1-16-2015 12:00 AM

Variable Glide Formation in Hexagonal French

Stephanie A. Kelly, *The University of Western Ontario*

Supervisor: Dr. François Poiré, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in French

VARIABLE GLIDE FORMATION IN HEXAGONAL FRENCH

Monograph

by

Stephanie Kelly

Graduate Program in Linguistics

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

This thesis examines phonetic and phonological aspects of gliding in Hexagonal French. In particular, we ask: Are glide phenomena as predictable as portrayed in modern descriptions? Do all three glides /j, w, ɥ/ or corresponding high vowels /i, u, y/ behave alike in all potential glide contexts? Given the duality of French glides (vowel and consonant), we use the term *vocoid* and the archiphoneme convention /I, U, Y/ in our discussion of glide contexts and glide phenomena. Our historical survey shows the glides of French (/j, ɥ, w/) evolve separately and during this period the high front vocoid /I/ occurs early and is involved in greater variety of contexts showing considerable variability. The other two glides emerge later, primarily through diphthongisation, and show less variability.

In a study of glide contexts in the spontaneous speech of native speakers from three regions of France (data from the *Phonologie du Français Contemporain* project), we examine the distribution of all three high vocoids and their surface realisations. For 3415 tokens identified, we determine if HVV (high vocoid plus vowel) tokens are realised with dieresis, with syneresis, or with the high vocoid deleted. Our findings show glide contexts are consistently distributed at a rate of about 85% lexicalised and 15% derived. The limited variability in lexicalised contexts involves mainly the non-round vocoid /I/ realised with dieresis. Distribution across the three-glide inventory of French shows that lexicalised glide contexts follow a general markedness hierarchy: I ≫ U ≫ Y. Tokens involving the front non-round vocoid /I/ are most prevalent followed by the back rounded vocoid /U/ and finally the front rounded /Y/. Derived contexts include word medial tautomorphemic high vowel + vowel /HV+V/ sequences resulting from suffixation or inflection, and also cross-word-boundary /HV+V/ sequences which have very rarely been studied before; we show that cross-word-boundary data largely follow the same phonological constraints as derivational data. In each of these contexts the general markedness hierarchy observed above is changed, giving preference to the front rounded /Y/ over the back rounded /U/ while /I/ remains most prevalent.

Keywords

Hexagonal French, high vowel, vocoid, glide, semi-vowel, semi-consonant, diphthong, sandhi effects, liaison, elision, epenthesis, syneresis, dieresis, variation, derived, derivation, inflection, lexicalised, lexical exceptions, syllabification, syllable theory, distinctive features, spontaneous speech, phonetics, French phonology, PRAAT, acoustic analysis, autosegmental analysis, lexically preassociated segments, Markedness Hierarchy, Optimality Theory, la phonologie du français contemporain (PFC)

Acknowledgments

My sincere thanks to my supervisor François Poiré for his time and patience in guiding my doctoral research to a successful conclusion, and for his intellectual leadership during his years as Canadian director of the *Phonologie du français contemporain* project. Our Windsor PFC survey left me hungry for more fieldwork and linguistics. I would also like to express my profound gratitude to my second thesis reader Ileana Paul, as well as to my other committee members and examiners, Jeff Tennant, Yasaman Rafat and Svetlana Kaminskaïa, for their detailed comments and insightful, constructive feedback on this thesis.

This research would not have been possible without Jacques Durand (Université de Toulouse II - Le Mirail) and all the collaborators of the *Phonologie du français contemporain* project who gave access to their data.

For professional support and academic advice, I am also endebted to Ioana Chitoran, Wlad Cichocki, Gabor Turcsan, Sheryl Sawyer, Sylvain Navarro, Géraldine Mallet, and to Paul Boersma for all his help with PRAAT while creating the judgement task.

For their administrative, professional and personal support, I am grateful to Chrisanthi Skalkos, Deb Smith, Mirela Parau and Tony Purdy of the French Studies department, and to Paula Menzies-Cameron of the School of Graduate and Postgraduate Studies, as well as to my TAs in the Linguistics Graduate Program at the University of Western Ontario.

This research was supported by a doctoral scholarship (767-2005-1834) from the Social Science and Humanities Research Council of Canada, and by a mobility scholarship from the French government. Other support received from the Department of French Studies (UWO) and from the ERSS-CNRS (Université de Toulouse II - Le Mirail) was also instrumental in allowing me to conduct my doctoral research.

A certain number of faithful friends have accompanied me throughout this process, and made the experience more bearable: Darcie Blainie, Anne-Josée Villeneuve, Rose Ricci, Anna Moro, Valérie Prat and Olga Kharytonava. Grazie, merci d'avoir été là.

I am grateful to my parents Jean and Jim Kelly for their constant and unquestioning support, to my sisters Lynn and Lois for understanding, and for the strong female role models provided by my grandmothers, Mary Claude Brown and Marie Aurore Laporte.

Words and space are insufficient to express what David, il cuore mio, brings to all my successes. Wherever we are, you are my home.

This thesis is dedicated to the fond memory of Negrita, my constant companion and unconditional confidant for many years.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviations used in this thesis:

V	Vowel
HV	High Vowel
G	Glide
GV	Glide +Vowel
OL	Obstruant + Liquid
C	Consonant
( )	Segment contained between parentheses is optional; for example HV(G)V indicates the optionality of epenthetic glides[1]
CW	Cross-word-boundary
#	Word boundary
$	Syllable boundary
'.'	Syllable boundary in phonetic transcriptions

[1] Parentheses are used in our empirical study to indicate assimilation observed in the segmental realisation (devoicing).

List of Appendices

Chapter 1

1 Introduction

The goal of this thesis is the study of glides in French with a view to advancing a better description as well as an improved understanding of glide phenomena in various contexts of Modern French. By glide, we are referring to such sounds as [j] (*bière*, *yeti*), [w] (*boire*, *ouïe*), and [ɥ] (*puis*, *huile*). While these are often referred to as semi-vowels or semi-consonants, we prefer to use glide, –a term which to us seems much more theoretically neutral than those other names that are traditionally used.

French provides a particularly fertile terrain for the study of glides and gliding phenomena. Standard French (SF) has a three-glide inventory /j w ɥ/ each of which corresponds to a high vowel /i, y, u/. A three-glide inventory is typologically unusual since the vast majority of glide languages include only two ([j] and [w] as in English yet and wet, for example). We discuss implications for the three-glide inventory of Modern French in Markedness Theory in Chapter 4.

French glides figure prominently in the descriptive literature and in theoretical accounts of Modern French (Schane 1968, Dell 1973, Kaye and Lowenstamm 1984, Durand and Lyche 1999, Bullock 2002, Hall 2006, cf. Chapters 3 and 4). Descriptions and theoretical discussions often focus on a set of glide data that when examined in isolation sometimes lead to generalisations based on their apparent predictability in a given environment. Sandhi effects, for example, can provide diagnostic insights as to a speaker's treatment of a particular glide. These data show that the very same glide may pattern with vowels or it may pattern with consonants (*les yeux* [lezjø] 'eyes' but *les yetis* [lejɛti] 'yetis'). In some cases, the duality of glides may be attested for a single word (*la ouate* [lawat] but *l'ouate* [lwat] 'cotton batting').

Glide formation (GF) is a process typically described in word medial contexts where suffixation or inflection results in a stem final high vowel followed by another typically non high vowel. Throughout this thesis, we use the term 'derived' as in derived glide or derived glide context to refer to any tautomorphemic high vowel vowel sequence (HVV,

cf. List of Abbreviations xxvii). A tautomorphemic HVV occurs as a result of a morpheme final high vocoid meeting a vowel-initial morpheme which can occur either in a case of inflection or (morphological) derivation or in contexts where word-final HV is followed by a vowel initial word (cross-word-boundary). Derived glide contexts can be distinguished from lexicalised glide contexts where we observe that the lexicalised glide occurs in a monomorpheme. In derived contexts, a rule of glide formation may operate as a strategy to avoid hiatus. Thus, a HVV sequence results in syneresis (GV, glide plus vowel). In the case that GF does not apply the HVV sequence remains (realised with dieresis with or without epenthesis). In an empirical study of glides in the spontaneous speech of continental French speakers, we examine glide phenomena in three contexts: lexicalised (*pied*, *loi, lui*), word medial derived (*lier*, *jouer, suer*), and also potential glide formation across word boundaries (*qui est, ou on, vu aussi)*. For each of the contexts outlined here, we examine outcomes for glide contexts with a goal of describing any variabilty observed for each context and for each high vowel or glide in evidence. Our data come from semidirected interviews of the international collaborative project *la Phonologie du Français Contemporain* (PFC, cf. Durand, Jacques, Bernard Laks & Chantal Lyche 2002, 2009, also Chapter 5).

Our research questions are as follows. Are all glides alike: equally distributed, equally stable (or variable) in all contexts? Are all glide processes as predictable as is described? And is this the case for all contexts (and by extension the specific environments in each context)? And where variability is observed, is it observed equally for all three glides? If we do observe variable realisations (syneresis versus dieresis), how can the variability of glide contexts inform a theoretical model of glide data in French?

In lexicalised contexts, for example, glides are described as invariably realised with syneresis (Tranel 1987). Derived glide contexts are also typically described realised with syneresis as a result of a rule of glide formation which operates equally on all three high vocoids.

Cross-word-boundary contexts present significant challenges since, to the best of our knowledge, no formal studies of glide formation systematically examine this context, and

as such, very little is known about glide phenomena in this environment. As a context where glide formation may occur, our literature review (Chapter 1-4) shows that although cross-word-boundary glide formation is noted on occasion in descriptions of spontaneous speech (cf. Nyrop 1899, Passy 1913, Lyche 1979) discussion for implications of glide formation in cross-word-boundary environment is often relegated to footnotes. There are no known studies that treat cross-word-boundary glide formation extensively.

Our thesis is structured as follows. In the rest of Chapter 1, we illustrate the depth of the complexity of French glide data as we discuss the details of glide descriptions of Modern French. We examine derived and lexicalised contexts, and any variability that may be observed in these contexts. By their nature derived glides result from phonological operations while lexicalised glides can be linked to historical processes in operation during the evolution of the French language. In a survey that spans almost 300 years of research on glides including the early grammatical and early phonetic traditions, we examine the emergence and evolution of glides in French and show that glides have shown considerable variability throughout the history of the language. Their relative uniformity in Modern French contrasts with their variable history.

In Chapter 2, we discuss the emergence of the classical tradition of phonetics and its implications for glide descriptions. At the inception of the International Phonetic Association (IPA), the work of Passy (1913) and Jones (1932) are at the forefront of their field of phonetic description. We show that the duality of glides, at times vowel-like while at other times consonant-like, presents persistent difficulty for a phonetic description focussed on assigning all sounds in language to one or the other of consonant or vowel classes. The phonetic descriptions based solely on describing the phonetic qualities perceived proves ill equipped to adequately describe glides. By the mid 20th century glide treatments are improved with the emergence of the phonological tradition exemplified in Pike (1943) and become more consistent as the goals of phonetic description as redefined with the expanding study of phonology. Finally in this chapter, we discuss the emergence of the phonological tradition in Prague School phonology.

In Chapter 3, we examine research on glides in the linear generative framework (Chomsky and Halle 1968, Schane 1968, Dell 1973, Morin 1976, Lyche 1979). The various discussions surrounding glide treatments in the linear framework bring to light the difficulty that glides and their duality present for a model that relies on distinctive features to distinguish between consonants and vowels.

In Chapter 4 we examine the early non linear generative models that follow the linear generative model, beginning with glide analyses in Kaye and Lowenstamm (1984). This is the first model to represent the role of the syllable in distinguishing between high vowels and their corresponding glides. Further discussions of glide treatments for French and for other languages including, Guerssel (1986), Nespor and Vogel (1986), Encrevé (1988), Carreira (1989), Hannahs (1995), Lyche and Girard (1995) shed further light on the discussion of glide phenomena. Finally we turn our discussion to markedness (typology, cross-linguistic distribution, child language acquisition) before turning the discussion to glide treatments in Optimality Theoretical analyses (Rosenthall 1997, Durand and Lyche 1999, Bullock 2002, Hall 2006).

In Chapter 5, we describe the goals, methodology, and results for a judgement task devised as a methodological tool to aid in our primary study of glide formation in a spontaneous speech sample of Continental French. This judgement task, based on a PRAAT script, wherein three linguistically informed judges who are native speakers of Continental French, are asked to determine whether audio samples of HVV tokens are realised with syneresis or dieresis. Our judgement task includes 782 /HVV/ tokens (both lexicalised and derived). The goals of the judgement task are: to establish the extent of variability observed for /HVV/ and /HV+V/ surface realisations in Continental French; and to establish, from those tokens that result in categorical judgements from our panel of judges, which acoustic cues serve to reliably distinguish a realisation with syneresis from a realisation with dieresis on a spectrogram.

The data for each of our studies come from the *Phonologie du français contemporain: usages, variation et structure* (PFC) corpus (Durand, Jacques, Bernard Laks & Chantal

Lyche 2002, 2009), a collaborative research project that aims to expand and renew the empirical foundation of the description of French.

In Chapter 6, we discuss the methodology and present the results of our primary spontaneous speech study. With past analyses in mind, the scope of our study is not limited to that artificial construct known as Standard French (Morin 2000) but rather we analyse glide phenomena in the speech of three geographically separate regions (Île de France, Normandy and the Rhône Alpes regions) to determine to what extent glide phenomena are homogeonous and regularised in Continental French at the turn of this new millennium. We identify 3415 tokens for /HV(+)V/ contexts in interviews from 31 speakers in six PFC surveys. Using the acoustic cues identified in our first study, we code tokens for realisations with syneresis and with dieresis in PRAAT. We examine the prevalence of glide formation (syneresis), realisations with dieresis and deleted high vocoids in /HV(+)V/ contexts for all three of the high vowels of French /i y u/.

