are based on casual, unsystematic observations and quite often prejudice. Passy (1892:119) claims 'on fait infiniment plus de liaisons dans la Suisse romande, par exemple, que dans la région parisienne.' By contrast, Brun (1934: 45) asserts that liaison is considerably less frequent in Marseille than in 'le français commun' and that this is a sign of the sloppiness of southern speakers. Surely, such beliefs need to be tested on a large scale!

<sup>4</sup> See also Borrell and Billières (1989).

It has been known for a number of years that when corpora are examined, they do not necessarily confirm Selkirk's generalisations. Regarding (1a), we owe to Encrevé (1983, 1988) a thorough analysis of unlinked liaisons where the liaison consonant fills the coda of the last syllable of W1 (the first linking word in a W1 W2 sequence such as *trop intéressant*). Such an unlinked realisation is said by Encrevé to occur exclusively with variable liaisons and is undeniably typical of the speech of politicians and people who regularly express themselves publicly. In Chirac's television address of November 2005, one could note at least 10 examples of unlinked liaisons, as for example in "il faut[t] intensifier l'action contre les filières". We return to this issue in §3.3.

The regular link to syntax has been challenged by Ågren (1973) and De Jong (1994) who give numerous instances of distinct liaison behaviour within the same syntactic context. To take just one example, De Jong notes that the different forms of the imperfect of *être* behave quite distinctively within identical contexts: *était/étaient* link to the following word in 20% of the recorded instances while *étais* triggers a liaison in only 5.3% of the possible instances. Morin has long questioned a unified approach to liaison, reintroducing a morphological dimension into the analysis, and the status of the liaison consonant as underlying has been rejected within a number of theoretical frameworks (Klausenburger, 1978; Tranel, 1881, *inter alia*). But, what can be said of the data since Selkirk's claims of (1972)? Our knowledge of liaison behaviour has certainly benefited from numerous pertinent observations, and it has been enriched by large corpus studies. Nevertheless, these contributions fail to provide a solid base for constructing an accurate typology of liaison in French. The truly large corpus studies such as Ågren's, Encrevé's or the Orléans corpus used by De Jong are either too restricted or not sufficiently controlled to allow a *bona fide* description of all the liaison contexts and of general liaison behaviour. Ågren's corpus, which can be claimed to represent an older stage of the language, is exclusively based on radio

recordings and does not permit a clear distinction between different registers. Whether the observations can be extended to the casual use of French is indeed questionable. Encrevé's corpus limits the study to politicians, whose speech can hardly be considered representative of the 'Frenchman in the street'. The Orléans corpus, which was collected according to standard sociological practice, does not suffer from the same limitations. On the other hand, it does not provide any information on the origin of the speakers and, by definition, it is restricted to one city in northern France. The PFC project that we will now briefly describe, proposes to combine geographical representation with controlled registers on a large scale, and thus should offer a truer picture of liaison behaviour in modern French.

## **2. PFC-protocol and coding liaison**

The PFC project, a collaborative endeavour which brings together some fifty researchers from a variety of countries, aims at the elaboration of a large reference oral corpus suited for phonological analyses of different varieties of French. The corpus includes the recording, partial transcription and coding of over 600 speakers from the francophone world on the basis of a common protocol including two reading tasks and two conversations, thus adopting a strict (classical) Labovian method. The PFC methodology is explained in detail by Durand and Lyche (2003) and the full protocol can be downloaded from the PFC website ([www.projet-pfc.net](http://www.projet-pfc.net)). Consequently, we will confine ourselves to a presentation of its main features.

### **2.1 The PFC-protocol**

The phonological perspective of the project focuses on the elaboration of the phonemic/allophonic inventory of speakers (and thereafter of locations), and on the collection

of robust data concerning two central phenomena in French phonology, schwa and liaison. The speakers, selected according to a network principle (Milroy, 1980), are asked to read aloud a wordlist and a passage, and are recorded during a semi directed and an informal conversation, each lasting from 20 to 30 minutes. The reading tasks were integrated within the protocol for two reasons: (i) they guarantee a full comparability of the results and (ii) they are required once the phonological goals have been clearly formulated. No oral corpus, regardless of its size, can claim to include all the data the linguist is looking for, but the reading tasks give us systematic access to much of the phonological information we seek and, in addition, to a formal register. They also allow the testing of hypotheses concerning the relationship between speech and writing. The distinction between semi-directed and informal conversation may often be fuzzy in a recording situation where the speaker is not at ease, and some would argue that the so-called informal register is by no means informal (Gadet, 2003). It turns out however, that work on automatic speech recognition done by Adda-Decker and Boula de Mareüil (LIMSI laboratory, Paris) clearly shows that their models cope better with the PFC semi-directed conversations than they do with the informal ones. The latter show more reductions and a faster speech rate thus confirming a dichotomy between the two.<sup>5</sup>

Within each location, the group of speakers includes 10 to 12 persons equally distributed for sex within well defined age ranges and ideally including three generations of families (Durand and Lyche, 2003). So far within the project, we have focussed on geographical variation, recording and analysing cohorts of speakers from as many different locations as possible in the French-speaking world. In the protocol, we minimize the social diversity requirement, aware that it is less easy to achieve with small groups of speakers. We favour family networks which allow for better comparison of age-grading, especially when the social background of the informants has remained relatively stable. Complete social

<sup>5</sup> For some of the work done on the PFC corpus from the automatic speech recognition perspective, see Boula de Mareüil, M. Adda-Decker & C. Woehrling (2007).

diversity cannot be achieved given that the protocol with its two reading tasks requires a certain level of literacy from the speakers, thus excluding completely illiterate speakers, a decision that we find defensible within France. Outside of France, on the other hand, there exist regions, like Louisiana, where French has been maintained nearly exclusively orally and where the majority of speakers do not read nor write French. Excluding illiterate speakers and choosing to concentrate on the few literate ones would seriously distort the linguistic picture of the location. To accommodate this particular situation, the protocol has been modified in such communities, and the reading tasks replaced by translation tasks easily performed by most speakers (Klingler, 2006; Lyche, 2007). The lack of explicit social sampling within individual survey points has fortunately not had an adverse effect on the whole database, the latter comes out as socially balanced when all the speakers are taken into account.

Once the recordings are made, they are analysed using Praat.<sup>6</sup> Exploiting this well-known software devised by Boersma and Weenink at the University of Amsterdam, we propose an orthographic transcription aligned to the signal and kept as close as possible to standard spelling (Durand and TARRIER, 2007). For every speaker, we thus transcribe and align the 94 items of the wordlist, the text, and roughly 10 minutes of each conversation. The transcription and alignment of each word of the wordlist prepares the data for phonetic studies. Nguyen and Espesser (2004) for example, extract 46 items (mid-vowels, /A/, nasal vowels) and detail the procedures used for the automatic extraction of the formants together with the methods applied to minimize a certain amount of variation due to speakers' physiological differences. These methods can be used to obtain the average phonemic system of each speaker and then to sketch out the main characteristics of the local variety of French, thus fulfilling our first phonological objective within the project (establishing phonemic inventories of varieties of French and a specification of the main allophones). Our choice of

<sup>6</sup> See <http://www.fon.hum.uva.nl/praat/>

software has proved felicitous and we have taken advantage of Praat's flexibility which allows, for example, the duplication of tiers. In addition to a transcription tier, we select two other tiers for annotation purposes: a schwa tier and a liaison tier. The orthographic transcription is duplicated and coded for schwa on the second tier, and for liaison on the third tier. The procedure is applied to the text, 3 minutes of each conversation for schwa and 5 minutes of each conversation for liaison. As there was agreement within the project on the need to go beyond the segmental domain and to incorporate a study of prosody, several attempts have been made to define a coding system for prosodic factors as well, and a fourth tier has been added in a few survey points, where we select five speakers whom we partially code for prosodic factors summarised further down (Lacheret and Lyche, 2006; Lacheret, Lyche and Morel, 2005).

It was decided at the outset of the project that information on schwa and liaison behaviour could best be obtained through a systematic coding scheme based on a common methodology which has been adopted for prosody as well. All three systems are claimed to be theory-independent and allow an initial qualitative and quantitative sorting out of the data, and all three reproduce the coder's perceptions. Schwa coding is made up of four digits added to any graphical *e* and to any final pronounced consonant, considered as a potential schwa site (Durand and Lyche, 2003). It relies on the classical work of Dell (1973/1985) and provides information on the realisation or not of the vowel, its position within the word/clitic, and on the right and the left context.<sup>7</sup> While both schwa and liaison can be annotated for the presence/absence of a segment, prosody is characteristically a gradient phenomenon, and its complex nature challenges attempts to devise a simple coding system based on a broad consensus. As a consequence, we chose to limit our investigations to the influence of prosodic factors on the presence/absence of schwa and at this stage, the primary aim of our prosodic

<sup>7</sup> The PFC methodology and codings are extensively exemplified in two recent PhDs: Eychenne (2006), Sobotta (2006). For a quantitative treatment of schwa, see Durand and Eychenne (2004)

coding is to enrich the schwa coding system. In spite of large disagreements within the community as to definitions and models, most researchers would agree with our theoretical premises: both the lexical word and the syllable are pertinent segmentation units, but the stress bearing unit is the syllable; French is a language where stress is assigned at the group level and not at the word level, fulfilling a demarcative function (in Trubetzkoy's sense), but not a contrastive one. Subscribing to these assumptions entails that the prosody tier should offer a segmentation into syllables. We depart here from the other two coding systems in that the annotated unit is not the graphic word but a syllable transcribed in the SAMPA phonetic alphabet. Each syllable is followed by four digits indicating the presence/absence of a perceived prominence, presence/absence of a pause and the position of the syllable within a rhythmic group (Lacheret and Lyche, 2006).<sup>8</sup>

Devising annotation systems is a difficult task and we fully agree with Gadet (2006): '[...] tout geste méthodologique a des conséquences théoriques. Si, dans la collecte de données, comme dans les premiers temps d'exploitation, un geste ne fait pas l'objet d'une décision mûrement pensée, il risque de colporter des choix d'autant plus pernicieux qu'ils demeurent implicites.' We believe the liaison coding system to which we now turn provides an adequate basis for further descriptive and analytical work.

## **2.2 Coding liaison**

Our claim is that liaison can be appropriately described, at least pretheoretically, by using an orthographic transcription as a starting point. The coding system we use is a set of alphanumeric symbols which are added to each potential linking word of French. The coding

<sup>8</sup> The prosodic annotation has already been applied to the study of schwa; the prosodic study of liaison being our next objective.

is auditory (just like for schwa and prosody).<sup>9</sup> It is kept as simple as possible in order to minimize the number of errors that naïve coders are bound to make and, more specifically, because it aims at a global description of the data. The coding system was conceived as a means to organise the data and construct a sound typology of liaison contexts in French. In other words, it aims at providing answers to the following questions: in which contexts is liaison always present (categorical liaison), in which contexts is it optional (variable liaison), and in which contexts is it totally or virtually absent (erratic or non attested liaison)? For coding purposes, by liaison we mean the pronunciation of any graphic consonant when the word (W2) following a linking word (W1) is vowel-initial: *mes amis* (*mes [z] amis*), *petit ami* (*petit [t] ami*), *toujours ami* (*toujours [z] ami*). Since we code any graphical consonant which is potentially a liaison consonant, we should stress that there are cases of liaison which are not indicated in the orthography - e.g. *quatre enfants* (*quat' [z] enfants*), *il va à Paris* (*i' va [t] à Paris*). These liaisons are taken into account as well. We speak of ‘epenthetic liaison’ here but without committing ourselves to a theoretical description in terms of epenthesis. The labels we use (e.g. latent or epenthetic consonant) should not be confused with an analysis which can be couched in a variety of frameworks (see §4). On the other hand, it is not the case that absolutely all final graphic consonants are coded for potential liaisons and we follow here the classical typology proposed by Delattre (1951, 1966). Delattre classifies as ‘forbidden’ liaisons occurring after the conjunction *et* and liaisons following a singular noun. Thus in *Pierre et// Anne, le savant// étudie les mouvements de la terre*, no liaison is ever expected. The PFC protocol excludes these two contexts from systematic liaison coding, but asks the coder to record such liaisons if they should happen to be made by a speaker. This decision was not triggered by a particular theoretical approach: it reflects our concern to minimize the

<sup>9</sup> The annotations by a first coder are systematically checked by a second person, and, when resources allow, a third one in cases of disagreement.

number of possible errors made by trained but nevertheless naïve coders who would probably simply *forget* to code instances viewed by them as totally impossible.

All our codings are fixed sequences of alphanumeric symbols, minimally two, positioned after the linking word.

## (2) Liaison coding system

- Field 1:
  - 1 = one syllable
  - 2 = two syllables or more
- Field 2: is drawn from the set{0, 1, 2, 3, 4} where:
  - 0 = absence of liaison
  - 1 = liaison enchaînée (forward linked liaison)
  - 2 = liaison non enchaînée (liaison consonant present but not forward linked)
  - 3 = uncertainty
  - 4 = ‘epenthetic’ liaison

The first coding position involves a decision as to whether the linking word is phonetically monosyllabic or polysyllabic. The other symbols that we use in the transcription indicate the nature of the liaison consonant, the presence of pauses, hesitations or glottal stops, cases of unexpected liaison (in relation to spelling) and liaisons involving nasal consonants. Given Encrevé’s well known (1988) work on liaison ‘non enchaînée’, we have paid attention to the possibility that a case of liaison or non liaison might involve either a pause, a hesitation or the presence of a glottal stop. We lump all these phonetic features under the symbol **h**. The symbol **h** indicates that there is a non smooth transition which can be the object of further fine-grained phonological and phonetic research. A presentation of all the

features of our coding would require much more space than is available here. Instead of this, we illustrate our notation using the first paragraph of the PFC text. If the following pronunciations were to be observed (in broad phonetic transcription within square brackets, with a full stop indicating a syllable boundary):

"Le maire de Beaulieu - Marc Blanc – est [ɛ.tã] en revanche très [trɛ.zɛ̃.kjɛ] inquiet. La cote du Premier Ministre ne cesse de baisser depuis les [lɛ.zɛ.lɛk.sjõ] élections. Comment [kɔ.mã.ã], en plus, éviter les manifestations qui ont [õt.ʔy] eu tendance à se multiplier lors des visites [vi.zit.ʔɔ.fi.sjɛl] officielles ?"

the PFC coding would be :

"Le maire de Beaulieu - Marc Blanc – est11t en revanche très11z inquiet. La cote du Premier Ministre ne cesse de baisser depuis les11z élections. Comment20, en plus, éviter les manifestations qui ont12th eu tendance à se multiplier lors des visites20h officielles?"

The interpretations of the above codes are as follows:

- *est11t en* = monosyllable (1) + liaison enchaînée (1) with [t]
- *très11z inquiet* = monosyllable (1) + liaison enchaînée (1) with [z]
- *les11z élections* = monosyllable (1) + liaison enchaînée (1) with [z]
- *Comment20, en* = polysyllable (2) + absence of liaison (0)
- *Ont12th eu* = monosyllable (1) + liaison non enchaînée (2) with [t] followed by a pause, a hesitation or a period of glottal closure (h)
- *visites20h officielles* = polysyllable (2) + absence of liaison (0) + presence of a pause, a hesitation or a period of glottal closure (h).

The coding is applied to all speakers within each survey point for three styles: the passage, semi directed conversation (5 minutes) and free conversation (5 minutes). The PFC text alone presents 35 potential liaison sites and provides valuable information when attempting to establish the characteristics of a formal register, and when determining geographical differences, if any. All the codings can be extracted by various tools and analysis can begin with much interaction between data and hypotheses. For this paper, we have made use of the tools provided on the PFC website by Atanas Tchobanoff (<http://www.pfc-projet.net>), as well as a stand-alone 'platform' devised by Julien Eychenne on which all our figures rely.

### **3. Ten investigation points**

The empirical base for the liaison results presented here consists of ten PFC investigation points (100 speakers) selected for their geographical spread, although at various points we will go beyond these survey points to test further hypotheses. Varieties of Midi French are represented by four survey points: 11a (Douzens, Aude, 10 speakers), 13a (Marseille, Bouches du Rhône, 10 speakers), 13b (Aix-Marseille, Bouches du Rhône, 8 speakers) and 64a (Biarritz, Pyrénées Atlantique, 12 speakers). The northern part of France is covered by the remaining six points: 42a (Roanne, Loire, 9 speakers), 50a (Brécey, Manche, 11 speakers), 54b (Ogéville, Meurthe, 11 speakers), 75c (Paris, 12 speakers), 91a (Brunoy, Essonne, 10 speakers), 85a (Treize-Vents, Vendée, 7 speakers). Within each group, rural zones are present (Douzens, Brécey, Treize-Vents), and the Paris region, crucial in the definition of standard French, is covered both through a survey in the inner city and one in the suburbs (Brunoy). This selection of points attempts to capture some sociological diversity as well, the speakers

having various backgrounds and education levels with the exception of Paris (inner city) where all our speakers belong to the upper class/nobility. In addition, in two of the regions - Brécey and Douzens - a local dialect has survived and is still in use. Brécey, a small village of 2113 inhabitants, is located in the southwest of Normandy, in a traditionally prosperous agricultural area. In this oil region, which has maintained a strong linguistic identity (Lepelley, 1999), a Norman dialect is still spoken by the oldest speakers. Douzens on the other hand, a small village of 600 inhabitants in the middle of a wine-growing region, is situated in the heart of Languedoc. The oldest speakers express themselves in a local variety of Occitan which remains strongly anchored in the population (Durand and Tarrier, 2003). Douzens departs from Brécey in that younger Douzens speakers (in the 30-60 age group) exhibit a certain competence in their local dialect, which is not the case in Brécey. A local lexical influence however, prevails in both regions. Taking into account all these factors, we feel confident that the 100 speakers selected for this study are reasonably representative of the French population within 'l'Hexagone', a claim which will have to be assessed later by reference to the entire database.

### **3.1 Categorical liaisons**

We pointed out in §2.2 that our presentation of the liaison coding system in the protocol rests on Delattre's (1951, 1966) detailed classification. Although later work (e.g. Léon, 1992) makes a few diverging observations, Delattre's four page tableau presents, still to this date, the most extensive picture of liaison contexts and provides much of the basis for current pedagogical material, as for example Walker (2001).

The PFC results concur systematically with Delattre's observations for non attested liaisons or erratic liaison, but they show some deviation for categorical liaisons. Our data does not provide any instance of liaison after a singular noun, or after the conjunction *et*.

The reader should recall that we chose to exclude these particular contexts from the coding, and that only realised liaisons are noted. We coded however non personal pronouns ending in a nasal in construction such as *trouvez-en une*, and, as described by Delattre, no liaison appears in this environment either ([\*truveẏnyn]). The situation is not as clear-cut when studying categorical liaisons. Delattre considers that a liaison is compulsory between a determiner and a substantive, a personal pronoun and a verb or another clitic, and again, the PFC data confirms this recommendation. Our results contradict Delattre however, on three points: (i) a monosyllabic preposition does not systematically entail a liaison and (ii) neither does a preposed adjective nor (iii) the impersonal construction *c'est+*. The first context was already observed by Léon (1992: 155) who remarks<sup>10</sup>: ‘La liaison tend à être *obligatoire avec les formes monosyllabiques*, qui sont inaccentuées et entrent ainsi dans la règle de cohérence syntagmatique comme dans : *en[n]* effet, *en[n]*avant, *dans[z]* une heure [...]’. Our own observations corroborate this statement and offer certain precisions: *en* induces a nearly categorical liaison, and out of 1124 relevant occurrences, we note two only examples without liaison: *en// un quart d’heure*, *manqué beaucoup de confiance en// elle*.<sup>11</sup> All other prepositions vary in their usage, although liaison instances are overwhelming: there is no liaison after *dans* in 5% of the occurrences (245), including fixed expressions like *dans// un sens*. *Chez* shows more variation with nearly 12% of cases of no liaison affecting systematically a lexical word: thus there is no liaison in *chez// un copain*, *chez// un patron*, *chez// Anne et Pierre*, but liaison is categorically present when followed by a monosyllabic pronoun *chez [z]elle*. We hypothesise that this discrepancy stems from prosodic factors and the clitic nature of the pronoun. We thus predict that *chez// Al* or *chez// Yves* is possible, as *Al*

<sup>10</sup> Léon (1992) is slightly inconsistent on this topic. He includes (p.154) examples like *chez[z]eux* among instances of compulsory liaisons, but points out in the next page that liaisons only tend to be systematic after monosyllabic prepositions.