Chapter 7 provides some analyses of our results for realisations of glide contexts and conclusions. Distribution of tokens across three high vocoids follows a general markedness hierarchy wherein /I/ dominates /U/ which in turn dominates /Y/. While /I/ also dominates the markedness hierarchy in derived glide contexts, we consistently observe a changed hierarchy wherein the front rounded high vocoid dominates the back rounded: /I/ ≫ /Y/ ≫ /U/. We examine different mechanisms for representing variation – lexical exceptions versus variable OT grammars. Across all our data we observe that the high vocoid /I/ is always most prevalent and most variable. All variability observed in lexicalised contexts, for example, involve /I/.

1.1 French Glide Data: derived and lexicalised contexts

Following from the descriptive literature (Tranel 1987, Walker 2001) and theoretical accounts (Schane 1968, Kaye and Lowenstamm 1984), we distinguish two categories for glides in French: derived and lexicalised. Lexicalised glides are present in the lexicon from their acquisition and as such they are most often considered invariable, whereas derived glides occur when, as a result of derivation or inflection involving a vowel initial morpheme, a high vowel is in a hiatus /HV+V/. In the following sections, we discuss in detail these categorisations and their implications for our study of variable glide formation (GF) in Hexagonal French.

In section 1.1.1 we discuss the issue of phonemicity for French glides and our categorisation of the glides studied: lexicalised and derived glide contexts. In sections 1.2 and 1.3, we discuss derived and lexicalised glides while providing examples for each context described. These are not intended to be exhaustive inventories of derived and lexicalised glide data, but rather an illustration of their presence, prevalence, and distribution in the inventory of Modern French glide contexts.

1.1.1 Phonemicity and categorisation of glides in French

Given the complementary distribution between high vowels and corresponding glides, French glides are not usually listed among phonemes. According to Levi, while consensus on the status of glides is not always clear, we often observe glides listed in the phonemic inventories for languages (2004: 1).

Phonemes typically exhibit an unconstrained distribution while their allophones are found in predictable environments. The predictability of derived glides combined with alterability with a corresponding high vowel in French leads to the position that glides and their corresponding high vowels are best considered two allophones of a single phoneme (Martinet 1945, Schane 1968, Kaye and Lowenstam 1984, Durand 1987).

Postvocalic yod presents a context apart, because this is the glide that is most often considered an underlying segment. According to Martinet “If it is necessary, in French, to

distinguish a phoneme /j/ from a phoneme /i/, this is because *paye* and *abeille* are not identical to *pays* and *abbaye*" (1960:72). French glides are not all alike and can be distinguished by their distribution and their underlying source; yod [j] alone occurs in word-final and syllable final position, for example. Given these facts, yod is sometimes described as a phoneme in French (Martinet 1960, Léon 1978). Levi (2004) discusses at length the difficulties of including glides in phoneme inventories:

> When faced with a near-minimal pair such as [taj] and [ta.i], it is necessary to look further into the language to determine the true representation of this contrast. Processes that target features and either group [j] and [i] together or differentiate between them help determine which analysis—'glide' or 'vocalic'—is better suited to the language. Similarly, finding the basic pattern of the language—hiatus or gliding—can also clarify which solution is better.
>
> (Levi 2004: 5)

In final position, *yod* [j] can be seen to pattern with consonants having the effect of closing a syllable and triggering a vowel adjustment (*payer* [pe̱je] 'to pay' ~ *paye* [pɛ̱j] 'pays'). For this reason we might include word-final yod among consonant phonemes in French. However, the status of word-final yod is not at issue here, and is therefore counted in the category of lexicalised glides (*rien* [ʁjɛ̃]*, loi* [lwa]*, nuit* [nɥi], for example). Unlike the glides we categorise as derived, these glides are not the result of a rule of glide formation that takes place after processes of suffixation. Thus we describe as lexicalised all glides present in the lexicon, while derived glides are the result of tautomorphemic high vocoid + vocoid sequence (i.e, a stem final high vowel meets another vowel initial morpheme). Often this environment is the result of suffixation (derivational or inflectional), or, as mentioned in Nyrop (1899), Passy (1913, section 2.1), Lyche (1979, section 3.4.2) glide formation can result across a word boundary when the /HV+V/ environment occurs.

Typically lexicalised glides are considered invariable. Evidence of high vowel glide alterability in lexicalised contexts is rare. Pike (1943, cf. section 2.5.1)) brings to the forefront the question of high vowel and glide alterability relative to distribution within syllabic constituencies while also coining the term vocoid to ameliorate the confusion of naming glides given their relationship with a corresponding high vowel (semi-vowel or

semi-consonant). It is only after Kaye and Lowenstamm (1984) that the high vowel phonemes of French are more precisely described as the high vocoid series /I Y U/. Each high vocoid /I Y U/ represents both possible allophones: high vowel and corresponding glide: /I/→[i] or [j]; /Y/→[y] or [ɥ] finally /U/→[u] or [w] (cf. Kaye and Lowenstamn 1984, Chapter 4). In this perspective, the high vowel occurs in syllable peaks while the glide occurs in non-peak positions only. In this model, a high vocoid may occur in all positions within the syllable (onset, nucleus, coda); its phonetic realisation to be determined at the surface level after syllabification or resyllabification. It is a particularly elegant account since some glides behave like vowels while others behave like consonants. In this model, derived glides are the non-nuclear realisations of /I, U, and Y/.

In the following section, we discuss further our categorisation of derived glide contexts with an examination of the glide data described for each of the various contexts observed. In 1.2.1, we illustrate how the process of glide formation occurs and its relationship to the high vowels. In 1.2.2, we examine the process of glide epenthesis and compare it to glide formation since both are strategies for hiatus avoidance. Since both of these processes are optional we will discuss the variability attested in section 1.2.3. We will show that glide formation and epenthesis, triggered by the surrounding phonological environment and permissible syllable structure, transform a high vowel in hiatus to a more desireable surface form ([GV] syneresis or [VGV] dieresis).

1.2 Derived Glides

French high vowels and glides are inherently linked in derived contexts. According to Gougenheim, for example, "[l]es *semi-consonnes*, que l'on appelle aussi *semi-voyelles*, ne sont autres que les trois voyelles extrêmes, *i*, *u*, *ou*, prononcées comme des consonnes [...] " (1969: 29). According to Durand " [...] in so many languages, the simplest analysis should treat these semi-vowels simply as the high vowels in a non-syllabic disguise (1989: 95)". Like Gougenheim and Durand, others advocate for a position that glides /j ɥ w/ are merely consonantal versions of the corresponding high vowels /i y u/ (Martinet 1945, Schane 1968, Kaye and Lowenstamm1984). Following Rey (1967) and Warnant

(1968), we use [ʀ] to represent the rhotic phoneme in our own phonetic transcriptions however many authors also use [r] which we leave unchanged in their examples.

This position that all glides are simply non-syllabic realisations of high vowels is supported by the fact that glides are never syllabic (*[bwd], *[ljs], *[fɥt], for example). High vowel~glide alternations of derived contexts are therefore considered consonantal associated to an available onset position (1):

(1) Derived forms

a) i~j: *étudie* [etydi̲] 'studies', *étudiant* [etydj̲ɑ̃] 'student' *étudier* [etydj̲e] 'to study',
b) u~w: *joue* [ʒu̲] 'plays', *jouable* [ʒw̲abl] 'playable', *joueur* [ʒw̲œːʀ] 'player',
c) y~ɥ: *attribut* [atʀiby̲] 'attribute', *attribuable* [atʀibɥ̲abl] 'attributable', *attribuer* [atʀibɥ̲e] 'to attribute'),

Syllabicity is widely seen as the primary motivating factor in glide formation in derived contexts (cf. Kaye and Lowenstamm 1984). However, it cannot not account for lexicalised glides such as *trois* [tʀwa] *[tʀua], *truite* [tʀɥit] *[tʀyit].

In the following sections, we will discuss the details surrounding the processes of glide formation (section 1.2.1) and glide epenthesis (section 1.2.2) and the tendency to under report variability in descriptions of these data. In section 1.3.3, we discuss the extent to which variability is observed in the realisation of derived glides. Lexicalised glides are by contrast much more regular, with some exceptions attributable to stylistic variation, particularly for j~i (Martinet 1945 cf. section 2.2.2, Bullock 2002 cf. section 4.7). We examine these cases in section 1.3. where, for each of the various categories outlined in (1), we will discuss their origins in the diachronic evolution from Latin to French. We will examine evidence of normative prescriptive pressures influencing treatments of these forms by way of the 17th century grammatical tradition as well as some empirical evidence from the *Atlas Linguistique de la France* published at the turn of the last century surveying the non-standard speech of rural regions across France in the 19th century.

1.2.1 Glide Formation in Hexagonal French

Glide formation as shown in (3), is one possible outcome when phonological derivation results in a potential hiatus. According to Tranel, "[a]s a rule when a vowel-initial suffix is added to a word ending in a closed vowel, the closed vowel changes into the corresponding glide by a process called gliding ('semi-vocalisation')" (1987: 118). Following Morin (1976) and Rosenthall (1997), we prefer the term glide formation (GF) to refer to the process that Tranel and others call 'gliding' or 'semivocalisation'. Glide formation is most commonly demonstrated with first group (*–er*) conjugations as shown in (2).

(2) Glide formation (syneresis): /HV+V/→[GV]

a. [i→ j]: *lier* /li+e/→ [lje] 'to tie'
b. [u→w]: *jouer* /ʒu + e/ → [ʒwe] 'to play'
c. [y→ɥ]: *suer* /sy + e/ → [sɥe] 'to sweat'

In (2) above, we show that all three French high vowels [i y u] undergo glide formation. Glide formation must be considered optional, however, since hiatus sequences do occur in derived contexts but also in lexicalised forms (*maïs* [mais] 'corn', *poème* [pɔɛm] 'poem', for example, cf. Tranel 1987, Noske 1993, Rialland 1994). We note also that many of these lexicalised forms might still undergo modification that resembles glide formation. Since hiatus is not banned by the grammar the application of glide formation cannot be a strict grammatical imperative. But does the outcome (syneresis versus dieresis) have phonological implications?

In (3) we show that glide formation is prevalent in derivations that involve a variety of vowel initial suffixes.

(3) Derived glide contexts /HV+V/→ [GV]

a. –able: /j/- *aimiable* ‘kind’ [amjabl], *fiable* ‘reliable’ [fjabl],
/w/- *louable* ‘laudable’ [lwabl], *jouable* ‘playable’ [ʒwabl], *avouable*
/ɥ/- *évaluable* ‘assessable’ [evalɥabl], *attribuable* ‘attributable’ [atribɥabl],
–ant:
/j/- *confiant* ‘confident’ [kɔ̃fjɑ̃], *conciliant*, ‘conciliatory’ [kɔ̃siljɑ̃],
/w/- *n/a*
/ɥ/- *puant* ‘smelly’ [pɥɑ̃], *remuant* ‘restless’ [rəmɥɑ̃], *suant* ‘sweaty’[sɥɑ̃],

b. –euse: /j/- *envieuse* ’envious’ [ɛ̃vjøz], *prodigieuse* ‘prodigious’ [prɔdiʒjøz],
/w/- *joueuse* [ʒwøz] ‘player’
/ɥ/- *tortueuse* ‘tortuous’ [tɔrtɥøz]

(Juilland 1967)

Given the nature of derived contexts, glide formation necessarily occurs across morpheme boundaries (stem + suffix). In these morphologically derived contexts, we take the stem final high vowel of the root morpheme as the underlying form from which glides may be derived resulting from the addition of a vowel initial suffix (3a.): /ami + able/ → [amjabl], /ʒu + able/ → [ʒwabl], /atriby + able/ → [atribɥabl]. In the following examples glide formation results in monosyllabics (4) as well as plurisyllabics (5) preceded by any single consonant (#CG).

(4) Glide formation resulting in monosyllabic forms

Preceding phonological environment: $C

a. /i/→[j]: *fie* [fi] 'relies on', *fier* [fje] 'to rely on'
lie [li] 'ties', *lier* [lje] 'to tie'
nie [ni] 'denies', *nier* [nje] 'to deny'
scie [si] 'saws', *scier* [sje] 'to saw'

(Juilland 1965)

b. /u/→[w]: *doue* 'equips', *douer* [dwe] 'to equipe'
loue [lu] 'praises' , *louer* [lwe] 'to praise'
noue [nu] 'knots', *nouer* [nwe] 'to knot'
roue [ʀu] 'beats', *rouer* [ʀwe] 'to beat'
voue [vu] 'vows', *vouer* [vwe] 'to vow'

(Rey 1967)

c. /y/→ [ɥ]: *hue* [hy] 'shouts/cries/hollers', *huer* [ˈhɥe] 'to shout/cry/holler'
mue [my] 'molts', muer [mɥe] 'to molt'
pue [py] 'stinks/repugnates', *puer* [pɥe] 'to stink/repugnate'
rue [ry] 'kicks', *ruer* [rɥe] 'to kick'
sue [sy] 'sweats' , *suer* [sɥe] 'to sweat'
tue [ty] 'kills', *tuer* [tɥe] 'to kill'

(Juilland 1965)

The data in (4) show that the preceding environment is unrestricted for consonant class; Glide formation occurs for all three high vowels when they are preceded by any single consonant: stop, nasal, fricative, liquid.

The data in (5) are similarly unconstrained for preceding environment, glide formation occurs for all three high vowels in a number of plurisyllabic words.

(5) Word medial glide formation

Preceding phonological environment: $C

a. /i/→[j]: *défie* [de.fi] 'challenges', *défier* [de.fje] 'to challenge'
concilie [kɔ̃.si.li] 'reconciles', *concilier* [kɔ̃.si.lje] 'to reconcile'
circonstancie [sir.kɔ̃.stɑ̃.si] 'details', *circonstancier* [sir.kɔ̃.stɑ̃.sje] 'to detail'

b. /u/→[w]: *secoue* [sə.ku] 'shakes', *secouer* [sə.kwe] 'to shake'
échoue [e.ʃu] 'fails' *échouer* [e.ʃwe] 'to fail'
désavoue [de.za.vu] 'disavows', *désavouer* [de.za.vwe] 'to disavow'

c. /y/→ [ɥ]: *dénue* [de.ny] 'deprives', *dénuer* [de.nɥe] 'to deprive',
continue [kɔ̃.ti.ny] 'continues', *continuer* [kɔ̃.ti.nɥe] 'to continue'
discontinue [dis.kɔ̃.ti.ny] 'discontinue', *discontinuer* [dis.kɔ̃.ti.nɥe] 'to continue'

(Juilland 1965)

In (5), the examples show verb inflection resulting in glide formation for all three high vowels. In section 1.2.3, we will examine other derived /HV+V/ contexts in a discussion of the variability of realisations in these contexts which is very often omitted in descriptions and theoretical accounts.

We have seen from the data in (4) and (5) that glide formation may occur when preceded by any single consonant: liquid, nasal or obstruent. According to observations from Gougenheim (1935) and Fouché (1959) glide formation is blocked when preceded by an Obstruent + Liquid (OL) sequence, resulting in dieresis with or without epenthesis as shown in (6).

(6) Blocked glide formation (dieresis): /OLHV+V/→[OLHV(G)V]

a. /j/: #OL – *plier 'to fold'*, [pli] → [pli.(j)e],
$OL – *déplier* [de.pli.je] 'to unfold'

b. /w/: #OL – *trouer 'to pierce'*, [tʀu] → [tʀu.(w)e]
$OL – *écrouer* 'to imprison', [ekʀu] → [ekʀu.(w)e]

c. /ɥ/: #OL – *fluer 'to flow'*, [fly] → [fly.(ɥ)e]
$OL – *inflouer* [ɛ̃.fly.(ɥ)e] 'to influence'

(Juilland 1965, Rey 1967, Martinet and Walter 1973)

In contrast to the data in (6) above, glide formation does occur when preceeded by an s+C sequence (7).

(7) Glide formation after a preceding s+C sequence

Preceding phonological environment

a. #(s)CG

/i/→[j]: skie [ski] 'skies', *skier* [skje] 'to ski'

b. $(s)CG

/i/→[j]: *amnistie* [am.ni.sti] 'pardons', *amnistie* [am.ni.stje] 'to pardon'

/y/→[ɥ]: *conspue* [kɔ̃.spy] 'shouts down', *conspuer* [kɔ̃.spɥe]'to shout down'

(Juilland 1965)

In view of the data in (5, 6 and 7), one might want to propose that OLG sequences are banned. Such a generalisation must be ruled out since OLG sequences do occur in French (*trois* [tʀwa] 'three', *truite* [tʀɥit] 'trout', for example, (cf. Chapter 3 and 4).

Finally, we discuss another environment for glide formation that is rarely reported: glide formation across word boundaries. Contrary to Tranel 1987 and Hannahs 1995, we observe that when a word-final high vowel meets an immediately following vowel initial word, glide formation may occur across a word boundary such that the word-final high vowel becomes a glide and forms a single syllable with the vowel following. An excellent example of across-word-boundary glide formation is attested in PFC word-list data. Number six in the PFC word-list, 6. *fou à lier*, is very often pronounced in two syllables [fwa.lje] versus a more careful pronunciation for each word resulting in three syllables: [fu.a.lje] (cf. Chapters 5 and 6)[2]. This less conventional form attests to the possibility of glide formation across word boundaries for some speakers.

Passy (1913) also reports this articulatory phenomenon resulting in diphthongs with off-glides [VG] in the spontaneous speech of a Northern variety.

---

[2] As part of the PFC project protocol, participants are asked to read, in the presence of an interviewer, from a numbered list including 94 words each time stating the number that precedes the word. Given the rather formal nature of the list task, we can expect that the word-list task elicits a more formal register from participants. We expect therefore an idealised pronunciation particularly when at the begining of the list.

> À la vérité, dans un parler rapide, il y a des cas où, deux voyelles étant juxtaposées, la moins sonore ou la moins accentuée perd son rôle syllabique, devient consonante, et forme avec l'autre une *diphtongue* ou réunion de deux voyelles en une syllabe ; ainsi dans certaines rencontres de mots comme *à outrance* (aŭtrɑ̃ːs), *il a écrit* (ilaĕkri), […] *il est si oublieux* (ilɛsiŭbliø). (Passy 1913:59)

While Passy's examples describe falling diphthongs in rapid speech, Nyrop describes spontaneous speech outcomes with on-glides [GV]: "[l]a voyelle finale d'un mot peut se changer en consonne devant la voyelle initiale du mot suivant ; dans la prononciation familière, *qui est* devient [kjæ]" (1899: 216-217, cf. § 288). Warnant observes that the context of the subject pronoun *tu* when followed by a vowel inital verb such as *est* is realised as [ɥ]; "[d]ans le parler de la conversation courante, la voyelle du pronom sujet *tu,* suivi d'une autre voyelle, se prononce [ɥ] : *tu es* [tɥɛ] *tu étais* [tɥɛ-tɛ]" (1968: XXII, note 1). We find similar phenomena attested in our findings for the spontaneous speech data of the PFC project. We will discuss cross-word-boundary data further in Chapter 6 where we examine theoretical implications as illustrated in modern treatments for glides.

In the following section, we turn the discussion to the realisation of glide epenthesis in derived contexts.

1.2.2 Glide Epenthesis

Descriptions of SF indicate that glide formation (syneresis) is more prevalent than epenthesis (dieresis) and that depending on the particular high vowel, hiatus may even be a more prevalent surface form than glide epenthesis. This observation is certainly borne out in the data presented in traditional descriptions (Léon 1978, Tranel 1987) as well as dictionaries such as Juilland (1965), Rey (1967), and Warnant (1968). Glide epenthesis in French is generally under reported and when it is reported it is done so more frequently for yod than for [w] *oué* and for [ɥ] *ué*. As we have seen for the descriptions of glide formation, the data is generally limited to the most commonly accepted forms. Glide epenthesis (like glide formation) is another strategy employed when high vowels [i y u] are in potential hiatus HV+V as shown in (1b.) reproduced here in (8).

(8) Glide Epenthesis (dieresis): /HV+V/→ [HV(G)V]

a. j: *plier* [pli] → [pli.je] 'to fold'
b. w: *trouer* [tʀu] → [tʀu.(w)e] 'to pierce'
c. ɥ: *fluer* [fly] → [fly.(ɥ)e] 'to flow'

(Rey 1963, Gaatone 1976)

Theoretical accounts (Kaye and Lowenstamm 1984, Encrevé 1988) describe glide epenthesis quite uncontroversially as the spreading of the high vowel features to a following empty onset resulting in a surface HV.GV sequence (dieresis). Glide epenthesis is most commonly described for those contexts where glide formation is blocked by a preceding OL sequence. As illustrated in (8) above, however, hiatus is not banned in French. Gaatone (1976) working in the linear rule ordered framework of early generative phonology describes epenthesis of [j] as glide insertion.

> Il est clair que l'on a affaire à un phénomène d'ordre purement phonétique, explicable en termes de conditionnement phonètique. Les distributions des deux formes sont faciles à établir: [ij] apparaît uniquement devant voyelle, [i] dans tous les autres contextes. Le yod ne faisant partie ni de la base, ni de la désinence ou du suffixe, est en somme un son de transition empêchant l'hiatus entre le *i* et la voyelle suivant et sera inséré par une règle prenant la forme en *i* comme point de départ. On peut formuler cette règle comme suit: *i* dégage yod devant voyelle.
>
> (Gaatone 1977:322-323)

Epenthesis is much less often described for [w] and [ɥ] as compared to [j]. Following from the infrequency with which they are reported we note these glides as optional using parentheses (G). Gaatone finds the under reporting unsatisfactory stating "[l]'examen du comportement des autres voyelles hautes du français, à savoir *ü* et *u*, dans des contextes identiques, fait apparaître le même types [sic] d'alternances que dans le cas de *i*" (1976: 323). In (9) below we provide examples of epenthesis for each of three high vocoids.

(9) Cross-inventory glide epenthesis (dieresis) after OL sequence

a. *trier*'to sort'*: trie* [tri], *trions* [trijõ]
b. *obstruer* 'to block'*: obstrue* [ɔbstʀy], *obstruons* [ɔbstʀyɥõ]
c. *trouer* 'to pierce' *: troue* [tʀu], *trouons* [tʀuwõ]

(Gaatone 1976: 323 adjusted IPA transcriptions)

Epenthesis is not limited to contexts where glide formation is blocked by a preceding OL sequence. Gaatone states, "[o]n peut en effet avoir deux prononciations, selon la rapidité du débit et le contexte, pour *lier* [lije], [lje], *nier* [nije],[nje], *etc.*...(1976: 327). Tranel observes that the nature of the consonant preceding an /HV+V/ context may also be a contributing factor in the application or non application of glide epenthesis. Comparing a series of derivations preceded by *r* and then *l* he states, "the second series is probably more generally accepted than the first" (Tranel 1987: 121).

(10) Glide epenthesis following simple onset (#C)
Preceding phonological environment

#CHV+V→ #CGV

a. #rj : *riant* [rijɑ̃] 'cheerful', *rieur* [rijœr] 'merry', and *rions* [rijõ] '(we) laugh'
b. #lj : *liant* [ljɑ̃] 'sociable', *lieur* [ljœr] 'binder', and *lions* [ljõ] '(we) link'

(Tranel 1987: 121)

"In sum, it seems that the consonant [l] more readily accepts gliding next to it than does the consonant [r]" (Tranel 1987: 121). According to Tranel then, when stem final [i] is preceded by [ʀ] dieresis is preferred (10a.), whereas when stem final [i] is preceded by [l] syneresis is preferred (10b.).

Klein (1991) following Gaatone (1976) states that for all hiatus in French (iV, uV and yV) the VV sequence is automatically broken by a homorganic glide ([ijV, uwV, yɥV]). Surveying various descriptions and dictionaries (Léon 1978, Juilland 1965, Rey 1967, Martinet and Walter 1973), we have no difficulty finding evidence of varying positions with respect to the prevalence of glide epenthesis in French.

Léon refers to all three French glides as "essentiellement un son de transition" (1978: 27, 30, and 32) which might be understood to mean a transitional sound between a high

vowel and a following vowel in hiatus (cf. also Gaatone 1976). However, taking 'son de transition' to refer to epenthetic glides only, we have difficulty explaining why he reports epenthesis exclusively for [j] (11).

(11) Glide Epenthesis (dieresis) after OL sequence: /OLHV+V/→ [HV(G)V]
   a. j: *crier* [kri.je] 'to cry', *plier* [pli.je] 'to fold'
   b. w: *troué* [tru.e] 'pierced', *ébloui* [eblu.i] 'dazzled'
   c. ɥ: *cruel* [kru.ɛl] 'cruel', *fluet* [fly.ɛ] 'thin', *truand* [try.ɑ̃] 'crook'

(Léon 1978: 27-32)

According to Tranel, epenthetic glide occurs for yod only and exclusively in contexts of [OLiV], "[t]here are special restrictions on the occurrence of glides after sequences composed of a stop or a fricative followed by a liquid. [...] *CLiV* are thus typically pronounced [CLijV], in two syllables ([CLi.jV]) (1987: 116). However, in identical contexts he reports no epenthetic glide for [CLu.V] nor for [CLy.V] (cf. Tranel 1987:116, Tables 7.11 and 7.12). Though preceding OL sequences may point to a potential yod epenthesis, it is not an indicator of glide epenthesis for the other two high vowels nor is it a necessary condition for glide epenthesis. Glide epenthesis may occur for any of the three high vowels when preceded by a simple [l] or [ʀ] though the likelihood for each of the three high vowels will vary.

According to Tranel, glide epenthesis depends on the quality of the high vowel in hiatus; "[i]t all depends on the nature of the vowel prevented from gliding. If the vowel is [i], then instead of [i] changing to [j], [j] is inserted between the [i] and the following vowel; if the vowel is [y] or [u], no change is made at all" (Tranel 1987: 119).

We have shown that, according to descriptions, /HV+V/ contexts may result in one of three different outcomes: hiatus [HV.V] (12abc.i); glide epenthesis [HV.GV] (12abc.ii); or glide formation [GV] (12abc.iii).

(12) Possible surface outcomes for /HV+V/ in French:

/HV+V/→ [HV.V] (hiatus), [HV.GV] (dieresis) or [GV] (syneresis)

a. *lier* ‘to tie’ :
i) /li+e/→[li.e]
ii) /li+e/→[li.je]
iii) /li+e/→ [lje]

b. *jouer* ‘to play’ :
i) /ʒu+e/→[ʒu.e]
ii) /ʒu+e/→[ʒu.we]
iii) /ʒu+e/→[ʒwe]

c. *suer* ‘to sweat’ :
i) /sy+e/→[sy.e]
ii) /sy+e/→[sy.ɥe]
iii) /sy+e/ → [sɥe]

Continuously borrowing from one another, descriptions of French contribute to the perpetuation of a descriptive tradition that privileges the notion of a homogeneous standard variety. We refer to this notion as the normative tradition in which glide formation is systematically described for all three high vowels in /HV+V/ sequences when preceded by a simple onset whereas glide epenthesis is reserved for yod (and then almost exclusively for contexts preceded by an OL sequence).

In the following sections, we examine variability in the realisation of glide formation (section 1.2.3) and glide epenthesis (section 1.2.4) in French. While many descriptions of glide phenomena in French acknowledge some variability, we show that most descriptions privilege a single form.

1.2.3 Variability of glide formation (syneresis versus dieresis)

We examine an array of phonologically derived contexts from a variety of sources expressly to examine the conventions most commonly adopted in these derived contexts. We show that even when an author recognises the potential for variability (Juilland 1965, Warnant 1987, for example) the data are still regularised. It is on this regularised version of data that descriptions will most often be centered. Nonetheless, conflicting reports

from various sources illustrate the degree to which these data can vary from prescribed norms. In this section we examine the variability described for glide realisations in derived /HV+V/ contexts.