<sup>11</sup> Out of the 1124 codings for *en*, liaison ‘enchaînée’ is attested in 1102 occurrences; lack of liaison in the remaining 22 cases corresponds to either ‘h aspiré’ (9 occurrences, e.g. *en hockey*), hesitation (11 occurrences) and our 2 clear examples of non liaison.

or *Yves* are treated as lexical words, but that clitics (*elle, eux*) will always be linked to the preceding monosyllabic prepositions. We reach here the limits of corpus studies which are never controlled enough or rich enough to allow investigating all hypotheses and which need to be supplemented by specific tests. We do not have enough tokens to comment on other frequent monosyllabic prepositions which might trigger liaison: *dès, hors, sous, vers*. Of the two occurrences of *sous*, there is one example with liaison (*sous [z] un autre nom*). *Vers*, in accord with the norm, does not appear to trigger liaison (5 examples).<sup>12</sup>

Our data disagrees with Delattre's classification for prenominal adjectives and the impersonal construction *c'est+*, two contexts viewed likewise by Léon (1992) as triggering categorical liaison. The two conversations fare poorly in providing abundant data on prenominal adjectives for the simple reason that *petit* overwhelmingly represents this particular context and that other adjectives occur only sporadically in a liaison environment. We can nevertheless test the sequence 'prenominal adjective + noun' in the PFC text read by all our speakers through two expressions: *grand émoi, grand honneur*. We assumed that in a reading task, all 100 speakers would link the adjective to the following noun, which is not what we observe: six speakers do not make the liaison and two pronounce a [d] instead of the expected [t]. Four of the six cases concern *grand émoi*, suggesting that the lack of familiarity with the construction impacts on the absence of liaison.

Data on prenominal adjectives present a particular interest due to the theoretical debate concerning their treatment. The current literature opposes a morphological approach (Steriade, 1999 inter alia) to a phonological one (Féry, 2003). Steriade (1999) argues that masculine and feminine adjective allomorphs are listed in the lexicon, and that a hiatus situation is resolved by lexical conservatism 'a class of grammatical conditions [...] promoting the use of pre-existing familiar expressions or parts of properties of such

<sup>12</sup> Polysyllabic liaisons are briefly dealt with in §4.2.

expressions.’ When hiatus occurs between an adjective and a noun, lexical conservatism requires that before inserting new segments that would solve the problem, one should look within the paradigm for possible solutions. Since the feminine allomorph of an adjective usually ends in a consonant, it implies that in a hiatus situation, the masculine allomorph will take the shape of the feminine allomorph.<sup>13</sup> In the phonological approach, defended by Féry (2003), the proper ranking of syllabification constraints suffices to account for the liaison form of the adjective. Both analyses treat liaison as a means of avoiding hiatus, and both propose to explain the presence of a consonant in examples like *sot ami*, *sot aigle*. We will not dwell here on the numerous counterexamples to a NO HIATUS constraint in French<sup>14</sup>, but will instead consider the data the analyses are based upon. Morin (1987) already pointed out the artificial character of *sot ami*, and a search through the entire PFC database should bring a few answers concerning pronominal adjectives. We note, that in his elicitation work involving nasal vowels, Sampson (2001) was unable to trigger liaison for adjectives placed in prenominal position and tested without success *fin*, *hautain*, *lointain*, *malin*, *mignon*, *souverain*. He concludes that, outside the usual inventory<sup>15</sup> (*un*, *mon*, *ton*, *son*, ... *bon*, *plein*), ‘the available evidence suggests that ZERO-liaison may already be established, or be well on the way to becoming established, as the default arrangement’ (p.255).

Prenominal adjectives appear in large numbers in the base, but only rarely in a liaison environment, and when they do, the liaison is not categorical. The adjective *gros* will serve to illustrate this point: we record 139 occurrences of *gros* in the base, but only 8 in a liaison context, equally distributed between liaison and no liaison. The liaison is then present with the plural form of the adjective (*gros [z]ouvrages*), but not in the singular (*gros// immeuble*). This particular example shows the strength of the plural marker for liaison, although we should

<sup>13</sup> Steriade restricts the number of possible liaison consonants to a few, which explains why in a liaison context *grand* is not pronounced with a [d].

<sup>14</sup> See Morin 2005 and the references therein.

<sup>15</sup> We return to these items in §4.

show caution in drawing hasty conclusions. Disyllabic adjectives like *premiers* vary between liaison and no liaison; *grands* links to the following word; and *petits* is pronounced several times with a [t] instead of the expected [z], as in *beaucoup de p(e)tits [t]hotels*. Most interestingly, although so-called elementary adjectives (in the terminology of traditional grammar) do occur regularly as expected, in a prenominal position, they rarely do so in a liaison environment. In addition to *gros*, already mentioned, we observe the same phenomenon with *ancien(s)*, *dernier(s)*, etc., frequent with slightly less than 100 occurrences each, but never in a potential liaison site. In other words, speakers seem to talk without difficulty about *un gros type*, *un gros chien*, but not about *un gros homme*, *un gros âne*. The PFC base not only throws doubt on the categorical character of liaison in prenominal adjectives, but it suggests that speakers systematically avoid a situation where they will be compelled to make a decision concerning the presence or not of a liaison (Lyche, 2003). To speak in terms of Optimality Theory, in such contexts, the winning candidate is the null candidate.

We will conclude this review of compulsory liaisons with *c'est+*, where our data blatantly contradicts Delattre. With only 30% of the liaisons realized, we have here a case of true variation. The examples in (4) taken from Brécey, make this clear.

- (4) C'est// agréable                    *it's pleasant*  
       C'est// un architecte                *he's an architect*  
       C'est// en cours                    *it's being done*  
       C'est [t]un agriculteur              *he's a farmer*  
       C'est [t]à mes fils                  *it belongs to my sons*

The differences cannot be attributed to distinctions of sex, age or education, the same speaker shifting from one form to the next within the same conversation, a type of inherent variation well described by Encrevé (1988).

The PFC data challenge the standard classifications (both Delattre's and Léon's) and point to a more restricted usage, which we examine in §4.1. Although we acknowledge a considerable distance between the PFC results and the situation in a dialect like Ranrupt where liaison is restricted to the determiner (Aub-Büscher, 1962), we witness a reduction of liaison contexts in daily conversation. Considering the gap between orthoepists' recommendations and the everyday use of so-called compulsory liaisons, we are entitled to expect even fewer liaisons when examining environments generally presented as optional. However, we should remind the reader that figures such as the ones given here require a great deal of care in their interpretation. We return to this issue in §4.

### **3.2 Variable liaison**

Returning to Delattre's classification, we firmly eliminate from potential liaison sites plural NP-VP constructions (*les parents [z]attendent*), already excluded by some linguists (for recent reviews of the situation, see Bonami, Boyé and Tseng, 2004, 2005). The PFC text could have triggered a liaison in *Quelques fanatiques[z]auraient même entamé*, but none of our 100 speakers makes a liaison in what we consider the most formal register. There is a general consensus in the literature that variable liaisons are located to the right of a head, linking it to its complement or modifier. The two prototypical instances are liaison after a plural noun and liaison after a conjugated form of the verb. Morin and Kaye (1982) for example, underline the morphological character of the liaison consonant [t] which they view

as a verb marker.<sup>16</sup> We know however (De Jong, 1988; Encrevé, 1988) that certain forms are more apt to link than others, that lexical items must be considered individually. In our data, for example, we find 1130 codings for *est* and we do observe true variation: 563 liaisons and 576 no liaisons. Looking at *était*, we obtain a totally different picture: 391 coded tokens with 36 liaisons (9.2%) and 355 no liaisons. These results are expected, although they indicate a further reduction in usage compared to De Jong (1994) who notes 20% of liaisons for the same item. There does not exist however any systematic study of geographic discrepancies. As the linking power of *être* is well accepted, we compare in (5) the third persons (present/imperfect, personal/impersonal) in the two rural areas described in §3, and oppose an oïl dialect village (Brécey) to an oc dialect village (Douzens).

(5) Variable liaison, *est* vs *était*: Brécey, Douzens

	Brécey		Douzens	
	liaison	no liaison	liaison	no liaison
Est	5	14	25	5
C'est	15	37	34	43
Était	1	20	9	6
C'était	1	16	13	10

These forms of the auxiliary *être* occur regularly in the data. Brécey appears to be moving towards a state of attrition for three of the forms and supports our classification of the impersonal *c'est* as a variable liaison site. Douzens speakers, on the other hand, conform to

<sup>16</sup> Our data presents for example two occurrences of a [t] insertion in the text : *qui mène [t] au village*.

the general pattern for the present form *c'est*, but link more frequently all the other forms of the verb. The variation is then item determined, but also geographically determined. A close scrutiny of the data brings forth instances of liaisons in [t] without counterpart in Brécey: *avait, allait, travaillait, voulait*, ('had', 'went', 'worked', 'wanted') which all trigger liaison at least once in the Douzens corpus. Note however that, if we conclude to a stronger propensity for liaison usage in Douzens which would represent some sort of conservative dialect (assuming that we view liaison as remnant of an older state), we still would have to account for the systematic lack of liaisons in first persons (*avais, voulais, etc.*). To test further this distinction north/south, we expanded our search to contrast the four southern French investigation points mentioned earlier (11a, 13a, 13b, 64a) with six non southern surveys used earlier on (i.e. 42a, 50a, 54b, 75c, 85a, 91a). A number of word-forms are not frequent enough to allow extrapolation from the data. For this reason, we have selected two forms of the verb *être* (i.e. *(c')est* and *(c')était*), and one form of *avoir* (i.e. *avait*). We have not used the text but only the formal and informal conversations. The results are as follows:

(6) *c'est, c'était, avait*, north vs south

	North		South	
	liaison	no liaison	liaison	no liaison
(C')est	155 33.91%	302	146 42.69%	196
(C')était	10 5.34%	177	26 15.75%	139
avait	0 0%	126	11 11.34%	86

On the basis of this data, the southern data seems to be nearer to the orthoepic norm than the northern one. Interestingly, the three Belgian surveys we examined gave the following results (in between north and south) for the same forms: *(c')est* 47.33% of liaisons realised (142/300 tokens), while *(c')était* comes close to the north: with 4.61% of realised liaisons (6/130 tokens), and *avait* 0% (0/54 tokens). We do not wish to draw hasty conclusions from this sample data but we wonder whether Brun did have any solid comparative evidence when he asserted of Marseille French: ‘Les liaisons sont donc beaucoup moins fréquentes qu’en français commun. Cette négligence, ainsi que la paresse à articuler les groupes de consonnes donne au parler du provençal ce caractère de vulgarité qui choque le nouveau-venu.’ We need to wait for the PFC basis as a whole to be available to make more trustworthy comparisons. At this stage, it seems to us likely that mainland French, and possibly European French, will reveal a common sociostylistic structuring given the close interaction between the varieties and the similar normative pressures which weigh on them. We do not of course exclude the possibility of local discrepancies but blanket assertions on Swiss French, Belgian French, Southern French or Corsican French are more likely to reflect prejudices than conclusions based on observations.