> En effet, même si l'on écarte les prononciations régionales, provinciales, ou dialectales d'abord les prononciations régionales Parisiennes de caractère argotique ou populaire ensuite, il reste beaucoup de mots dont la prononciation soignée varie non seulement de sujet à sujet, mais aussi chez le même sujet dans des situations différentes.
>
> (Juilland 1965:VII)

We can assume the data reported in conventional dictionaries follow the most accepted norms found in traditional descriptions. Rey (1963) states "[l]es variations individuelles et l'appauvrissement du système phonétique [...] font qu'il est toujours utile de donner une norme" (Rey 1963: XIII). According to Warnant (1968) forms involving [i], [u] and [y] in hiatus generally show syneresis but for these words he offers that "[a]près un mot qui se termine par une consonne, il y a parfois diérèse (*par liaison* paʀ-li-ɛ-zɔ̃, *une liaison* yn-li-ɛ-zɔ̃" (1987: 237 cf. *liaison*).

Tranel proposes that a preceding consonant can influence the realisation of glide epenthesis.

> [I]t is worth noting that the vowel [i] of the verb stem *rire* 'to laugh' relatively rarely turns into the corresponding glide [j] when followed by a vowel-initial ending. Thus for *riant* 'cheerful', *rieur* 'merry', and *rions* '(we) laugh', for instance, the pronunciations [rijɑ̃], [rijœr], and [rijõ] are much more common than [rjɑ̃], [rjœr], and [rjõ]; that is, while no speaker will find the first set impossible, many speakers will find the second odd. [...] If we now consider the case of the verb stem for *lier* 'to link', we also find two similar patterns of pronunciation for words like *liant* 'sociable', *lieur* 'binder', *lions* '(we) link': (i) [lijɑ̃], [lijœr], and [lijõ], and (ii) [ljɑ̃], [ljœr], and [ljõ]; except that this time, it seems that the consonant [l] more readily accepts gliding next to it than does the consonant [r].
>
> (Tranel 1987: 121)

Based on Tranel's position, the implication is that there is a constraint against glide formation (syneresis) following an /ʀ/ but not following /l/. It is interesting to note that Morin also observes a disparity in the applicability of 'relaxation' for /Olj/ versus /Orj/ (cf. 1976).

(13) Variable Gliding

#LV+V→ #LGV

i) *lier* 'to tie': [ljɑ̃], [ljœr], and [ljõ] syneresis

ii) *rire* 'to laugh' ?[rjɑ̃], ?[rjœr], and ?[rjõ]

#RV+V→ #RV.GV

i) *rire* 'to laugh' [rijɑ̃], [rijœr], and [rijõ] dieresis

ii) *lier* 'to tie': ?[lijɑ̃], ?[lijœr], and ?[lijõ]

(Tranel 1987: 121)

Although syneresis resulting from glide formation is reported quite systematically for any /HV+V/ preceded by any single consonant, when allowing for stylisitic, or individual variation, any one of the three possible outcomes in (syneresis, dieresis with or without epenthesis) may result (13). Juilland (1965) provides some perspective on the considerations that must be made in planning a dictionary.

> Dans tous ces cas, il nous a fallu ou bien choisir une prononciation "préférée" parmi celles que les dictionnaires et manuels considèrent acceptables, ou bien attribuer à chaque mot autant de transcriptions, donc de rubriques dans le dictionnaire, que de prononciations admises par les autorités. [...] il suffira de dire que nous avons préféré la prononciation naturelle à la prononciation artificielle, la prononciation usuelle ou courante à la prononciation littéraire, la prononciation familière à la prononciation formelle.
>
> (Juilland 1965: IX)

Other dictionaries must of course make similar considerations but often it seems they prefer to align pronunciations with prescriptive norms. Consequently, we find a much greater degree of variability in Juilland's dictionary. Let us look now at the /HV+V/ data from Juilland (1965) for one of the more productive suffixes of French (other than first conjugation verb morphology) –age, as shown in (14).

(14) Nominal suffix *–age*

a. /i + age/

#Ci+age→ #CVV

liage [li.aʒ] ‘linkage’

$Ci+age→ #CGV

alliage [a.lja ʒ] ‘alloy’

reliage [rə.ljaʒ] ‘hooping’

maniage [man.jaʒ] ‘handling’

#OLi+age→ #OLVV

pliage [pli.aʒ] ‘folding’

criage [kri.aʒ] ‘public announcing’

triage [tri.aʒ] ‘sorting’

b. /u + age/

#Cu+age→ #CVV

nouage [nu.aʒ] ‘knotting’

rouage [ru.aʒ] ‘cog’

#Cu+age→ #CGV

louage [lwaʒ] ‘renting’

touage [twaʒ] ‘towing’

$Cu+age→ $CVV

ébouage [e.bu.aʒ] ‘garbage removal’

affouage [a.fu.aʒ] ‘wood cutting rights’

$Cu+age→ $CGV

échouage [e.ʃwaʒ] ‘beaching’

tatouage [ta.twaʒ] ‘tatooing’

#OLu+age→OLVV

clouage [klu.aʒ] ‘nailing’

renflouage [rɑ̃.flu.aʒ] ‘a bail-out’

c. /y + age/

#Cy+age→#CGV

nuage [nɥaʒ] ‘cloud’

suage [nɥaʒ] ‘condensation’

$Cy+age→$CGV

écobuage [e.kɔ.bɥaʒ] ‘clearing’

cocuage [kɔ.kɥaʒ] ‘cuckoldry’

remuage [rə.mɥaʒ] ‘stirring/shaking’

ressuage [rə.sɥaʒ] ‘flushing’

$OL

engluage [ɑ̃.glɥaʒ] ‘sticking’

(Juilland 1965: 425-426)

The suffix *–age* provides a variety of data wherein all three of the high vowels occur in the target context. Not only do these data show the full range of possible phonological contexts preceding a gliding environment, but they also show an array of potential outcomes (with the exception of epenthesis). In (14a.), for example, for the same preceding consonant *l-* we observe syneresis (#C) and dieresis ($C) while the blocking effect for a preceding OL sequence yields categorical hiatus. With minor differences the same is true for (14b.), where, for a single preceding liquid or obstruant, we observe both syneresis ([u] →[w]/[t, l] __+age) and dieresis ([u]/[n, r] __+age). Dieresis is, otherwise, most prevalent in (14ab.) whereas syneresis is categorical in (14c.) even for /OLHV+V/ contexts. None of the data in (14) shows the level of homogeneity that has been described

for the verb inflection data in section 1.2.1. Instead, we observe a high degree of variability not only across the three high vowels but also for a single preceding consonant (cf. -*l*- in 14a.) or a single class of consonants (liquids and obstruants in (14b.)). Glide data in Juilland (1965) do not necessarily follow the conventional norms observed in other dictionaries (Rey 1967, Warnant 1965), but they are based on attested data (cf. Martinet and Walter 1973, Martinet 1945).

> En déterminant les prononciations acceptables, nous nous sommes appuyés sur les ouvrages de Passy, Martinon, Fouché et Delattre. En dernière instance, nous avons généralement adopté la solution suggérée par le verdict statistique inhérent aux résultats de l'enquête entreprise par André Martinet dans *la prononciation du français Contemporain.*
>
> (Juilland 1965: IX-X, original emphasis)

Linguistic data from dictionairies offer insight into the accepted norms and descriptions of the day. Like Juilland (1965), Rey (1963) and Warnant (1968) acknowledge the influence of these same sources in choosing the pronunciations to cite for a given word. Most striking about the choice of data given in Juilland is the prevalence of glide formation for the high vowel /y/ in hiatus particularly when glide formation is regularly reported for /y/ in hiatus even when preceded by an OL sequence which is said to block glide formation. Take, for example, the unexpected forms such as *engluage* [ɑ̃.glɥaʒ] (14c.).

In fact, we find that in Juilland glide formation (syneresis) is quite prevalent for derivations involving the suffix –*ant*; variability is observed across high vowels exclusively for [y] in the context of a preceding OL sequence (15).

(15) Adjectival suffix –*ant* (categorical syneresis)

/OLy+ã/→[OLɥã]

a. #OL *gluant* [glɥã] 'sticky',
bruant [brɥã] 'sparrow',

b. $OL *affluent* [aflɥã] 'tributary',
effluent [eflɥã] 'stream',
influent [ɛ̃flɥã] 'worn',

(Juilland 1965)

Given the apparent blocking effect of a preceding OL sequence that was discussed earlier (Gougenheim 1935, Fouché 1959 and Tranel 1987), most descriptions and standard dictionaries do not cite glide formation as a possibility in this context. According to Klein however some idiolectal variation may occur particularly in the cases of /YI/ sequences which he refers to as pseudo-morphology: *altruisme* [altʀyism] or [altʀɥism]; *incongruité* [ɛ̃kɔ̃gryite] or [ɛ̃kɔ̃grɥite], each containing a suffixal boundary (1991:47).

While discussions of variation are present in traditional descriptive literatures, these discussions generally center on regional varieties. Table 1 below illustrates variation across regional varieties.

Table 1: Pan dialectal variation in glide formation

Variety	*nier*	*nouer*	*nuer*	
a. Parisian French	[nje]	[nwe]	[nɥe]	syneresis
b. Northern French	[nje]	[nwe]	[ny.e]	dieresis
c. Midi French	[nje]	[nu.e]	[ny.e]	
d. Media French	[ni.e]	[nu.e]	[ny.e]	

Heap and Kelly 2005, based on Bullock 2002:20 after Tranel 1987:122 and Lyche and Girard 1995.

Tranel's original figure illustrates propensity to glide across the inventory of 'closed' (i.e. high) vowels. Bullock adapted the original in order to illustrate that propensity of glide formation can been seen to have regional implications. Heap and Kelly (2005) add Media French based on Lyche and Girard (1995) to illustrate that variation can be represented as a continuum with categorical syneresis and categorical dieresis at the two extremes.

Table 1 shows that Parisian French, the defacto standard for descriptions (Table 1a.), purports to be homogenous. HV+V sequences are resolved through glide formation resulting in uniform syneresis. At the other extreme is Media French (Table 1d.) showing across-the-board hiatus (dieresis).

Data from Martinet and Walter (1973) illustrate a slightly different picture. In an empirical study of cultivated speakers from Paris, Martinet and Walter show that there is much more variability between glide formation (syneresis) and hiatus (dieresis) than is generally described. Martinet and Walter ellicit the pronunciation of target words presented to participants in carrier phrases. According to the authors: "[o]n n'a pas hésité

à donner fréquemment à ces énoncés un tour résolument badin qui s'est révélé très efficace pour atténuer la monotonie des enregistrements." (Martinet and Walter 1973: 25). The data in (16) show a variety of derived contexts for each of the three high vowels.

(16) Variability in Parisian French (Martinet and Walter 1973)

a. i~j
   i) SCIER *v.* 'to saw' [sje] 15/17 syneresis, [sije] 2/17 dieresis with epenthesis
   ii) SCIEUSE *nf.* 'sawyer' [sjøz] 14/17 syneresis, [sijøz] 3/17 dieresis with epenthesis
   iii) SCIEUR *nm.*'sawyer' [sjœr] 16/17 syneresis, [sijœr] 1/17 dieresis with epenthesis
   iv) SCIAGE [sjaʒ] 'sawing' 15/17 syneresis, [sijaʒ] 2/17 dieresis with epenthesis
   v) SCIANT [sjɑ̃] 'sawing' 15/17 syneresis, [sijɑ̃] 2/17 dieresis with epenthesis

b. u~w
   i) LOUER *v.* 'to rent' [lwe] 16/17 syneresis, [lue] 1/17 dieresis (hiatus)
   ii) LOUEUR *nm.* 'lessor' [lwœr] 7/17 syneresis, [lu(w)œr] 10/17 dieresis (hiatus)
   iii) LOUEUSE *nf.* 'renting agency' [lwøz] 8/17 syneresis , [lu(w)øz] 9/17 dieresis (hiatus)
   iv) JOUER *v.* 'player' [ʒwe] 16/17 syneresis, [ʒue] 1/17 dieresis (hiatus)
   v) JOUABLE *adj.* 'playable' [ʒwabl] 14/17 syneresis, [ʒuabl] 3/17 dieresis (hiatus)
   vi) JOUAILLER *v.* 'bit player' [ʒwa/ɑje] 12/17 syneresis, [ʒua/ɑje] 5/17 dieresis (hiatus)

c. y~ɥ
   i) INFATUÉ *adj.*'infatuated' [ɛ̃fatɥe] 16/17 syneresis, [ɛ̃fatye] 1/17 dieresis (hiatus)
   ii) TUAGE *nm.* 'killing' [tɥaʒ] 6/17 syneresis, [tyaʒ] 11/17 dieresis (hiatus)
   iii) TUER *v.* [tɥe] 'to kill' 11/17 syneresis, [tye] 6/17 dieresis (hiatus)
   iv) TUANT *adj.* 'killing' [tɥɑ̃] 10/17 syneresis, [tyɑ̃] 7/17 dieresis (hiatus)
   v) BUÉE *nf.* [bɥe] 'foggy'11/17 syneresis, [bye] 6/17 dieresis (hiatus)
   vi) PUANT *adj.* 'smelly' [pɥɑ̃] 3/17 syneresis, [pyɑ̃] 14/17 dieresis (hiatus)
   vii) PUER *v.* 'to smell' [pɥe] 13/17 syneresis, [pye] 4/17 dieresis (hiatus)

(Martinet and Walter 1973) [3]

[3] We have replaced identification of the participants as given in the original with ratios of the number of speakers (N) out of 17 speakers overall (syneresis N/17 participants and dieresis N/17).