### **3.3. Liaison non enchaînée**

Ever since the groundbreaking work of Pierre Encrevé (1983, 1988), the question of ‘liaison non enchaînée’ (unlinked forward liaison) has proved a thorn in the flesh of French specialists. As will be recalled, the central claim of Encrevé’s article is the observation that liaison does not always involve the linking forward of the liaison consonant (*il est [t]inquiet*) as traditionally assumed. The fact that this forward linking does not always take place had occasionally been pointed out in the literature but such examples had always been treated either as a mark of emphasis or as performance errors without major significance. For instance, Coustenoble and Armstrong (1934: 142-143) had pointed out that when words begin

with a vowel an ‘intensive stress’ could be achieved by stressing the first syllable of the word beginning with a consonant (*Je suis en<sup>h</sup>chanté de vous voir*) but they ask ‘How does a French speaker avoid carrying over a liaison consonant when he places emphatic stress on the first syllable?’. Their answer was the following: ‘The liaison consonant is pronounced but it does not function as a liaison consonant, the speaker inserting the glottal plosive which starts the emphasized syllable in its stead: **ʒ sɥiz<sup>h</sup>ʔãʃtɛ d vu<sup>h</sup>vwar; s ɛt<sup>h</sup>ʔapsɔlymã ɛpsibl<sup>h</sup>**’.

On the basis of his corpus of high level political speeches, Encrevé contends that traditional assumptions are not correct. For a start, ‘liaisons non enchaînées’ are not marginal: for 11 out of 21 politicians he examines, the percentage of ‘liaisons non enchaînées’ is about 11%, and, in one of Jacques Chirac’s speeches (5-10-1981), 33.7% of the optional liaisons are ‘non enchaînées’. In addition, instrumental analysis of Encrevé’s data failed to reveal any systematic tendency to use such a device for emphasis. Encrevé makes also the further claim that in categorical liaison only forward linking is attested and that it is solely in the case of optional liaison (and indeed fixed final consonants) that there is variation as to the eventual anchoring of the liaison consonant.

It is not clear whether earlier specialists had failed to notice the importance of ‘liaison non enchaînée’ or whether the latter is a modern phenomenon. Passy (1892: 42), who considers linking forward as part of a natural French rhythm, seems to be describing a ‘liaison non enchaînée’ (in *C’est une idée*) in the following passage: ‘Et de fait, on cesse les liaisons dès qu’il y a arrêt. Rien de plus risible qu’une liaison faite mal a (sic) propos. *C’est une idée* prononcé (set, ynide), fait croire qu’on a le hoquet. Un professeur prononçait des phrases comme *la première est excessivement facile* en s’interrompant après *est*; (la prəmjerɛ, "teksɛsivmã 'fasil). La première fois que nous l’avons entendu, ça a été un éclat de rire général.’ The negative description of this pronunciation may indicate that it was present in the

speech of some individuals but interpreted as a sign of poor verbal command. However that may be, Encrevé was absolutely correct in stressing that unlinked forward liaison is part and parcel of the speech of what are called in France ‘les professionnels de la parole’: politicians, news readers, teachers, lawyers, and so forth.

On the other hand, other work on liaison has failed to attribute the same significance to ‘liaison non enchaînée’: Laks’ (1983) work with Villejuif adolescents, for instance, makes no room for this phenomenon. Acquisition work also shows that before schooling, liaison is either present, and if so linked forward, or absent (see Chevrot, Dugua, Fayol, 2005, Wauquier-Gravelines and Brau, 2005, and the references therein). A corpus such as the PFC one is therefore essential in establishing the role played by ‘liaison non enchaînée’ outside the sphere of political meetings, conferences and broadcasting. For this, we have checked data from 17 surveys made in mainland France (the ten surveys mentioned earlier plus 21a Côte d’Or, 31a Haute Garonne, 38a Isère, 44a Loire Atlantique, 69a Rhône, 75x Paris, 92a Hauts de Seine). The geographical and social range, as well as the number of codings (22737 codings for all styles) and informants (156, in total) obviously insures the representativeness of our corpus.

The PFC coding for non enchaînement has been explained in §2.2 and is symbolised by the figure 2 placed in second position in the alphanumeric liaison notation. There are 26 occurrences of ‘liaison non enchaînée’ in the reading of the text. Leaving aside errors in the coding, a close examination of the cases reveals marked hesitations in most of the cases. The most salient case of ‘non enchaînement’ is the sequence *Il s’est, en désespoir de cause*, often read as *il s’est[t] [ʔ]en désespoir de cause*. If we turn to the conversations, there are 45 examples of the ‘non enchaînement’ coding (as opposed to 7818 examples of forward linking). We have examined all the other cases individually. The coders seem to have had extreme difficulty in establishing what ‘non enchaînement’ involves. In most of the cases they

wrongly identified liaison contexts where a consonant through strong hesitation straddles W1 and W2 but ends up being onset of W2. Quite a few codings involve a liaison consonant before a strong hesitation *euh* (*elle a vingt[t]euh...*). Some of the codings involve a consonant before a consonant, which requires another code in the PFC system (e.g. *C'est[t] l'ancien nom*). Most of the examples are in fact hesitations involving repetitions (e.g. *On forme pas un chercheur en, en cinq ans*), to which we return below. Finally, nine of the codings involve the word *quand*. When we listened to them: eight of them were of the form *quand[t]euh*. But one coding of the word *quand* provides the only true *clean* 'liaison non enchaînée' we could find in our corpus: *quand[t][?]on envoie la balle*.<sup>17</sup>

The presence of liaison non enchaînée in the reading of the text contrasted with its quasi absence from spontaneous speech is in line with what many specialists feel. 'Liaison non enchaînée' is inextricably linked to the orthographical system and typically occurs in reading aloud or in situations where highly literate speakers are called upon to produce an elevated register. This knowledge and influence of the orthographical system is part of the so-called Buben effect (after Buben 1935) which has been described by a number of specialists (e.g. Chevrot et Malderez, 1999; Laks, 2005). Adequate psycholinguistic models need to be devised to reflect the connection between knowledge of speech and writing displayed by many speakers of French. In particular, it is clear that the presence of blanks between words in the written system helps French reinforces the autonomy of words. It does not however create it! Lyche and Girard (1995) have given a range of arguments showing that there are various generalizations conspiring to give words their identity. From the fact that stress is not a property of words but of phrases or rhythm groups in French (as stressed by Laks 2005), it does not follow that that there are no other cues as to word boundaries. In fact, much

<sup>17</sup> The frequency of *quand* pronounced [kãt] before consonants in our own observations outside the PFC corpus suggests that this word may well have been reanalysed as consonant final by a number of speakers. For discussions, see Encrevé (1988: 276-278) and Morin (1990).

experimental work on enchaînement tends to show the opposite (cf. Nguyen et al. to appear, Spinelli and Meunier 2005).

At this juncture, we return to some of the hesitations we have observed while examining ‘non enchaînement’. A number of them show the liaison consonant thrown backwards on to W1: e.g. *en, en cinq ans* [ãn ãsekã]; *tout, mais tout-à-fait* [tut metutafɛ]; *un, un Aveyronnais* [œn œnaverone] (by a Southern speaker). These examples seem to us extremely interesting: despite the clear predominance of ‘liaison enchaînée’ in our corpus, they provide possible evidence against an analysis which simply treats a liaison consonant as an onset of W2.

The previous remarks lead us naturally into some speculation concerning ‘liaison non enchaînée’. The latter, as far as we can see, arises within contexts of marked linguistic tension, whether it is the result of thought-construction or the normative effect of the linguistic market (in the Bourdieu sense). It seems to us that there is a continuum between examples such as *en, en cinq ans* [ãn ãsekã] and Encrevé’s prototypical examples such as *j’avais[z] j’un rêve*. We do wonder whether Encrevé did not present a somewhat idealized picture of ‘liaison non enchaînée’ which gave pride of place to the latter example at the expense of more murky cases. One of the reasons we ask this question is that the clear separation that Encrevé establishes between categorical liaison which for him always links forward and variable liaison which allows unlinking does not fully correspond to our own observations. Thus in Chirac’s November 2005 televised address, we observe the following instances of ‘non enchaînement’.

(7) Chirac, 14<sup>th</sup> November 2005

(i) celles qui connaissent de grandes difficultés doivent[t]# en revanche être activement soutenues

- (ii) il faut[t]# intensifier l'action contre les filières ...
- (iii) beaucoup[p]# a déjà été entrepris
- (iv) et plus particulièrement[t]# aux plus jeunes
- (v) notamment[t]# les jeunes en difficulté
- (vi) j'ai créé un service civil volontaire associant[t]# accompagnement et formation
- (vii) les représentants [...] doivent[t]# eux aussi refléter la diversité de la France
- (viii) mais sachons[z]# aussi nous rassembler pour agir
- (ix) ils font[t]# honneur à la République
- (x) les[z]# handicaps dont souffrent les plus vulnérables
- (xi) cet[t]# engagement financier important de la France

The last two examples do show that 'non enchaînement' is not confined to variable liaison and we have noticed in Encrevé's own examples at least one clear instance of unlinked categorical liaison: *dans son interprétation*, which he transcribes [dāsõnəʔẽtɛrpretasjõ] (1988: 38, 192-194), beside an example such as *une norganisation* [dõkünənɔrganizasjõ] (p. 194).

One of the ways in which the data may be idealized is the interpretation of the transitional vocalic elements in the 'non enchaînement' space as schwas which fill in a phonological place. While it has to be admitted that the spectrogram of [dāsõnəʔẽtɛrpretasjõ] given p. 37 (sonagramme 6) is remarkably clear and *prima facie* favours the classification of this 'epenthetic' vowel as a schwa, our own observations reveal much more 'hesitation' schwas of variable quality (within the IPA space [ø, œ, ə, ɜ, ɞ]) than clear 'phonological' schwas. A final complication is that different consonants do not appear to have the same linking capabilities. It seems to us that among the three main liaison contenders (t, z, n), the plosive [t] allows non linking and glottal stop insertion more easily than the fricative [z] and, in turn, that the latter is a better pseudo-coda than [n]. We have neither heard nor seen reported

pronunciations such as *certain*[n] [ʔ]atout, *bien*[n] [ʔ]utile. This is perhaps possible with structures such as *J'en ai un bon, enfant* [ʒãneœbɔn | ãfã] cited by Côté (2005) from Tranel (1990). But this pronunciation is not part of our linguistic experience, as opposed to the alternative [ʒãneœbɔ̃ | nãfã] supplied by Côté (ibid, p. 74, note 6) which we find much more natural. In the same way, we conjecture that in corpora of elevated speech one can contrast [r] (a strong forward linker if variable liaison is realized) with [p] (a frequent non linker in attested variable liaisons). A detailed phonetic investigation is warranted but if different consonants do not behave identically with respect to syllabic affiliations, it does complicate further the interpretation of 'liaison non enchaînée'. While the latter is hardly attested within our corpus, we are convinced that the whole gamut of hesitations combined with evidence from experimental phonetics and a careful consideration of morphophonological facts demonstrate that liaison consonants cannot be treated as pure and simple onsets of W2, which would be the naïve interpretation of the PFC codings providing 'liaison enchaînée' in approximately 99.5 of the cases and apportioning the rest to speech errors.<sup>18</sup>

#### **4. Elements for a theoretical treatment**

In this paper, we will not attempt a proper theoretical treatment of French liaison. Hundreds of articles, books, book chapters and oral presentations have been devoted to this question and it would be foolhardy to believe that a completely novel account can be offered. We have insisted on the importance of data in the preceding sections but, of course, while a theoretically motivated account must ideally describe and explain the data, it cannot be derived from it in a mechanical way. As stressed in Durand (1993), the data collected in large corpora such as traditionally sociolinguistic ones does not as such represent anybody's system. Even in close-knit communities, the grammars of individual speakers do not fully

<sup>18</sup> For further discussion of 'liaison non enchaînée', see Laks (to appear) which reached us too late to be integrated to our discussion.

coincide. The problem is even more severe for a database such as the PFC one which encompasses varieties and speakers sometimes widely separated in the geographical and social space. On the other hand, our database suggests that there are commonly shared patterns and it gives an approximation to the kind of variation most children come across in building their grammatical systems (whether this is the result of innate factors or not). It is surely more reliable than intuitions about potential sequences which would arguably never be produced spontaneously. Having said this, we do not believe in the necessity of an account which is uniform for all speakers and believe that the liaison grammars internalized by individuals may well vary at crucial points. For this reason, like Côté (2003), we will include dialectal variation as pointers to possible structural facts. We must also briefly clarify the relationship between the figures we obtain from our surveys and internalized grammatical systems.

If one examines the results derivable for a large database such as the PFC one, there are few if any generalizations which are truly categorical. It might therefore be argued that the grammar should mirror this and defend a statistical approach offering continuous scales from nearly obligatory to practically unattested. This seems to us unwarranted. Consider a sequence such as *les étés* (i.e. DET + N, where N begins with a vowel). In the varieties we have studied, the only examples we have of liaison not being made are in fact hesitations (see further down) in which speakers change their minds and move on to other structures. Equally, there are variations with words like *handicaps* (*les handicaps*). But these examples should not threaten the notion that there is a system-categorical constraint requiring liaison between e.g. *les* and the following noun if it begins with a vowel. In the case of *handicap*, we know that for most varieties there is a class of words or expressions beginning with a phonetic vowel (so called ‘h aspiré’ examples, varying between individuals and regions) which block liaison, whatever the ultimate formalization of this: e.g. *les // haies*, *les // haricots-paille*, *les hauts lieux*, in the PFC

corpus. In the examples of hesitation, which are by no means rare in a large corpus and which we have already mentioned, a proper treatment of any phenomenon must surely provide for an interaction between the grammatical system and other kinds of factors (memory, tiredness, thought-construction, etc.) as stressed by Chomsky in many writings. In the case of the determiner *les*, we checked against a database of 21 surveys (all the previous ones plus 4 Belgian investigation points) containing 28893 codings, there are (i) 1210 occurrences of liaison enchaînées (*les entrées, les autres*, etc.); (ii) 14 examples where liaison is not observed and which all contain either words or sequences treated as ‘h aspiré’ (*les oui, les haies, on fête les un an, les r minuscules, les a*); (iii) and, finally, 3 cases where a [z] is produced before a word beginning with a consonant and which all involve a clear hesitation: *les10z euh maladies des artères; s’est dit oui, comme tous les12z, tous les enfants* and, in the reading aloud of the text, *indiquerait que les, des activistes12z des communes voisines*. One does not have to adhere to all the Chomskyan tenets on the nature of the language-faculty to believe that there is a difference to be established between the linguistic system and its deployment in language-use (cf. distinctions such as ‘langue’ vs. ‘parole’, or ‘competence’ vs. ‘performance’, or ‘I-language’ vs. ‘E-language’). Ideally a terminological distinction should therefore be established between what belongs to the system and what is observed in the corpus. One could revert to the terms ‘obligatory’, ‘optional’ and ‘forbidden’ but these are so influenced by the norm that it seems preferable to avoid them. We will speak here of system-categorical and system-variable; system-absent is simply the complement of the liaison structures specified by the grammar. This will allow us to stress the difference between the observables for which we will speak of categorical, variable and erratic or unattested.

#### **4.1 System-categorical liaison**

For most varieties that we are aware of the following liaisons should be treated as system-categorical:

(8) Categorical contexts

(i) Det + vowel-initial X within an NP (un, les, des, mon, son, ton, mes, tes, ses) : *les enfants, les autres enfants, mon enfant...*

(ii) Proclitics (ils, elles, on, nous, vous, en): *on [n]en [n]avait parlé, il y en [n]a...*

(iii) Enclitics : *De quoi parle-t-[t]on ? Comment dit-[t]on ? Encore faut-[t]il travailler...*

(iv) Compounds and fixed phrases : *tout-[t]à-fait, toujours est-[t]il, pot-[t]au-feu...*

Within (8)(i), numerals pose a particular problem. For example, rather predictably *vingt ans* occurs several times as *vingt [t]ans* in our corpus but we have also observed examples such as *vingt [z]-employés* outside the corpus. The numerals are by no means straightforward and require a separate investigation and classification. Canadian varieties also demand further attention: e.g. *ils* before a vowel-initial verb can be realized without a plural [z] (e.g. [ijɔ̃] for *ils ont*); and Encrevé (1988: 53, note 39) also points out that *on* in *on est* is realized without liaison in 514 cases out of the 4087 cases examined by Tousignant and Sankoff (1979). In studying (8)(ii) and (iii), it must be recalled that a proper account must not deal with sequences but with structures. As correctly pointed out by Dell (1985: 43), a sequence like *allez+vous+écouter* cannot be pronounced [alevɛkute] if it corresponds to the imperative sentence *Allez vous-écouter!* On the other hand, [alevɛkute] is the normal realization of the interrogative sentence *Allez-vous écouter?* Apart from the intonation, liaison can therefore give strong cues as to the phrase structure of sentences.

The strong linking forward of the liaison consonant in categorical liaison has sometimes been attributed to the construing of the liaison consonant as a prefix of W2. While this insight is compatible with a subset of the data, it does not cover the whole range of possible examples. If we take an example like *les amis*, there is *prima facie* only one analysis

which seems to us wrong: the classical generative one (Schane 1968, Dell 1973/1985) which assigns the final consonant of *les* to the coda of the base. If this were the case, the vowel in systems where the ‘loi de position’ is alive such as the conservative variety of southern French described by Durand (1976, 1988, 1990, 1995) would be an [ɛ]. However, the normal southern realization of this sequence is [lezami] and not \*[lezami]. The consonant must be either extrametrical or floating (/le.z#ami/), or be introduced epenthetically between W1 and W2 (/le#z#ami/), or be treated as a special prefix of W2 (/le#z+ami/). The choice of a plural example does complicate matters here since in many cases the plural marker /z/ can be treated as a property of the construction as a whole. We return to this issue below. In the meantime, it is worth observing that other examples (e.g. *mon ami*, *ton ami*, *son ami*, *un ami*) do not require the final consonant to be a prefix. The conservative southern variety of French mentioned earlier does give a possible proof that the liaison consonant is not extrametrical, epenthetic or W2-prefixal. In that variety, the sequences *mon ami* (selected here as representative of the whole set) is pronounced with a non nasalized vowel [mɔnami], despite the highly nasal environment. Durand (1988) argues that the underlying structure of words like *mon* (pronounced [mɔŋ] or [mɔ̃ŋ] in isolation or before ‘h aspiré’ words) must be a VN sequence (which we could symbolize /mɔ<sup>N</sup>/), whether the nasal element is part of the coda, the nucleus or the second part of a complex heavy vowel. If the nasal was extrametrical, epenthetic or a prefix of W2 we would expect the vowel in such sequences to be mid-high, which again is not the case in the variety in question: \*[monami]. This suggests that a final consonant can be part of W1 and still linked forward in categorical liaison. If we contrast what we have said about *mon ami* with what was said earlier about *les amis*, we can see why a single solution to liaison phenomena is not the most profitable avenue to pursue. The liaison consonant can be a syllabic part of W1 (*mon*, *ton*, *son*, *un*) or external to it (*les*, *des*, *ces*, ...).

Let us now turn to plural cases of categorical liaison. In a number of insightful papers, Morin (2003a,b inter alia) has argued that some examples of spontaneous generalizations cannot be explained by assuming that liaison was epenthetic (e.g. Tranel 1981) or part of a construction (as defended e.g. by Bybee 2001, 2003). The [z] in examples such as *quatre-z arbres* could indeed be analyzed as generalization of the following construction put forward by Bybee:

(9) [NUMBER – z – [vowel]-NOUN ]<sub>Plural</sub>

on the basis of phrases such as *les [z] arbres, deux [z] arbres, trois [z] arbres, ... grands [z] arbres*. But this schema does not directly explain generalizations such as *C'est quoi comme, z-arbres? On prend quoi comme z-affaires, ...qui consiste en z-éléments indépendants, Je préfère ça (dans la) version z-années soixante*. These point to a prefixal [z], not to mention the following dialogue observed by Morin (2003 : 11):

(10) A: - Ça fait plusieurs mois que je mange des papayes.

B : [correcting A's statement]

- Plusieurs z-années, plutôt

C : [not convinced]

- Z-années !?, je ne crois pas.

These examples are extremely convincing. It should however be observed that when noun phrases consist of plural nouns in initial utterance position which are not echo repetitions of some other utterance, it does not seem that one observes plural [z] markers. For instance, the literature we have consulted does not seem to mention examples such as the following where an initial [z] intuitively seems very odd to us:

(11)

(ii) Etudiants (?[zetydjã]), venez tous nombreux à notre soirée théâtrale!

(iii) Avocats ou pas (?[zavɔkaup], je les déteste.

(iv) Enfants (?[zãfã]), femmes, vieillards, tout le monde a quitté la salle à toute vitesse.

Equally, similar sequences with adjectives preceded by a plural [z] seem odd to us:

(12) Utiles (?[zytil]) ou pas, il faut jeter toutes ces vieilles casserole.

These examples seem to suggest that for the plural [z] to be ‘wrongly’ assigned, it should ideally be in the air within phrases, as in the Damourette and Pichon (1911-1940) example quoted by Morin (2003: 11):

(13) ... il est pris de convulsions, d’abord *z-oculaires*, puis généralisées

Constructions where plurality can be shared by several lexical units are the ideal context for ‘over-spreading’ of the regular plural marker.

These remarks are not intended to provide a knock-down argument against Morin’s proposal. As a side piece of interesting evidence, Morin (2003: 14) quotes the dialogue below drawn from one of Luc Baronian’s surveys showing that in some varieties of Louisiana French the surface forms seem to be perceived as prefixal by naïve speakers:

(14) - Comment vous dites *cold*?