The first and perhaps most important observation is that, for 17 educated speakers of the Parisian norm, we find no homogeneity, though preferences clearly emerge in these data. Syneresis is by far the most prevalent outcome and where syneresis is not realised, outcomes depend on the particular high vocoid. While we observe dieresis with epenthesis for [i] (16a.), tokens involving [u] and [y] show dieresis with hiatus only (16cd.). No one glide, nor any specific context, shows categorical results. Most consistant are the data in (16a.) where tokens involving [i] strongly privilege syneresis showing, at most, three tokens for dieresis with epenthesis. In (16bc.) however, these nearly identical environments show variabity to extremes; in (16bc.i), we observe that speakers privilege syneresis (16:1) then in (16bc.ii) results are flipped with as much as 60% favouring dieresis.

Looking at the data in (16bc.), we are struck by the degree to which results vary for the same high vowel in an /HV+V/ environment. All that distinguishes these various examples is the suffix, which raises the question of the significance of lexical frequency and/or vowel quality for the second vowel segment ([e] versus [ø, œ] or [a, ɑ̃], for example). First conjugation *–er* forms, are predominantly realised with syneresis while other suffixes: *–eur* and *–euese*, show slightly more dieresis than syneresis. In (16c.) *–age* (syneresis 6 dieresis 11) and *–ant* (syneresis 4 dieresis 13) show exactly the inverse outcomes as compared to the verbs *infatué:* syneresis 11 dieresis 6 (16c.i)*, puer*: syneresis 13 dieresis 4 (16c.vii). The degree of variability attested here is particularly significant given that the sample studied represents the so-called standard variety for which this kind of variability is rarely described.

These data and all of the variability described support the observation that "[e]verything else being equal, of the three closed vowels of French, [i] is the one that glides the most generally, next comes [u], and last, [y]" (Tranel 1987:122). This assymetrical application of glide formation (or glide epethesis) suggests an implicational hierarchy amongst glides in French. The implication is such that for any glide phenomenon the presence of [ɥ] implies [w] and [j], whereas the presence of [w] implies [j] only. Tranel (1987) states it this way "[...] [y] –gliding implies [i]-gliding and [u]-gliding, and [u]-gliding implies [i] gliding, but not vice versa" (1987:122). Bullock (2002) based on Tranel proposes the

following markedness hierarchy: j » w » ɥ (i.e. [j] is less marked than [w], which in turn is less marked than [ɥ]. We will come back to this in our discussion of markedness and Bullock's proposals for an OT account of French glides in Chapter 4.

We turn now to variability in glide epenthesis contexts. The variable realisation of glide epenthesis (as we saw for glide formation data) can be related to the same implicational hierarchy.

1.2.4 Variability of glide epenthesis (dieresis)

Variability of glide epenthesis is most prevalent when an /HV+V/ environment is preceded by an OL sequence. It is observed that a preceding OL sequence (unlike a preceding sC cf. *skier*) blocks glide formation, thus providing an ideal context for glide epenthesis and by extension to observe the possibility of variable glide epenthesis. Martinet and Walter (1973) show that where high vowels are found in hiatus, yod epenthesis is indeed most prevalent (17a.) and, though marginally attested for [w] (17b.i. ii) glide epenthesis is negligible for [ɥ]. In (17) we show results from Martinet and Walter (1973) for various contexts of OL sequences preceding a /HV+V/ context.

(17) Variability in OLHVV (hiatus versus epenthesis)

a. iV~ijV:
 i) PLIER *v.* 'to fold': [pli.je] 16/17 epenthesis; [pli.e] 1/17 hiatus
 ii) PLIAGE *nm.* 'folding': [pli.jaʒ] 16/17 epenthesis; [pli.aʒ] 1/17 hiatus
 iii) PLIEUSE *nf.*'folder' ': [pli.jøz] 16/17 epenthesis; [pli.øz] 1/17 hiatus
 iv) PLIOCÈNE *adj.* 'Pliocene' : [pli.jɔsɛn] 15/17 epenthesis; [pli.ɔsɛn] 2/17 hiatus

b. uV~uwV:
 i) TROUER *v.* 'to pierce': [tʀu.e] 17/17 hiatus
 ii) TROUÉ(E) *adj.* 'pierced': [tʀu.e] 16/17 hiatus; [tʀu.we] 1/17 epenthesis
 iii) TROUÉE *nf.* 'holed': [tʀu.e] 15/17 hiatus; [tʀu.we] 2/17 epenthesis

c. ɥ:
 i) FLUER *v.* ' to creep': [fly.e] 17/17 hiatus
 ii) REFLUER *v.* 'to reflux': [fly.e] 16/17 hiatus; [f(l)ɥe] 1/17 syneresis with OL simpification
 iii) FLUET *adj.* 'frail': [fly.ɛ/e] 16/17 hiatus; [fly.ɥe] 1/17 epenthesis

(Martinet and Walter 1973, syllable breaks '.' added)

Similar to patterns observed in (16 above), epenthesis after an OL sequence is most prevalent for yod. As shown by these data, epenthetic [w] and [ɥ] are much less frequently attested than [j]. The labial velar [w] is observed three times (17b.ii) *troué(e)* [tʀu.we] and twice in (17b.iii) *trouée* [tʀu.we], while we find no [ɥ] epenthesis in these data. A single participant in this study stands out as the consistant outlier with respect to glide epenthesis; tokens involving the high vowel [i] show that a single speaker gives a realisation with hiatus for all /HV+V/ contexts with OL preceding. Interestingly, results in (17b.i.) for the verb form *trouer* show categorical hiatus, however, adjectival and noun forms show realisations with epenthesis (17c.ii) and (17c.iii.). In (17c.ii) we find a single token for glide formation resulting in [ɥ] preceded by OL. Note that the token is realised with simplification of the OL preceding. We will discuss more on OL simplification in our discussion of dialect data in Morin (1976) in section 1.3.

Other empirical evidence from Martinet and Walter (1973) corroborates this finding (18).

(18) OLy$V~OLɥV:

a. TONITRUANTE 'thunderous' *adj.*:[tɔ.ni.try.ɑ̃t] 15/17; *[tɔ.ni.trɥɑ̃t] 2/17
b. TONITRUER *v.* 'to thunder': [tɔ.ni.try.e] 14/17; *[tɔ.ni.trɥe] 3/17

(Martinet and Walter 1973, syllable breaks '.' added)

In (18) we observe two examples of an /HV+V/ context with OL preceding realised consistently with syneresis by two of these 17 speakers (note that for (18b.) syneresis is given by three speakers). Anomalous, though it may be, this evidence combined with the prevalence described by Juilland (1987) does suggest more irregularity of glide formation data for this environment than is most often described.

Juilland's transcriptions for the adjective formed with *–ant* (see 16 above) contradict what is traditionally described for preceding OL contexts, reporting categorical glide formation after an OL sequence but exclusively for the front rounded high /y-ɥ/ vowel~glide alternation. This is particularly interesting since the front rounded /y/ is traditionally reported to be the high vowel the least prone to glide formation in French (cf. Table 1 above, also Tranel 1987, cf. also also Bullock 2002 and our discussion of markedness in Chapter 4).

(19) Glide formation after a preceding s+OL sequence

a. #sOLG: /y/→[ɥ]: *obstrue* [ɔb.stry] ‘obstructs’, *obstruer* [ɔb.strɥe] ‘to obstruct’

b. $sCG: /y/→[ɥ]: *désobstrue* [dez.ɔb.stry] ‘unobstructs’
désobstruer [déz.ɔb.strɥe] ‘to unobstruct’

(Juilland 1965)

Throughout Juilland’s dictionary the high vowel [y] in a /HV+V/ environment consistently results in glide formation (19 above and 20 below). Data from various sources in (20) illustrate that pronunciation for these contexts are not regularised across sources.

(20) Varying reports of pronunciation of OLHV+V

Warnant (1968)

a. Hiatus (exclusively) cf. *prier* ‘to pray’ [pʀi-e], *clouer* [klu-e] ‘to nail’, *cruel* [kʀy-ɛl] ‘cruel’

Rey (1967), Tranel (1987), Centre National de ressources Textuelles et Lexicales (CNRTL on-line)

b. [j] epenthesis only cf. *prier* ‘to pray’ [pʀije], *trouer* [tʀue] ‘to puncture’, *cruel* [kʀyɛl] ‘cruel’

Juilland (1965)

c. Glide formation cf. *désapproprier* [dezapʀɔpʀje] ‘to dispossess’, *trouer* [tʀue] ‘to puncture’ but *confluer* [kɔ̃flye] ‘to flow together’ and *influer* [ɛ̃flye] ‘to influence’

We would suggest then that (20b.) represents a convention for the most commonly accepted ‘standard’ pronunciations. We have shown that dictionary sources must decide which pronunciation(s) they wish to list and, like descriptions of French, are generally biaised towards the prescribed norms. With respect to glide epenthesis Warnant (1968) notes this very issue in his preface:

> Lorsque deux voyelles dont la première est un [i] sont en contact, comme dans *prier*, deux prononciations du mot sont possibles, soit [pʀi-e], soit [pʀi-je], avec un léger [j] intervocalique. Nous avons seulement noté la première prononciation [...].
>
> (1968: XVI)

Most descriptions and dictionaries report prononciation with hiatus when glide formation is not possible. Epenthetic yod is reported most commonly for /OLi/ after which we may observe some [w] and still fewer [ɥ] (20b.). The online resource CNTRL notes this is a more recent convention:

> **PRONONC. ET ORTH. − 1. Forme phon. :** [buklie] ou [-klije]. Le mot est transcrit sans yod dans Fér. 1768, Fér. *Crit.* t. 1 1787, Land. 1834, Gattel 1841, Nod. 1844, Fél. 1851, Littré et Warn. 1968. Il est transcrit avec yod dans *DG,* Dub., *Pt* Rob. et *Pt Lar.* 1968.
>
> (Centre National de ressources Textuelles et Lexicales (CNRTL) online)

Morin reports asymmetry for glide epenthesis among the three glides, which he attributes to dialectal variation:

> [i]n some dialects, only *i* is followed by a transitional glide, viz. yod, e.g. in Standard French […]. In some other dialects, only *i* and *u* are followed by transitional glides, e.g. Belgian French. In still others, all three high vowels are followed by transitional glides, e.g. Québec French.
>
> (Morin 1976: 39)

Gaatone, in regards to under reporting of epenthetic [w] and [ɥ], notes:

> Sur ce point, la description des données présentées ici est en désaccord avec les autres descriptions existantes des mêmes faits. En effet, toutes les transcriptions de ces formes ou de formes analogues, proposées soit dans les dictionnaires soit dans les traités ou études de phonétique, notent assez régulièrement le yod de transition, dans *crier* par exemple, mais jamais le [ẅ] de *obstruer* ou le [w] de *trouer*.
>
> (Gaatone 1976: 323)

Gaatone speculates that though these glides are unmistakably present he would attribute their absence from descriptions to the general imperceptibility of [ɥ] and [w] as compares to yod (cf. section 5.3 Judgement Task results).

> Cependant l'observation attentive de la prononciation de mots de cette espèce par des informants francophones ne laisse aucun doute sur la présence d'une semi-voyelle de transition entre les deux voyelles. Il est vrai que la perceptibilité, sur le plan acoustique, de [ẅ] et [w], est moindre que celle du yod, ce qui peut sans doute expliquer leur omission dans les transcriptions phonétiques. Mais elles n'en sont pas moins présentes et, comme yod, assurent la transition articulatoire entre une voyelle haute et une voyelle suivante.
>
> (Gaatone 1976: 323)

1.2.5 Derived glide contexts

In (21) below, we summarise the major derived glides contexts which will form the basis of our analysis of derived glides contexts in spontaneous speech (cf. Chapter 6). The process by which they result allows us to distinguish between two different contexts for derived glides: glide formation (21a.) and glide epenthesis (21b.).

(21) French glides

Derived Glides: /HV+V/→[GV] or [HVGV]

a. Glide formation (high vowel~glide alternations): /HV+V/→[GV]
   i. [i]→ [j]: *lier* /li+e/→ [lje] 'to tie'
   ii. [u]→ [w]: *jouer* /ʒu + e/ → [ʒwe] 'to play'
   iii. [y]→ [ɥ]: *suer* /sy + e/ → [sɥe] 'to sweat'

b. Epenthetic glide (high vowel spreading): /OLHV+V/ → [OLHV(G)V]
   i. [i] → [i.j]: *plier* /pli+e/ → [pli.je] 'to fold'
   ii. [u] → [u.w]: *trouer* /tʀu+e/ → [tʀu.(w)e] 'to pierce'
   iii. [y] → [y.ɥ]: *fluer* /fly + e/ → [fly.e], [fly.(ɥ)e] 'to flow'

(Rey 1963, Juilland 1965)

In (21) above we distinguish two derived glide contexts: glide formation (21a) and epenthetic glide (21b.). Glide formation (21a.), is most often a result of derivation or inflection, which occurs when a vowel initial suffix is added to the high vowel-final root morpheme. Glide formation results in syneresis (the glide and the following vowel remain tautosyllabic) while epenthesis in (21b.) results in dieresis (the high vowel remains in the nucleus but is then followed by a heterosyllabic glide). Epenthetic glides are most commonly attested for the high vowel [i] in hiatus particularly, though not

exclusively, where glide formation is blocked (30b.). Like glide formation, glide epenthesis occurs as a possible hiatus breaking strategy (*lier* [li.je], *jouer* [ʒu.we], for example) but since hiatus and glide formation are attested along with epenthesis in identical environments this is but one of three possible outcomes. Glide formation and glide epenthesis are related, in so far as they are both hiatus breaking strategies, thus they are inherently linked to the presence of high vowels yet they are distinct from one another. Glide formation and glide epenthesis are cross-linguistic strategies observed in the grammar to avoid a potential surface hiatus (cf. Rosenthall 1997, and our discussion Chapter 4.6.1).

We turn next to a discussion of lexicalised glides in Continental French. We will show that lexicalised glides though generally described as invariable (except for very particular instances of a performance style such as when reciting verse) may, in fact, show more variability than is described. In section 1.3 we will examine the historical origins of the lexicalised glides to show how various historical processes have influenced the distributional facts for glides in Modern French. In section 1.4, we provide examples for all lexicalised and glide contexts discussed in the literature. In section 1.5 we will discuss variablity that may be observed in realisations of the lexicalised glides.

1.3 Lexicalised glides in French

By lexicalised, we understand these forms to be present in the lexicon in the same phonetic form as can be observed in surface realisations. Lexicalised glides are therefore distinguished from their derived counterparts since their presence does not depend on a particular conditioning environment to trigger a glide realisation as is described for derived glides. Still, for some speakers, at least, high vowels are recoverable from glides in lexicalised contexts (Martinet 1945). Apart from spelling conventions that may aid in this connection, the presence of lexicalised glides is not based (synchronically at least) on the presence of a corresponding high vowel in hiatus (though diachronically that may have been the case for some of these glides). Links between a glide and a corresponding high vowel are powerful enough that for Saussure (1914; 1982) and Martinet (1945) the

phones [i] and [j] are but two realisations of a single phoneme /I/. Early generative models follow suit.

A cursory examination of these data shows that yod [j] is most prevalent, enjoys the widest distribution (occuring in initial, medial, and final position in a word) and is associated with the most varied spelling conventions. When compared to yod, the glides *oué* [w] and *ué* [ɥ] are much less prevalent (though instances of oué far outnumber those of ué), have a defective distribution (limited to initial and medial position), and show a more limited inventory of spelling conventions (hence the names *oué* and *ué)*. Most commonly seen as part of a true or nuclear diphthong marked by the orthographic convention -*ou*- (*ouest* [wɛst] 'west' or *ouate* [wat] 'cotton batting', *moi* [mwa] 'me' and *mouette* [mwɛt] 'seagull', for example), [w] is also found in absolute initial position in borrowings where it is identified by the orthographic convention *w*- (*watt*, *Wallon*, *week-end*). Similarly [ɥ] occurs in initial and medial positions as part of a true diphthong[4] [ɥi] (*huile, huître, truitre* [tʀɥit], *fuite* [fɥit] 'leak' for example) or occasionally occurs in intial position as *hu*- (*huis* [ɥi] 'door'*, huit* [ɥit] 'eight', *huée* [ɥé] 'boo', *huître* [ɥitʀ] 'oyster').

These categorisations could be simpified further to reflect common historical origins. Glides in absolute initial position (i.e. non nuclear) are largely the result of late borrowings, for example. The Latin words *hiatus* (Latin *hiare* dates to late 17th century) and *hier* (dates to 12th century originally *ier*, *h*- is added later) are late entries to the language and therefore show adaptation to the French spelling conventions rather than evidence of the classic evolution for Latin [i] initial words (cf. *iunctio→jonction*, for example).

In the following sections, we will illustrate how the various diachronic processes contribute to our Modern French three-glide inventory. In section 1.3.2, we examine processes that have contributed to the prevalence of lexicalised yod in Modern French. In

---

[4] According to Lyche and Durand 1999, these rising diphthongs are described as heavy nuclei with strong cohesion (les diphtongues à forte cohésion du français cf. Kaye et Lowenstamm 1984, Klein 1991) whereas word-final Vj is a falling light diphthong.

section 1.3.3, we are concerned with those processes that contribute to the Modern French occurrences of lexicalised *oué* [w]. In section 1.3.4 we will show that Modern French lexicalised *ué* [ɥ] occurs late in the evolution of French from Latin (only after the emergence of the corresponding high vowel [y]) and in much more constrained environments. Following the discussion of historical factors in the distribution of glides in Modern French, we discuss some aspects, both diachronic as well as synchronic, of the variability of lexicalised glides in section 1.4.

1.3.1 Historical influences in the prevalence of lexcalised glides

Although Classic Latin does have [j] and [w], none of these persist as glides in Modern French (Carton 1974:162). According to Nyrop:

> On avait en latin la médio-palatale ordinaire [j] qui s'entendait dans jocus jam, junius, etc.; le même son remplaçait dans la langue vulgaire un *i* (*e*) en hiatus: rationem>ratjone, cavea>cavja (§ 262,3). Le yod latin a subi de nombreuses modifications: jam >[ža], ratjone >[ræzɔ̃], etc.
>
> (Nyrop 1899: 354)

Latin [w] also continues to evolve to its modern realisation as the voiced frivative [v] (Fouché 1966: 621). As we have seen for French, Latin shows a similar tendency to avoid hiatus. Glide formation in hiatus is well attested for yod which later results in palatalised consonants. Ué [ɥ] did not exist in Latin but is rather a French innovation in the 8th century. According to Carton, ué [ɥ] "provient de ŏ latin combiné avec un [j] qui le suit, ou d'un [y] français devant voyelle, le yod influençant tout l'entourage" (1974: 163).

The prevalence of lexicalised yod as compared to lexicalised oué [w] and ué [ɥ] in Modern French can be explained, at least partially, by the various origins in processes that occur at different stages in the evolution of French from Latin. We observe that yod as compared with *oué* and *ué* is involed in considerably more contexts. In the following sections, we will illustrate the various evolutionary processes and their contribution to the distinct distributions of the three-glide inventory /j ɥ w/ of Modern French. Lexicalised

glides like their derived counterparts are almost exclusively related to the presence of high vowels in a HVV sequence that is later resolved through the consonantisation of the high vowel segment. Lexicalised glides do not however behave as their derived counterparts do.

1.3.2 Historical origins of Modern French yod [j]

Carton (1974) surveying Fouché (1966), Straka (1964), (1965) and Bourciez (1967) discusses the various processes of palatalisation of Latin consonants often, though not always resulting in yod. We include only some of these processes here to illustrate the prevalence of yod in the evolution of Latin to Modern French. Many outcomes will continue to evolve and are indeed present in Modern French (nj > ɲ *montanea > montagne; fingente > feignant,* for example), while others will disappear before Modern French (tj > t' > ts' jts (intervocalic -jdz- rationem > raison, kj > k̯ > t̯ avancé > ts' > ts > s facia > face, for example). According to Nyrop, "J peut se fondre avec la consonne précédente (suivante) en un son mouillé; ce développement a lieu dans les groupes nj, lj, rj, sj, tj, ssj, stj, qui se changent en [ñ], [ɫ], [r'], [z'], [s']", (1899: 355). Various examples of the processes described by Carton are shown in (21):

(22) Processes resulting in yod (remains in Modern French)

a. lj > λ > j in palea > [paλ] > [paj], *paille* 'straw'*, travail* 'work' [travaj]
b. kl, gl > λ > j in solic(ŭ)lu > *soleil* [sɔ'lɛλ], [sɔ'lɛj] 'sun'
c. –dj > jj > j dans radiu(s) > rai 'ray'
d. –gj > jj > j as in exagiu(m) > *essai* 'attempt'
e. Intervocalic *k* and *g*: a_a > j [5]
  i) pacáre > paiier, *payer* 'to pay'
  ii) nēcā́re > lat. vulg. *negáre > neier, *noyer* 'to drown'

(Carton 1974:157-162 glosses added)

---

[5] Carton observes that the obstruents /k g/ are vulnerable by virtue of their intervocalic position; thus are subject to articulatory weakening (1974: 162).

Among the various instances of palatalisation, Carton lists: true palatalisations (21ab.), false palatalisation, *renforcement du yod* according to Carton (21cd.), and *résolution en yod* ou *relâchement en yod,* the case of lenition of [k] and [g] resulting in yod (21e.). Intervocalic [g], for example, when followed by front vowels [i] or [e] is weakened (se mouille/palatalises) under the influence of the front vowels [i] or [e] immediately following: g + i, e > j pagense finishes in hiatus from anc. fr. *païs,* mod. Fr. *pays* and *païen*. Both velar consonants [g] and [k] are subject to this inertvocalic weakening (intervocalic [k] undergoes sonorisation [k] > [g] before further leniting to [j]). As shown in (21), various processes can account for much of the lexicalised yod data observed in Modern French.

Nyrop describes later developments in the evolution of the French language that also contribute significantly to the prevalence of lexicalised yod in Modern French:

> Cependant un nouveau yod s'est développé en français, où ce phonème est maintenant d'emploi fréquent ; il provient surtout d'un *i* en diphtongue ou en hiatus, et d'un [λ] réduit (§ 351): *pied* [pje], *yeux* [jø], *viande* [vjɑ̃:d], *payer* [pɛje], *fille* [fi:j], etc.
>
> Nyrop (1899: 420)

Many Modern French instances of lexicalised yod are the result of the process of Romance diphthongisation (22).

(23) Romance diphthongisation ę > ęę > ẹę > íę

a. ĕ (ouverts) tonic and free (open syllable)
péde > pié > *pied* 'foot'

b. ĕ tonic and closed
tĕrtiu > *tier(s)* 'third(s), *pĕttia > *pièce* 'room', nĕptia > *nièce* 'niece', fĕl > *fiel 'bile'*, mĕl > *miel* 'honey', rĕm > *rien* 'nothing'

(Carton 1974:139 glosses added)

In (23) we show instances of glides issued from yod in hiatus as is mentioned by Carton.

(24) Yod resulting from hiatus
violon [vjolɔ̃] 'violin', piano [pjano] 'piano', lierre [ljɛːʀ] 'ivy'

(Carton 1974:139 glosses added)

In Modern French, we find the stem changing verbs *venir* → *vient*, and *tenir* → *tient* that appear to be vestiges of the historical changes (Romance diphthongisation). These verbs show alternating stem forms as indicated by their alternating orthography -*e*-/-*ie*-. The alternation depends on the grammatical person evident in the forms in *–ie-* whereas the infinitive is unchanged (*venir* [vəniʀ] 'to come' → *vi<u>e</u>nt* [vjɛ̃] 'comes' → *v<u>e</u>nez* [vəne] 'you come' pl.). We take these GV sequences to be lexicalised diphthongs, the result of Romance diphthongisation of the Latin tonic short ĕ.

> Cette alternance continue à fonctionner en moyen français, dans la plupart des cas : nous avons cependant noté quelques cas de réduction de la diphthongue, mais ce n'est jamais systématique […] Pour bien d'autres verbes, l'alternance continue à fonctionner parfaitement : *venir*, *tenir*, *quérir* et leur composés.
> (Marchello-Nizia 1997: 275)

The alternating forms *venir* and *tenir* provide evidence of tonic accent conditioning in these historical processes.

Thus far, we have only discussed the various processes acting upon words but many French suffixes originate in Latin, and are therefore subject to some of these very same kinds of articulatory pressures. One of the most prevalent of these (24), the French suffix *–ier*, from the Latin suffix *-arium* the accusative of –ārius gives several outcomes according to Lanly (1971).

(25) The suffix *–ārĭu*, *–ārĭa* > –ier, –ière
a.*–aria* > *aire* : parĕa > *paire* 'pair'
b.*–aria* > *ère* : brucā́rĭa > *bruyère* 'heather'
c.*–aria* > *ière* : primā́rĭa > *première* 'first'
(Lanly 1971:48-51 glosses added)

We turn now to the origins of Modern French oué [w] and ué [ɥ]. Whereas the Modern French yod results from some forms of palatalisation, diphthongisation and hiatus resolution, the French oué [w] and ué [ɥ] occur later and are involved in far fewer processes. As a result of their late appearance, these two glides are less frequent and exhibit more restricted distributions.

In section 1.3.3 we discuss the processes that contribute to Modern French lexicalised oué [w]. Two late diphthongisations result in *oi* prononced [wẹ] before finishing as [wa] thereby accounting for the prevalence of the orthography *oi* and the diphthong [wa] in Modern French.

1.3.3 Historical origins of Modern French oué [w]

In contrast to processes contributing to yod in Modern French, the processes primarily responsible for lexicalised oué operate relatively late. Diphthongisation by coalescence (12th century) and what we will call French dipthongisation (13th century) occur after the fall of the Roman Empire making them unique to to French.

French diphthongisation, the process Carton calls *la deuxième diphtongaisation,* operates on Latin long ē when tonic and in an open syllable (25).

(26) French diphthongisation ḗ > ói > óẹ > uẹ́ > wẹ́: [wa] *oi*

tḗla > *toile* 'canvas', mḗ > *moi* 'me'

(Lanly 1971: 200-201 glosses added)

Diphtongisation by coalescence contributes significantly to instances of lexicalised [w] (part of the diphthong [wa]) in Modern French. A diphthong by coalescence arises from the coalescing of neighbouring segments (rather than the splitting of a single segment). According to Carton, "la voyelle ne se segmente pas, mais au contraire elle s'unit à un son voisin", (1974:174).

(27) Diphtongaison by coalescence: Vj→[wa] *oi*, [wẽ] *oin*(C)

a. ẹ+j > oi : nĕcā́re > noyer 'to drown'
b. au+j > oi : gaudia > joie 'joy'
c. ē+j > ei > oi : tḗctu > toit 'roof'
d. ō+j > oi : vṓce > voiz, voix 'voice'
e. ọ+jn > *óĩn > *óẽn > *wẽ́(n)

(Lanly 1971: 200-201, cf. also Fouché 1969: 254-320 glosses added)

In (26) we show a variety of contexts for dipthongisation by coalescence. Among these, (26e.) results in a nasalised diphthong results [wɛ̃] (cf. 1971:200 also *éloigner* 1971:201).

Coalescence is a widely observed process that occurs in many conditioning environments. Though frequently the coalescence is between two vocalic segments it is not exclusively vocalic (calidu > chaude, for example). For presentation purposes, we have greatly simplified the evolutions here (for more detailed accounts cf. Fouché 1966: 273). Both French diphthongisation (25) and diphthongisation by coalescence (26) result in *oi* [wa] a lexicalised diphthong in Modern French. These two processes, resulting in the same outcome, account for the lexicalised form *oi* [wa; wɛ̃].