- [lœfrɛ]. C’est-à-dire, c’est selon l’histoire [lœ distwar] /c’est-à-dire le contexte/. Tu vois, pour un n-exemple, t’as *larbre*, *narbre*, *arbre* ou *zarbre* : *un narbre*. Tu vois, t’as /hésite et se reprend/ des fois t’uses le mot *larbre*, *narbre*, *arbre* ou *zarbre*. *Zarbre* veut dire « plus [plys] qu’un ». En anglais, t’uses **un** mot. Ça ne me gêne pas si y en a un ou i n’n a dix, c’est toujours le même mot. Et en français, t’as quatre mots pour *un narbre* ou *un n...* /hésite/ *un* /se reprend/ *larbre*. C’est, selon, comment t’appelles /hésite et se reprend/ dépeins ton discours, là tu uses le mot *zarbre*, parce que tu peux dire quand quelqu’un n’est pas habitué de parler en français un tas, il va dire *Regarde le gros narbre*. Non, non. Tu dis : *Regarde le gro arbre* [gro a:b], comprends?

From this standpoint, such varieties may come close to Michif, the mixed language of North American prairies which has a nominal system based on French and a verbal system based on Cree. According to Bakker (1977) a word like *ours* has three forms in Michif: [nu:r], [zu:r], [lu:r] which can be in free variation.

It should, however, be pointed out that the developmental scenario does lead to re-analyses by speakers which are more compatible with an epenthetic analysis according to some specialists (see Côté 2003: 68-70) than to a fully prefixal analysis. Moreover, while enchaînée it has been claimed that there are auditory cues below metalinguistic awareness which guide speakers in separating enchaînée liaison consonants from true onsets – an issue to which we return in the next section (Spinelli and Meunier, 2003). It should be pointed out that when a true consonantal prefix such as *r-* is added to a vowel-initial base: e.g. we have seen no evidence that the *r-* of *rajouter*, *rallumer*, *rouvrir* etc. was different from ‘true’ onsets as in *ramer*, *roussir* or *rivage*. There may therefore not be for all varieties of French or speakers of French a unique solution: while some systems may have moved to a prefixal stage as argued by Morin, we agree with Côté that system-categorical liaisons can still be plausibly be interpreted as ‘epenthetic’ for a number of varieties. It should nevertheless be recalled that the term ‘epenthetic’ is theoretically far from neutral and is not necessarily to be interpreted as a ‘process’ within a derivational (or more appropriately a transformational) approach. The plural [z] is an exponent of a morphosyntactic feature which varies according to specialists as to its source (is it the D or the N?) and which is usually be a property of phrases as wholes. This [z] can be interpreted as the result of rules of correspondence between independently motivated phonological and morphosyntactic structures à la OT or, within a monostratal sign approach like HPSG (see Bonami, Boyé, Tseng 2003), as a piece of morphosyntactic and phonological information which occurs at the juncture between words.

Let us next examine briefly the treatment of enclitics such as *en*, *y*, *il(s)*, *elle(s)*, which are found in imperative constructions and in questions with subject inversion such as: *vas-y*, *a-t-on*, *prends-en*, etc. Whatever the correct morphosyntactic analysis of such structures, the consonant is not phonetically part of its leftward lexical ‘source’: for instance, in the Midi French variety referred to earlier, the loi de position would once again not apply to ‘est’ in e.g. *est-on*, which is realized [etõŋ]. Whether ‘prefixal’ or not, the consonant in these structures is treated as an onset in the same way as final root consonants (fixed or ‘latent’) before vowel-initial suffixes (cf. the syllable-initial [t] in *pédantique* or *prophétique*). But, it is questionable, whether the obligatory liaison consonant in such structures can be simply carried over from the liaison consonant found elsewhere as assumed in many generative accounts. The case of the *a* form of *avoir* is of course well known (*Il a un ami*, *A-t-il un ami?*). But note that this seems far from exceptional. We pointed out earlier that in a sizeable portion of the PFC surveys liaison with *avait* was simply never made (e.g. *Il y avait // un mur*, *Elle avait // organisé la résistance*). By contrast, whenever inverted structures are attested, the liaison consonant is invariably present (e.g. *Peut-être avait-il mal compris?*). The same point could be made for other endings such as *-ez* or *-ons*.

Finally, one must deal with so-called compounds and fixed phrases such as *pot-[t]aux roses*, *tout-[t]à-fait*, *pied-[t]à-terre* vs. *pied//-à-pied*, *pot//-à-vin*. Laks (2005: 115-118) is correct in pointing out that neither the fixed nature of these units nor their liaison are predictable by general principle. Some expressions allow both treatments (e.g. *pas-à-pas*) with a high level of individual idiosyncrasy. Interestingly, the reading aloud of the text reveals that the sequence *jeux olympiques* (a prime example of obligatory liaison for Fouché 1959: 441) was read without liaison by 14 of the 195 speakers of the 21 surveys referred to earlier. All such cases contribute to the atomization of liaison and to a high level of lexical

coding as opposed to on-line computation. They also favour the treatment of the liaison consonant as an onset.

When we put together the points made in this section, our conclusion is that (quasi-) categorical liaison at the observational level is not the result of a single structure or strategy: as advocated by Côté (2003) it seems to us that the consonant can have its source either within the W1 word or in a position at the juncture of W1 and W2 either neutral as to syllabic affiliation or with clear onset characteristics.

#### 4.2 System-variable liaison

As we have stressed throughout, the use of corpora is an indispensable source of information regarding variable liaisons. Once again, we cannot conclude from the fact that a particular type of liaison is variable in the corpus as a whole that this is a property of the systems internalized by each speaker (in other words, that it is system-variable for each informant). If we consider liaison with *est* in the Douzens sub-corpus (excluding occurrences of *c'est-à-dire*), out of 166 relevant codings, liaison is attested on 102 occasions and not realized 64 times. There are two individuals for whom liaison is **always** attested (11ajp1 and 11aml1). But, while *est* liaison might indeed be system-categorical for these two informants, the problem is that there are only 12 codings concerning *est* for 11ajp1 and 17 for 11aml1. However, if we consult the corpus as a whole, out of 28893 codings for *est* (including all styles), liaison is made 1000 times and not made 1248 times. Earlier on, we noticed that for this particular parameter the southern varieties appeared to be nearer the orthoepic norm than the northern ones and this is compatible with the possibility that some speakers use liaison after *est* in a categorical manner. If we extrapolate from the database as a whole, it seems to us more likely that if we transcribed the whole recording for 11ajp1 and 11aml1 or made further recordings, we would find them not to use *est* liaison in a categorical manner. When there are too few tokens for individual speakers, it often seems to us a reasonable strategy to

use the database as a whole, all the more so as other results would support this decision (concerning *est*, see e.g. Encrevé 1988: 66-67). We do stress however that such decisions should not obscure the fact that individuals vary with respect to liaison items and liaison contexts. As correctly underlined by Encrevé (1988: 258):

Rappelons encore que, contrairement aux affirmations des ouvrages classiques, certaines liaisons sont “traitées” non pas catégorie par catégorie mais mot par mot pour certains locuteurs, qui lieront catégoriquement *c’est, pas* ou *très*, mais variablement *sont, dans* ou *chez*, sans qu’on puisse exclure que tel ou tel de ces choix soit plus “distinctif” (légitimant ou délégitimant), sur tel ou tel marché, qu’un taux moyen plus ou moins élevé sur l’ensemble des liaisons facultatives.’

One of the strengths of a corpus such as the PFC one, however, is that it allows us to be as fine-grained as we like from the data-set as a whole down to the performance of individuals for a given task.

In terms of the liaison consonants involved, only five consonants are attested in our corpus in liaison: /z, n, t, r, p/. Two of these /r/ and /p/ seem to be restricted to variable liaison. If we take the 195 speakers of the 21 surveys we had access to and leave the text aside, the figures we obtain are the following:

(15) Frequency of liaisons out of 2261 codings overall

[z]: 4544 > [n]: 3689 > [t]: 1665 > [r]: 13 > [p]: 9

Unlike many other earlier studies which treat [t] as second most frequent liaison consonant or as equal to [n] (e.g. Léon 1992: 152), but in accord with Green and Hintze (2001: 34), our figures place [n] in clear second position. The PFC annotation does not code structures involving enclitics (*Allez-y, Le veut-il?*) but even taking these into account does not substantially modify the overall picture. The drop between [z, n, t] and the other two players [r] and [p] also corresponds to a system-categorical vs. system-variable divide. We do not

have enough data to warrant our assertion but it seems that, for individuals who use [r]-liaisons and [p]-liaisons, these are restricted to variable environments. Thus, out of the thirteen [r]-liaisons, twelve involve *premier* (e.g. *premier [r]août*) and there is only one infinitive form (*m'installer[r]à Paris*). As for the nine [p]-liaisons, there are two with *beaucoup* (*beaucoup[p] occupé*) and seven with *trop* (e.g. *trop importantes*). The infinitive –*er* ending is particularly interesting: 677 verbs ending in –*er* in potential liaison environments do not liaise with the following vowel-initial word (as opposed to the one attested liaison cited above). It therefore seems quite likely that for many varieties or individuals, liaisons in [p] or in [r] are simply no longer options in the system (or may be variably restricted to a few collocations or items such as *premier* in line with Encrevé's quote given above).

If we extrapolate from our data to the systems internalized by speakers, system-variable liaison seems to be well-entrenched in the following contexts: Adj<sub>sg/pl</sub> + N, Adv + Adj, N<sub>pl</sub> + XP, Verb<sub>infl</sub> + XP, Prep + XP. Thus the corpus yields: *derniers [z]événements* vs. *paysans// importants*, *très [z]évident* vs. *très// âgée*, *petit [t]avantage* vs. *gros// immeuble*, *étaient [t]en vélo* vs. *étaient// avec moi*, *dans [z]un camping* vs. *dans// une soirée*. Whether individual grammars need to have recourse to such generic patterns has however been questioned by many specialists and in particular Bybee (2001, 2005, 2007) within construction grammar. Throughout the article we have given many examples showing the item and context-specificity of liaison which, on balance, favour Bybee's claim. If an inflectional ending such as –*ait* or –*aient* were uniformly coded for liaison we would expect similar liaison results for all verbs in the imperfect. However, what we find for most varieties is a complete dearth of realized –*ai(en)t* liaison outside the auxiliaries *éai(en)t* and *avai(en)t*! We note too that liaison with polysyllabic prepositions may have moved outside system-variable liaison for individual grammars. While our data needs to be supplemented, the number of attested liaisons for the following frequent prepositions is dismally low: *après*

(0/126 tokens), *avant* (2/9, with one occurrence of *avant-hier*), *assez* (1/33), *depuis* (0/13), *devant* (0/9), *pendant* (0/35).<sup>19</sup>

The status and affiliation of the liaison consonant in variable liaison remains a thorny issue. As for system-categorical liaison, the vowel which precedes the liaison consonant is not normally influenced by it and, in varieties like the traditional southern French mentioned earlier, the quality of the final W1 vowel is not compatible with an attachment to the coda of the last syllable of the word (e.g. *très occupé* [trez ɔkype], not \*[trez ɔkype]). It should, however, be noticed that the many treatments of such liaisons through simple epenthesis have never faced the fact that the information about the nature of the liaison consonant comes from W1 and not from W2 which only provides a free onset.<sup>20</sup> Moreover, once again nasal liaison poses a problem. In that variety, as indeed in standard French, *bon exercice* (selected here as representative of the whole set) is pronounced with a non nasalized vowel [bɔnɛgzɛrsisə] sequence (which we could symbolize /bɔ<sup>N</sup>/), whether the nasal element is part of the coda, the nucleus or the second part of a complex heavy vowel ending W2. If the nasal was extrametrical, epenthetic or a prefix of W2 we would expect the vowel in such sequences to be mid-high, which again is not the case in the variety in question: \*[bɔnɛgzɛrsisə]. This suggests that a final consonant can be part of W1 and still linked forward in variable liaison. At many points in this article, we have emphasized that the treatment of liaison consonants as simple onsets of W2 is somewhat reductionist even if it is fully compatible with a subset of the data.