Again, we find alternations in the modern language that give evidence of this stage in the historical evolution. The words *esperer* [ɛspere] ~ *espoir* [ɛspwar], for example, show an alternation *e/oi*. According to the CNRTL (on-line), the noun *espoir* is derived from the strong verb forms "[d]éverbal de *espérer** d'apr. les formes fortes de l'ind. prés. ([*j'*] *espoir,* [*tu*] *espoires,* etc.)". These strong forms no longer exist in Modern French (*j'espère*; *tu espère*s). As we saw in the previous section with the alternations e/ie the presence of the tonic accent has a significant role in the process of diphthongisation and by extension these alternating forms. These strong forms bear the tonic accent and therefore undergo the French diphthongisation: ẹ́ > ói > óẹ > uẹ́ > wẹ́: [wa] *oi*.

Finally one other source for [w] is observed in the consonantisation (glide formation) of the high vowel [u] in hiatus rendering a [w] (27 below, cf. also Lanly *poêle* (1971: 261); *louer* (1971: 205). For ease of presentation we have replaced the u̯ used by Carton with [w].

(28) [w] from hiatus *oui* [ui] > [ẅi] , *oue* [ue] > [we]

a. [we] : *fouet 'whip'*, *loué* 'rented'
b. [wi] : *enfouir* 'to bury', *Louis* (Carton 1974: 163)
c. [wɛ], [wa] : *poêle* 'skillet'

(Lanly 1971: 260, Fouché 1966: 441-2 glosses added)

Though hiatus resolution does occur in diachrony, it is not at all clear that syneresis is, in fact, generalised for these cases. Neither is it clear that hiatus is systematically resolved; counter-examples to (27c.) are plentiful *bohème, poème, poète,* and *poésie*.

We turn our discussion now to the origins of ué [ɥ]. Of the three French glides, ué is the least prevalent; in lexicalised contexts it occurs exclusively as a nuclear diphthong. Although the presence of ué [ɥ] can be linked to the late emergence of the corresponding front rounded vowel [y], the lexicalised diphthong [ɥi] results from a process of coalescence that involves the raising and palatalisation of [o]. Lexicalised [ɥ] occurs in restricted contexts thereby accounting for its scarcity as compared to other lexicalised glides.

1.3.4 Historical origins of Modern French ué [ɥ]

Ué and its corresponding vowel [y] were never part of the Latin inventory and as such are clearly Gallo Romance innovations. The case for an inherent link to the presence of the corresponding high vowel is made most apparent since *ué* does not appear until the corresponding high vowel [y] emerges; as early as the 8th century according to Lanly (1971:192). [y] occurs as the result of fronting or vocalic palatalisation (cf. Fouché 1966, Bourciez 1967, Lanly 1971, Carton 1974). According to Fouché:

> [...] avant la chute des voyelles finales, *úọi̯* a passé en effet à *úi̯* par suite de la fermeture de *ọ* médial sous l'action combinée de *ú* et *i̯*. [...] Dans la suite *úi̯* s'est palatalisé en *üi̯* [yi], prononciation du français primitif. Puis, dans le courant du XIIe siècle, l'accent s'étant déplacé sur l'élément le plus audible, *üi̯* est devenu ẅi, prononciation du français actuel.
>
> (1966: 328, cf. also Bourciez 1967:151-152).

Examples of this process for the words *cuir and nuit* are shown in (28):

(29) o+j > oj > yi, *ui* [ɥi]
  a) cŏrĭu > [kuọ̑irju] > [kyiʀ], *cuir* 'leather' [kɥiʀ]
  b) nŏcte > [nojt] > [nyi], *nuit* 'night' [nɥi]

(Based on Carton 1974: 163, cf. also Lanly 1971 *huis*: 190-192 glosses added)

Fouché describes the evolution of the verb *suivre* 'to follow' this way:

> [a]près la palatalisation de *u̯*, **siu̯re*, **siu̯*, **siu̯s*, **siu̯t* sont devenus **siu̯r̈e*, **siü̯* **siu̯s̈*, **siu̯ẗ*. […] Puis **siu̯r̈e*, **siü̯* **siu̯s̈*, **siu̯ẗ* se sont invertis en **süi̯sre, *süi̯, *süi̯s, *süi̯t;* d'où grâce au changement de *üi̯ en ẅi, *sẅire, *sẅi, *sẅis, *sẅit* "
>
> (Fouché 1966: 332).

According to Fouché, this evolution can be assumed for other words containing [ɥ] such as *fuir* (cf. Fouché 1966: 332, REMARQUE 1). Lanly discusses the emergence of [y] followed by the diphthongisation by coalescence in the context of the evolution of Latin ŌSTIUM to Modern French *huis* 'door'.

We note the displacement of the tonic accent discussed by Fouché above. Glide formation (consonantisation) takes place only after the tonic accent is displaced to the following vowel [i]. Placement of the tonic accent plays a significant role in the diachronic processes that result in the consonantisation of the high vowel in a HVV sequence that we would submit can also be observed in modern derived glide formation.

We turn now to a summary of observations from our discussion of the historical origins of each of the lexicalised glides of Modern French.

1.3.5 Observations on historical origins of lexcalised glides

Although diachronic observations do not inform the native speaker grammar, the various changes that occur throughout the diachrony can inform our understanding of the kinds of articulatory pressures that have brought about changes in the Modern French glide inventory. The prevalence of yod and by extension its broader distribution among the lexicalised glides in Modern French is unquestionably the result of an early emergence in the language, and origins in the prevalent Latin vowels [ĕ] (tonic and free) and [ā́]. We have shown that Modern French lexicalised yods result from early processes of palatalisation, diphthongisation or hiatus which account, in a large part, for the prevalence of yod in Modern French represented by the common orthography -*ie*- (*pièce* [pjɛs], *pied* [pje], *miel* [mjɛ], *rien* [rjɛ̃] as shown in 31d. and 31f. below).

By contrast oué and ué occur much later and in fewer contexts (diphthongisation primarily). We offer a summary of these in (30).

(30) Processes contributing to instances of lexicalised [w] and [ɥ] in Modern French

a. Diphthongisation by coalescence (12th century)

i) Vj→[wa] *oi*, [wẽ] *oin*(C)

ii) o+j > oj > yi, *ui* [ɥi]

b. French diphthongisation (13th century): ẹ́ > ói > óẹ > uẹ́ > wẹ́: [wa] *oi*

c. Hiatus resolution: *enfouir'bury'*, *Louis*

We have shown that apart from some hiatus resolution contexts (30c. below), lexicalised oué [w] results primarily from late processes of diphthongisation: diphthongisation by coalescence (30a.i) and French diphthongisation (30b.). These processes result in modern diphthongs: [wa], [wẽ], represented by the spelling convention *oi*(n)(C) thus explaining the prevalence of *oi* [wa] in Modern French. The front rounded high vowel [y] and corresponding glide [ɥ] are innovations in Gallo Romance emerging as late as the 8th century. The lexicalised diphthong [ɥi]) develops through a process of diphthongisation by coalescence (30a. ii) which is in turn represented in Modern French by the spelling convention *–ui*-.

All three Modern French lexicalised glide contexts occur in Modern French as a result of diphthongisation of a tonic mid back vowel in the evolution of the language. In each case of historical diphthongisation, the tonic mid vowel undergoes diphtongisation that results in one high vowel [i], [u], [y] element and a second non high element [a], [e], [ɛ] that are realised as a diphthong: [je], [wa; wɛ̃], [ɥi]. Traditionally only two nuclear diphthongs are described for Modern French: [wa] and [ɥi], when by all indications the sequence [je] resulting from Romance diphthongisation should also constitute a diphthong (cf. Heap, Nadasdi and Tennant 1992, hereafter Heap et. al. 1992). Instead [je] is generally treated as a GV sequence in Modern French treatments (Lyche and Durand 1999, Bullock 2002).

We will come back to this in our discussion of theoretical implications for glides in Chapter 4.

In the following section we summarise lexicalised glide contexts amongst French glide data.

1.3.6 Summary of lexcalised glide contexts

The categorisations outlined below provide a summary of all lexicalised contexts which form the basis upon which we distinguish between lexicalised and derived glide contexts in our empirical study (cf. Chapter 6). The former occur in the lexicon in their glide form thus they are lexicalised) while the latter may occur when a derived environment (inflection or suffixation) triggers a rule of glide formation that transforms the high vocoid in hiatus (dieresis) to a glide (GV, syneresis).

In (31 below), we elaborate the various categories of lexicalised glides. Our categorisation is based not only on distribution within the word (and syllable) but it is also based on most common orthographic conventions for these various environments since these shared orthographic conventions can indicate their shared historical origins.

(31) **Lexicalised: [GV], et [G] onset [j w], coda [j], nuclear [GV]**

a. Absolute initial [G]:

i) *hi*V- : *hier* [jɛːʀ] [i.jɛːʀ] [i.ɛːʀ] ‘yesterday’; *hiatus* [(ˈ)ja.tys] ‘hiatus’

ii) *i*V- : *iambe* [jɑ̃b] ‘iambe’; *iode* [jɔd] ‘iode’, *ion* [jɔ̃] ‘ion’; *iule* [jyl] ‘centipede’; *iota* [jɔ.ta] ‘iota’ ; *iouler* [ju.le] ‘to yodel’

iii) *y*V- : *yaourt* [jaurt] ‘yoghurt’, *yeux* [jø] ‘eyes’, *yéti* [jɛ.ti] ‘yeti’

iv) *w*V- : *wiskey* [wi.ski] ‘whiskey’, *wallon* [wa.lɔ̃] ‘walloon’, *watt* [wat] ‘watt’

v) *ou*V- : *ouest* [wɛst] ‘west’, *ouïe* [wi] ‘hearing’

b. Medial (onset): V$jV

i) -V*y*V- [j]: *payer* [pe.je] ‘to pay’, crayon [cʀɛ.jɔ̃] ‘pencil’, *moyen* [mwa.jɛ̃] ‘average’

ii) -V*ille*V-: *ailleurs* [a.jœːr] ‘elsewhere’, *tailler* [ta.je] ‘to cut’

c. Final (coda) $Vj

i) -V*il*: *œil* [œj] ‘eye’, *soleil* [sɔ.lɛj] ‘sun’, *ail* [aj] ‘garlic’, *bail* [baj] ‘lease’, *éventail* [evɑ̃.taj] ‘fan’

ii) -V*ille:* paille [paj] ‘straw’, *volaille* [vɔ.laj] ‘poultry’, *groseille* [gʀo. zɛj] ‘currant’, *oreille* [ɔ.rɛj] ‘ear’, *nouille* [nuj] ‘noodle’, *quille* [kij] ‘keel’

d. Lexicalised: CjV (nuclear)

i) -C*ien* [jɛ̃], -C*ier* [je]: *acier* [a.sje] ‘steel’, *dossier* [do.sje] ‘file’, *panier* [pa.nje] ‘basket’, *bien* [bjɛ̃] ‘well’, *chien* [ʃjɛ̃] ‘dog’, *combien* [kɔ̃.bjɛ̃] ‘how much’, *rien* [ʀjɛ̃] ‘nothing’, *soutien* [su.tjɛ̃] ‘support’

ii) C*iè-* [jɛ], C*ie-* [je]: *pièce* [pjɛs] ‘room’, *nièce* [njɛs] ‘niece’, *miel* [mjɛl] ‘honey’, *fiel* [fjɛl] ‘bile’, *pied* [pje] ‘foot’

iii) C*e-*/*Cie-*: venir/vient [vjɛ̃] ‘to come/comes’, *tenir*/*tient* ‘to take/takes’

e. Nuclear: [GV]

i) *-oi-* [wa]~[wɛ̃]: *proie* [pʀwa] ‘prey’, *coin* [kwɛ̃] ‘corner’

ii) *-ui-* [ɥi]: *huitre* [ɥitʀ] ‘oyster’, *truite* [tʀɥit] ‘trout’, *huit* [ˈɥi(t)] ‘eight’

f. Lexicalised suffixes

i) *–ier* [je] /*–ière* [jɛʀ] adj. *langagier/langagière* [lɑ̃.ga.ʒje; lɑ̃.ga.ʒjɛʀ] ‘language’, *ecolier/ecolière* [e.kɔ.lje; e.kɔ.ljɛʀ] ‘schoolboy’

ii) *–ier* : *palmier* [pal.mje] ‘palm tree’, *poirier* [pwa.ʀje] ‘pear tree’, *saladier* [sa.la.dje] ‘salad bowl’, *ouvrier* [u.vʀi.je; u.vʀi.jɛʀ] ‘worker’

iii) *–ien* [jɛ̃]/ –ienne adj. *parisien* [paʀizjɛ̃; paʀizjɛn]

iv) *–ième* [jɛm; ijɛm]: *dixième* [di.zjɛm] ‘tenth’; *quatrième* [ka.tʀi.jɛm] ‘fourth’

(Rey 1963, Juilland 1965, CRNTL online)

We distinguish lexicalised yod [j] and *oué* [w] in absolute initial position or onset (31a.i-iv) from those glides that are actually nuclear (diphthongs 31a.v). Yod [j] occurs in initial position represented orthographically by *hi*V- (31a.i), by *i*V- (31a.ii), and *y*V- (31a.iii). Oué [w] in initial position may be represented orthographically by *w*V- (31a. iii) or by *ou*V- (31a.iv). In (31b.) we show word medial yod represented orthographically by -V*y*V- or the sequence -V*ille*V-. This word initial yod associates with the syllable onset whereas in (31c.) each of the various orthographic representations (-V*il*, -V*ille*) represents word-final yod that necessarily associates with the coda. In (31d.) we distinguish a number of monomorphemes that share a common orthography for the rime. In (31d.i) we show word medial -C*ien*- [jɛ̃] or –C*ier*- [je] which we distinguish from word initial C*iè*- [jɛ], C*ie*- [je] in (31d.ii) and finally, from the stem changing verb forms *venir→vient*, and *tenir→tient* in (31d.iii) that are particularly interesting since this is a rare case of historical alternation that has been maintained in the modern language. The nuclear diphthongs -*oi*- [wa] or -*oin*-[wɛ̃] and -*ui*-[ɥi] in (31e.) occur as the result of early French diphthongisation and represent the only examples of non derived word medial *oué* [w] and *ué* [ɥ] in the modern lexicon. In (31f.) we include some of the glide initial (yod) suffixes of French.