With words like *trop* or *premier*, our observations of the conservative southern variety and many others are that the final W1 vowel is typically mid-high (tense): e.g. [tropabime]

<sup>19</sup> We have no liaison contexts for *concernant*, *durant*, *envers*, *nonobstant*, *selon*, *suivant*, *touchant*.

<sup>20</sup> Which is why the work of morphologists such as Bonami, Boyer and Tseng (2004, 2005) is to be taken seriously.

(*trop abîmé*) and [prømjɛrut] (*premier août*). But we certainly do not exclude the possibility of e.g. [prømjɛrut] *premier août*. If so, our own theoretical leanings would bring us closer to the morphological solution of e.g. Steriade (1999) than to Féry's (2003) phonological one (both summarised in §3.1). These difficult cases are better dealt with through appeal to suppletion and lexical storage than to rule or constraint reordering, which can obscure the marginality of some data.

Finally, a few words should be said about absence of liaison. As mentioned earlier, a theoretical account does not have to make provision for absent structures which are merely the complement of the structures which the grammar generates (i.e. specifies explicitly). It is nevertheless extremely useful to provide items or structures which prevent liaison as done, for example, in Walker (2001: 162-163) or Bonami, Boyé and Tseng (2005: 92-94). Our PFC data does not seem to reveal any case of singular noun (with a 'latent' consonant which is attested in derivation e.g. *bois – boisage – boiser – boisé - boiserie*) liaising with a following adjective (e.g. \* *un bois [z] immense*). Equally, we have not come across any liaison such as the following ones excluded by Bonami, Boyé and Tseng (2005: 94):

(15)(a) \* Paul dort [t] et Marie travaillait

(b) \* Paul doit acheter ces livres [z] ou les emprunter

But can we be sure that examples parallel to those in (15) would not occur among 'professionnels de la parole', given the apparent over-application of orthographically-based strategies by this group? And if so what would it mean for an account of the link between morphosyntactic structures and liaison? One should recall here that many of the predictions made by Selkirk (1972/1981) turned out to be indefensible. For instance, liaison is not blocked by traces left by Wh-movement and Clitic-movement. In an experiment partially duplicating Morin and Kaye's 1982 tests, Durand (1986: 166-167) found, in a reading task with ten subjects, that all of them made at least one liaison in sentences such as *Le courage*

*que sa présence donnait[t] à ces gens* or *Adolphe les mènerait[t] au pouvoir*, contrary to the prediction of trace theory. Nobody can deny the link between morphosyntax and liaison, the question is whether it supports any strongly articulated phrase structure models developed in the wake of X-bar and trace theories. At any rate, a comparative typology of liaison contexts in extensive corpora seems an essential task in the agenda of liaison specialists.

## **5. Conclusion**

In this paper, we have endeavoured to show that liaison phenomena require extensive data to be dealt with appropriately. The PFC approach, based on a stable methodology, seems to us a promising way forward and its extension beyond the survey points used here will allow us no doubt to (in)validate a number of current hypotheses. It also has the advantage of permitting an examination of the data ranging from the corpus as a whole to the performance of individuals via well specified regional varieties. We do not believe, however, that a motivated theoretical account can be mechanically extracted from the data as shown in §4 of this article. Future analyses will surely benefit from abstract model construction and a better understanding of the connections between the various ‘levels’ of description; but they will have to take explicitly into account the results of sociolinguistic surveys, acquisition studies, experimental phonetics as well as psycho-(neuro-)linguistic investigations, including the relationship between speech and writing. As stressed in Chevrot, Fayol and Laks (2005), these will have to acknowledge that French liaison is not a homogeneous locus but a multi-faceted phenomenon requiring us to accept, without demur, the crossing of disciplinary boundaries. Above all, future accounts will have to accept that intuition and recourse to past authorities are not sufficient to build a solid picture of French liaison. Philosophers of science are right to point out that theories are underdetermined by observations and that the greatest scientists have often flouted available observations to construct their models. On the whole, however, bad or insufficient data does not lead to good theories.



## References

- Ågren, J. (1973). Enquête sur quelques liaisons facultatives dans le français de conversation radiophonique. Uppsala: Acta Universitatis Upsaliensis.
- Armstrong, N. (2001). *Social and Stylistic Variation in Spoken French: a comparative approach*. Amsterdam: John Benjamins.
- Aub-Büscher, G. (1962). *Le parler rural de Ranrupt (Bas-Rhin). Essai de dialectologie vosgienne*. Paris: Klincksieck.
- Bonami, O., Boyé, G. and Tseng, J. (2004). An Integrated Analysis of French Liaison. In: G. Jaeger (ed.), *Formal Grammar 2004 Preproceedings*. Nancy, France, pp. 1-17.
- Bonami, O., Boyé, G. and Tseng, J. (2005). Sur la grammaire des consonnes latentes. *Langages* 158: 89-100.
- Borrell, M. and Billières, M. (1989). L'évolution de la norme phonétique en français contemporain. *La linguistique* 25 : 45-62.
- Boula de Mareuil, P., Adda-Decker, M. and Woehrling, C. (2007). Analysis of oral and nasal vowel realisation in northern and southern French varieties <http://www.limsi.fr/Individu/mareuil/publi/1240.pdf> , 6th International Conference on Phonetic Sciences, Saarbrücken, pp. 2221-2224.
- Brun, A. (1931). *Le français de Marseille*. Marseille : Institut historique de Provence. (Réédité en 1978, Marseille: Laffitte Reprints.)
- Bruneau, C. (1931). *Manuel de phonétique pratique*. 2<sup>nd</sup> edition. Paris: Berger-Levrault.
- Bybee, J. (2001a). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bybee, J. (2001b). Frequency effects on French liaison. In : J. Bybee and P. Hopper (eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, pp. 337-359.
- Bybee, J. (2005). La liaison: effets de fréquence et constructions. *Langages* 158: 24-37.

- Bybee, J. (2007). *Frequency of Use and the Organization of Language*. Oxford : Oxford University Press.
- Chevrot, J.-P. (2001). *Variation phonétique, développement, orthographe*. Dossier non publié présenté en vue de l'Habilitation à diriger des recherches. Vol. 1 : Synthèse. Université Stendhal – Grenoble 3.
- Chevrot, J.-P., Chabanal, D. and Dugua, C. (2007). Pour un modèle de l'acquisition des liaisons basé sur l'usage: trois études de cas. *Journal of French Language Studies* 17: 103-129.
- Chevrot, J.-P., Dugua, C. and Fayol, M. (2005). Liaison et formation des mots en français : un scénario développemental. *Langages* 158 : 38-52.
- Chevrot, J.-P. (2001). *Variation phonétique, développement, orthographe*. Dossier non publié présenté en vue de l'Habilitation à diriger des recherches. Vol. 1 : Synthèse. Université Stendhal – Grenoble 3.
- Chevrot, J.-P., Chabanal, D. and Dugua, C. (2007). Pour un modèle de l'acquisition des liaisons basé sur l'usage: trois études de cas. *Journal of French Language Studies* 17: 103-129.
- Chevrot, J.-P., Fayol, M. and Laks, B. (eds.)(2005). *La liaison: de la phonologie à la cognition*. *Langages* 158.
- Côté, M.-H. (2005). Le statut lexical des consonnes de liaison. *Langages*, 158 : 66-78.
- Coveney, A. (2001). *The Sounds of Contemporary French*. Exeter: Elm Bank Publications.
- De Jong, D. (1988). *Sociolinguistic aspects of French Liaison*. Unpublished Ph.D. Dissertation, Free University, Amsterdam.
- De Jong, D. (1994). La sociophonologie de la liaison orléanaise. In: C. Lyche (ed), *French Generative Phonology : Restrospective and Perspectives*, AFLS/ESRI, pp. 95-130.

- Delattre, P. (1951). *Principes de phonétique française à l'usage des étudiants anglo-américains*. Middlebury College.
- Delattre, P. (1966). *Studies in French and Comparative Phonetics*. The Hague: Mouton.
- Dell, F. (1973/1985). *Les règles et les sons*. 2<sup>nd</sup> edition, 1985. Paris: Hermann.
- Durand, J. (1976). Generative Phonology, Dependency Phonology and Southern French. *Lingua e Stile*, XI(1): 3-23. Also in *Essex Occasional Papers*, vol. 16.
- Durand, J. (1986). French liaison, floating segments and other matters in a dependency framework. In J. Durand (ed.) *Dependency and Non-Linear Phonology*. London: Croom Helm, pp. 161-201.
- Durand, J. (1988). Phénomènes de nasalité en français du midi. Phonologie de dépendance et sous-spécification. *Recherches linguistiques* 17: 29-54.
- Durand, J. (1990). *Generative and Non-Linear Phonology*. Londres: Longman.
- Durand, J. (1993). Sociolinguistic variation and the linguist . In: C. Sanders (ed.) *French Today: Language in its Social Context*. Cambridge: Cambridge University Press. 257-285.
- Durand, J. (1995). Alternances vocaliques en français du midi et phonologie du gouvernement. *Lingua* 95: 27-50.
- Durand, J. and Eychenne, J. (2004). Le schwa en français: pourquoi des corpus? In: T. Scheer (ed.) *Usage des corpus en phonologie. Corpus* 3. 311-356.
- Durand, J., Laks, B. and Lyche, C. (2002). La phonologie du français contemporain: usages, variétés et structure. In: C. Pusch and W. Raible (eds), *Romanistische Korpuslinguistik-Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen: Gunter Narr Verlag, pp. 93-106.

- Durand, J., Laks, B. and Lyche, C. (2005). Un corpus numérisé pour la phonologie du français. In: G. Williams (ed.), *La linguistique de corpus*. Rennes: Presses Universitaires de Rennes, pp. 205-217.
- Durand, J. and Lyche, C. (2003). Le projet 'Phonologie du Français Contemporain' (PFC) et sa méthodologie. In: E. Delais and J. Durand (eds.), *Corpus et variation en phonologie du français : méthodes et analyses*. Toulouse: Presses Universitaires du Mirail, pp. 212-276.
- Durand, J. and TARRIER, J.-M. (2003). Enquête phonologique en Languedoc (Douzens, Aude). *La Tribune Internationale des Langues Vivantes* 33 : 117-127.
- Durand, J. and Jean-Michel TARRIER (to appear). Deux thèses sur la transcription et le codage d'un grand corpus oral: le cas de PFC. In: M. Bilger (ed.) *Les enjeux de la transcription de la langue parlée*. Perpignan: Presses Universitaires de Perpignan.
- Durand, M. (1936). *Le genre grammatical en français parlé à Paris et dans la région parisienne*. Paris: D'Artrey.
- Encrevé, P. (1983). La liaison sans enchaînement. *Actes de la recherche en sciences sociales* 46 : 39-66.
- Encrevé, P. (1988). *La liaison avec et sans enchaînement*. Paris: Seuil.
- Eychenne, J. (2006). *Aspects de la phonologie du schwa dans le français contemporain: optimalité, visibilité prosodique, gradience*. Thèse de doctorat non publiée. Université de Toulouse 2 Le Mirail.
- Féry, C. (2003). Liaison and syllable structure in French. Postdam, Ms.
- Fouché, P. (1959). *Traité de prononciation française*. Paris: Klincksieck.
- Fougeron, C., Goldman, J.-P. and Frauenfelder U. (2001). Liaison and schwa deletion in French: an effect of lexical frequency and competition? In: *Proceedings of Eurospeech* 200. Aalborg: Danemark, pp. 639-642.
- Gadet, F. (2003). *La variation sociale en français*. Paris: Ophrys.

- Gadet, F. (2006). Réflexions sur les corpus oraux: que le linguiste sache ce qu'il fait. Colloque international, *Phonologie du français : du social au cognitif*, Paris, MSH, 7-9 décembre 2006. [http://www.projet-pfc.net/?u\\_s=4&u\\_a=101&](http://www.projet-pfc.net/?u_s=4&u_a=101&).
- Green, J. and Hintze, M.-A. (1990). Variation and change in French linking phenomena. In: J. Green and W. Ayres-Bennet (eds.), *Variation and Change in French*. London: Routledge, pp.61-88.
- Green, J. and Hintze, M.-A. (2001). The maintenance of liaison in a family network. In: M.-A. Hintze, T. Pooley and A. Judge (eds.), *French Accents: phonological and sociological perspectives*. London: AFLS and CILT, pp. 24-43.
- Kaisse, H. (1985). *Connected Speech. The Interaction of Syntax and Phonology*. Orlando: Academic Press.
- Klausenburger, J. (1978). Liaison 1977: the case for epenthesis. *Studies in French Linguistics* 1-2: 1-20.
- Klingler, T. (2006). PFC en terrain louisianais: défis et adaptation du protocole. Colloque international *Phonologie du français contemporain : données et enjeux théoriques*, Paris, MSH, 3-5 février 2006. [http://www.projet-pfc.net/?u\\_s=4&u\\_a=49&](http://www.projet-pfc.net/?u_s=4&u_a=49&).
- Lacheret, A. and Lyche, C. (2006). Le rôle des facteurs prosodiques dans l'analyse du schwa et de la liaison. In: A.-C. Simon (ed.), *Bulletin PFC* 6, ERSS, Université de Toulouse le Mirail et CNRS, pp. 27-49.
- Lacheret, A., Lyche, C. and Morel, M. (2005). Phonological Analysis of Schwa and Liaison within the PFC Project: how determinant are the prosodic factors? In *INTERSPEECH-2005*, [http://www.isca-speech.org/archive/interspeech\\_2005](http://www.isca-speech.org/archive/interspeech_2005), pp.1437-1440,
- Laks, B. (1980). *Différenciation linguistique et différenciation sociale: quelques problèmes de sociolinguistique française*. Thèse de doctorat, Université de Paris VIII-Vincennes.

- Laks, B. (1983). Langage et pratiques sociales. Étude sociolinguistique d'un groupe d'adolescents. *Actes de la recherche en sciences sociales* 46, 73-97.
- Laks, B. (2002). Description de l'oral et variation: La phonologie et la norme. *L'information grammaticale*: 5-11.
- Laks, B. (2005). La liaison et l'illusion, *Langages* 158 : 101-126.
- Laks, B. (to appear). Dynamiques de la liaison en français.
- Léon, P. (1971). *Essais de phonostylistique*. Paris : Didier.
- Léon, P. (1992). *Phonétisme et prononciations du français avec des travaux d'application et leurs corrigés*. Paris : Nathan.
- Lepelley, R. (1999). *La Normandie dialectale*. Caen: Presses Universitaires de Caen.
- Lyche, C. (2003). French liaison and Data. CASTL kick-off Conference, Universitetet i Tromsø, October 2-4, 2003.
- Lyche, C. (2006). Corpus en domaine francophone : la Louisiane. Ecole Thématique CNRS, *Linguistique de corpus oraux*, Nantes 19-24 June, 2006.
- Lyche, C. and Girard, F. (1995). Le mot retrouvé. *Lingua* 95 : 205-221.
- Malécot A., (1977). *Introduction à la phonétique française*. La Haye, Mouton.
- Martinet, A. and Walter, H. (1973). *Dictionnaire de la prononciation française dans son usage réel*. Paris : France Expansion.
- Martinon, P. (1913). *Comment on prononce le français. Traité complet de prononciation pratique avec les noms propres et les noms étrangers*. Paris: Larousse.
- Milner J.-C. and Regnault, F. (1987). *Dire le vers. Court traité à l'intention des acteurs et des amateurs d'alexandrins*. Paris: Editions du Seuil.
- Milroy, L. (1980). *Language and Social Networks*. Oxford: Blackwell.

- Morin, Y.-C. (1986). On the morphologization of word-final consonant deletion in French. In: H. Andersen (ed.), *Sandhi Phenomena in the Languages of Europe*. Berlin: Mouton de Gruyter, pp. 167-210.
- Morin, Y. C. (1987). French data and phonological theory. *Linguistics* 25: 815-843.
- Morin, Y. C. (2000). Le français de référence et les normes de prononciation. *Cahiers de l'Institut de linguistique de Louvain* 26-1: 91-135.
- Morin, Y. C. Morin, Y.-C. (2003) [1998]. Remarks on pronominal liaison consonants in French. In S. Ploch (ed.) *Living on the Edge. 28 Papers in Honour of Jonathan Kaye*. Berlin & New York: Mouton de Gruyter. 291-330.
- Morin, Y. C. (2005). La liaison relève-t-elle d'une tendance à éviter les hiatus. Réflexions sur son évolution historique. *Langages* 158 : 8-23.
- Morin, Y. C. and Kaye, J. (1982). The Syntactic Bases for French Liaison, *Journal of Linguistics* 18: 291-330.
- Morin, Y. C. (1990). La prononciation de [t] après *quand*. *Linguisticae Investigationes* 14 : 175-189.
- Nguyen, N. and Espesser, R. (2004). Méthodes et outils pour l'analyse acoustique des systèmes vocaliques. Version 1.0. In J. Eychenne and G. Mallet (eds.) *Bulletin PFC* 3, ERSS, 77- 85.
- Nguyen, N., Wauquier-Gravelines, S., Lancia, L. and Tuller, B. (to appear). Detection of liaison consonants in speech processing in French: experimental data and theoretical implications.
- Passy, P. (1892). *Les sons du français. Leur formation, leur combinaison, leur représentation*. Paris: Firmin-Didot.
- Pichon, E. (1938). Genre et questions connexes (sur les pas de Mlle Durand). *Le français moderne* 6 : 107-26.

- Sampson, R. (2001). Liaison, nasal vowels and productivity. *Journal of French Language Studies*, 11: 241-258.
- Schane, S. (1968). *French Phonology and Morphology*. Cambridge, Mass.: MIT Press.
- Selkirk, E. (1972). *The Phrase Phonology of English and French*. Ph.D. dissertation, MIT. (1980, New York: Garland).
- Smith, A. (1998). French variable liaison: a proposed simplification. *Francophonie* 17: 11-14.
- Smith, A. (1999). *Linguistic change on British and French public service radio*. Unpublished PhD thesis. University of Newcastle upon Tyne.
- Sobotta, E. (2006). *Phonologie et migration. Aveyronnais et Guadeloupéens à Paris*. Thèse de doctorat, LMU München/Paris X-Nanterre.
- Spinelli, E. and Meunier, F. (2005). Le traitement cognitif de la liaison dans la reconnaissance de la parole enchaînée. *Langages* 158: 79-88.
- Steriade, D. (1999). Lexical Conservatism in French Adjectival Liaison. In: B. Bullock, M. Authier and L. Reed (eds.), *Formal Perspectives in Romance Linguistics*. Amsterdam: John Benjamins, pp. 243-270.
- Tousignant, C. and Sankoff, D. (1979). Aspects de la compétence productive et réceptive : la liaison à Montréal. In P. Thibault (ed.) *Le français parlé. Etudes sociolinguistiques*. Edmonton: Linguistic Research Inc., pp. 41-52.
- Tranel, B. (1981). *Concreteness in Generative Phonology: Evidence from French*. Berkeley: University of California Press.
- Tranel, B. (1995). French final consonants and non linear phonology. *Lingua* 95: 131-167.
- Tranel, B. (1996). French liaison and elision revisited : a unified account within Optimality Theory. In: C. Parodi, C. Quicoli, M. Saltarelli and M.L. Zubizarreta (eds.) *Aspects of Romance Linguistics*. Washington, DC : Georgetown University Press, pp. 433-455.

- Tranel, B. (2000). Aspects de la phonologie du français et la théorie de l'optimalité, *Langue Française* 126 : 39-72.
- Walker, D. (2001). *French Sound Structure*. Calgary: University of Calgary Press.
- Wauquier-Gravelines, S. and Braud, V. (2005). Proto-déterminant et acquisition de la liaison obligatoire en français. *Langages* 158: 53-65.
- Watbled, J.-P. (1991). Les processus de sandhi externe en français de Marseille. *Journal of French Language Studies* 1: 71-91.
- Wioland, F. (1987). *Prononcer les sons du français*. Paris : Hachette.