In the next sections, we will examine variability attested for lexicalised forms. We begin with an examination of variability as discussed in the philological analyses of Marchello-Nizia (1977). Following this examination of metrics in the literary tradition, we continue with variability as discussed in 17th century grammatical tradition (Ménage 1675; 1972). Finally, we examine the state of variability for lexicalised glide contexts in Modern French.

1.4 Variability of lexicalised glides

As observed in section 1.3.1, changes in pronunciation as discussed above occur gradually. In this section, we examine some of the circumstances surrounding diachronic variability. The evolution of French grammar in the 13th - 17th centuries, in particular,

exhibits considerable high vowel/glide variability until many changes are regularised by the grammatical tradition (Morin 1976, Marchello-Nizia 1997). According to Morin: "there must have been a phonetic change in the nature of glides in this period, making them more constricted and therefore less capable of appearing after an OL-cluster" (1976: 41). Morin the later deveopement of relaxation of the tension in OLG clusters observed in data from the *Atlas linguistique de la France* (ALF).

Modern descriptions (Léon 1978, Tranel 1987), have lexicalised glides as regularised and invariable. Some lexicalised forms are more stable than others. Where variability is evident it is most commonly attributed to stylistic or regional variation (Léon 1978) but there is evidence which points to individual variation not often discussed in descriptive literature (Martinet and Walter 1973). We will show that modern French variability is linked at least in some way to the quality of the high vowel/glide and sometimes to preceding context; the ability to recover a high vowel realisation is more evident for [j] than for the two rounded glides [ɥ] and [w].

1.4.1 Variability in Metrical Analysis

We have shown that, although diachronic processes do not involve a high vowel at their inception, they do eventually result in a HVV, in which the high vowels will undergo consonantisation to be realised finally as a GV sequence. Of course these gradual processes during which considerable variability occurs before the glide pronunciation (GV) is generalised. Marchello-Nizia discussing the nature of hiatus reduction offers this: “même lorsque le compte des syllabes assure la fréquence de cette réduction, des graphies conservant l’hiatus subsistent longtemps encore, et les deux types de formes peuvent se rencontrer dans le même texte. ” (1997: 69).

Before consonantisation is generalised the HVV sequences born of the processes of diphthongisation are mostly realised as HVV sequences in dieresis but the phonetic and phonological systems are still evolving, thus a period of high vowel~glide variability ensues. According to Lanly “[e]n a. fr., du moins au XII^{e} siècle, le i de ier est réduit à yod (ier compte pour une syllabe) […] Mais au XVII^{e} siècle, il redevient pour un certain

temps une voyelle (-ier est alors dissyllabique)" (1971 : 39). There is evidence of a strong association between a glide and its corresponding high vowel even in lexicalised forms very early on.

In a survey of the metrical practices in the literary works from various authors, Marchello-Nizia identifies word forms which show periods of the instability across several centuries (13th-16th centuries). Variability is attested for words involving all three high vocoids. Words with spelling in *-ien* remain variable through the 16th c. with incongruency between popular usage and literary forms (Marchello-Nizia 1997:74), while those exhibiting ié-/ïé: [je]/ [iə] ... (Marchello-Nizia 1997:82) show stability as early as middle of the 13th c. Words with [we] / [wɛ] show evidence of a generalised Parisian usage in evidence from 13th -16th centuries that merits its own spelling convention: *troas, moas.* According to Marchello-Nizia, "Ce n'est guère qu'à partir du milieu du XVIII^{e} que cette prononciation [wa], d'abord dans un petit nombre de mots seulement, sera acceptée par les grammariens" (Marchello-Nizia 1997: 79-80). With respect to the diphthong [ɥi], pronunciations varied across contexts depending on the rime required, such that sometimes the high vocoid pronunciation is employed: *s'enfuir*: *reçut*, while other times it is the diphthong: *vi*: *bui*, "mais au XlIV^{e} siècle, il ne s'agit plus sans doute que d'une tradition, la diphtongue étant alors devenue [ɥi][…]" (Marchello-Nizia 1997:86).

Although textual variation during this period (14th and 15th centuries) demonstrates a general acceptance of pronunciation with dieresis, HVV sequences (*–ien* in *ancien*, *oï* in *Roïne* and uï in *fuïr* et *Juïf* are generally treated with dieresis HVV (disyllabic) sequences. Marchello-Nizia cautions that the usage found in texts should not be taken to reflect popular usage, "[m]ais ici aussi usage poétique et usage courant divergent sans doute déja" (Marchello-Nizia 1997:75). We must assume that whatever the role of such a literary device, these texts are understood by a public (baggage and all) and must represent pronunciations that exist in that era.

We turn now to evidence for variability as seen through 17th century grammatical tradition. Ménage (1675; 1972) devotes significant attention to the variability prescribed for words in *–ier* preceded by OL (*ouvrier*, *meurtrier*, for example).

1.4.2 Variation in grammatical tradition: Ménage (1675; 1972)

As a key point of interest, observations made by Ménage suggest a precise moment when the pronunciation of a high vowel (HV) preceded by an OL sequence and followed by another vowel changes from a single syllable: OL<u>GV</u>, syneresis, to two syllables: OL<u>HV.V</u>, dieresis):

> Tous nos vieux Poëtes, généralement, ont fait d'une syllabe de l'*I* précédé d'une mute & d'une liquide, & suivi de la syllabe ER […] Ceux qui ont succedé à Malherbe ; les Gombauds, les Racans, les Chapelains, les DesMarets, les Scuderys, & les le-Moines ; en ont usé de la sorte. Mais aujourdhuy cet *I* précédé d'une mute & d'une liquide, & suivi de la syllabe ER, est constamment de deux syllabes. Notre Poësie a cette obligation, avecque plusieurs autres, à M. Corneille; qui dans sa Tragédie du Cid a osé le premier faire le mot *meurtrier* de trois syllabes.
>
> (Ménage 1675: 498)

This particular variation takes a central position in the discussion that follows which we take as an indication of the moment of change where opinions are most charged on the subject. Ménage devotes more than four pages to this single issue when most other chapters in his text are comprised of no more than a single paragraph, often only several lines (cf. Chapitre CCLXIV p. 495, for example). Ménage's focus illustrates, not only that the variability exists, but also that the Academy continues to oppose such usage:

> Je say bien qu'il en a esté repris par Messieurs de l'Académie dans leurs Sentimens sur le Cid. Mais le temps a fait voir que ça esté injustement ; & qu'on le devoit louer de cette nouveauté, aulieu de l'en blâmer. Je suis un des premiers, aveque Monsieur de Vence, qui ay imité en cela Monsieur Corneille : ayant remarqué que les Dames s'arrestoient, comme à un mauvais pas, à ces mots de *meurtrier*, *sanglier*, *bouclier*, *peuplier*, & autres semblables, lorsqu'ils estoient de deux syllabes. Monsieur de Segrais qui a l'oreille fort délicate, & qui n'est pas moins bon Juge de la Poësie, que bon Poëte, se joignit aussi-tost à nostre parti : & dans la Préface son Poëme Pastoral, il fit une remarque des raisons qu'il avoit d'emploier ces mots de la sorte.
>
> (Ménage 1675: 498)

Ménage is of the opinion that this change in pronunciation is well underway such that to stray from the expected pronunciation elicits a strong reaction and as such it should be embraced as it is by Corneille and younger poets of the day. We cannot help but think that Ménage was a sociolinguist before his time while fighting grammatical prescriptivism he recognises the role of women who appear to privilege dieresis and as such can be seen to be at the forefront of this change.

> Tous les jeunes Poëtes, généralement, en usèrent ensuite de la meme façon. Et ça esté inutilement que Monsieur Des-Marets a voulu s'opposer à cet usage. Voicy comme il en parle dans la Préface de son Clovis : « Quelques poëtes de notre temps se font avisez de leur autorité privée. De faire de trois syllabes les mots d'ouvrier, bouclier, sanglier, meurtrier, levrier, et quelques autres semblables, pour les rendre de plus facile prononciation; quoyque depuis que l'on parle françois, on ne les ait faits que de deux syllabes : comme les mots de Guerrier, Courier, dernier, qui ne font pas plus faciles à prononcer. Mais Poëtes n'ont aucun droit, ny aucune autorité suffisante pour établir une loi nouvelle […].
>
> (Ménage 1972: 499)

Taking a cue from the example of the *Apendix Probi* (Baeherens 1922) from the Latin grammatical tradition, we can be reasonably certain that, when grammarians decry a particular usage, this is a strong indication that the usage under discussion is widely practiced amongst a certain population. Thus, we take Ménage's observations to indicate that this change in usage from syneresis to dieresis is a relatively recent development for poets and grammarians alike. These events would support the position taken by Lanly who observes that the *i* of *–ier* is reduced to yod in the 13^{th} century but at sometime during the 17^{th} century *i* is reanalysed as a vowel. Marchello-Nizia also stipulates that the metrics represented in literary works may diverge from the usage of the general public.

We turn now to an examination of variability in more recent sources of glide data. We will survey glide data from the turn of the 19^{th} century from the *Atlas linguistique de la France* (ALF) and then to more recent discussions from the late 20^{th} century.

1.4.3 Modern French variability

Modern descriptions have made much of the role of the preceding OL environment in glide formation. Morin (1976) surveys Fouché (1966) and Bourciez (1967) and finds that

"[g]lide formation was always more or less optional, with a tendency to become obligatory when the two vowels of the hiatus belong to the same morpheme" Morin (1976: 43).

While determining an exact timeline for the generalisation of dieresis in HVV contexts following OL may not be possible, Morin offers evidence for ordering of processes based on the evidence of schwa insertion patterns found in ALF data.

> We observe in some dialects words where OLG-clusters are relaxed through schwa-epenthesis, even though the glide is historically the initial high vowel of a hiatus, e.g. *prier* [pœrye: perye], *oublier* [ubelye], *trouer* [terwe], *truelle* [tœrẅɛ] (data from Gilliéron, Landreau and Bourulot). In these forms the schwa cannot be accounted for unless glide-formation applied first, and the resulting OLG-cluster was subject to tension-relaxation. In most cases (and in particular SF), however, the initial high vowel of a hiatus after an OL-cluster is now syllabic. This was to be expected: when the effect of the tension began, glide-formation was optional, and therefore every OLG-cluster where the glide is a reflex of an earlier high vowel had a nontense variant with a high vowel instead of the glide.
>
> (Morin 1976: 43)

Morin notes that relaxation does not apply to historical [ɥi] and [wa; wɛ̃] (true diphthongs as in *trois, coin* and *truite*, for example). Historic diphthongs [ɥi] and [wa] are found in contexts preceded by OL sequences, while there are no tokens of a monomorpheme with OLjV (Tranel 1987). The distribution of the diphthong [je] appears to be more constrained. With respect to variable OLG treatments Morin proposes three different grammars:

> (1) <u>tense dialects</u>, in which there are no exceptions to the relaxation of Oly-clusters (e.g. SF); (2) <u>lax dialects</u>, in which *–ions*, *–iez* can always be exceptions to the relaxation of Oly-clusters (e.g. Belgian French); and (3) <u>semitense dialects</u>, in which *–ions*, *–iez* can always be exceptions to the relaxation of Oly-clusters but never Ory-clusters (e.g. some forms of French spoken in Paris; cf. Dell 1973, 1973:258, and in Gaspésie, Québec).
>
> (Morin 1976: 44 our <u>emphasis</u>)

As discussed in Morin (1976), OLG relaxation includes instances involving the high vowel [i] transcribed above as *y*. In addition to the evidence in the preceding discussions, HV/G variability when preceded by an OL sequence is well documented in etymological sources.

Morin's position is that different strategies may be employed variably in order to relax an OLG sequence:

> The phonetic changes described here are all drawn from dialects where the changes are restricted to segments within the OLG-clusters, and in particular yod becomes syllabic only within OLG-clusters, and not elsewhere; epenthetic schwas are introduced between an obstruent and a liquid, only when a glide follows, etc. (Morin 1976: 42).

As a result of relaxations and following glide vocalisation, glide epenthesis may occur subject to the particular dialect, "[w]hen glide formation does not take place, a transitional glide may be observed between the high vowel and the following vowel […]" (Morin 1976: 39).

The Centre National de Ressources Textuelles et Lexicales (CNRTL, online resource of the Centre National de la Recherche Scientifique CNRS) provides commentary on both pronunciation and orthographic conventions. According to their analysis, at the beginning of the 12th century we begin with the form *bucler*, which by 1268, has become *bouclier* and by the 15th century varies between a form without *i* as in *levee de boucler* and a form with *i* as in *faire bouclier de*. The variability, discussed here, appears involve the morpheme –ier [ie].

Using data from the *Atlas Linguistique de la France* (ALF), Morin posits four possible forms of relaxation of OLG tension: "(1) vocalization of the glide, (2) loss of glide, (3) loss of the liquid, and (4) insertion of an epenthetic schwa between the obstruent and the following liquid" (1976: 42). Morin does not limit his discussion to derived data alone, the implication being that, since the phonetic change that brought about the OLG tension in the grammar, derived OLG sequences are blocked while some lexicalised OLG are relaxed. We turn now to an examination of ALF data in an effort to illustrate the degree of variability of glide realisations in lexicalised forms at the turn of the 19th- 20th century.

In the *Notice servant à l'intelligence des cartes* (1902), Gilliéron and Edmont explain that among the words and simple phrases that comprise a survey, their questionnaire includes "un certain nombre de mots, isolé également, que nous savions varier en multiples aires et comme tels plus particulièrement indiqués pour montrer la variété du vocabulaire

gallo-roman (1902: 5). Since we take Paris to be the epicentre of the Northern variety or the norm commonly referred to as Standard French, we examine the degree of variability observed for some common lexicalised forms from the ALF survey points *Seine* and *S. Et-Oise*. This region includes four survey points (217, 226 (Paris), 227, 239) corresponding to Paris and its surroundings.

Figure 1 below shows *Seine* and *S. Et-Oise* region situated in the centre of the northern region of France from ALF map No. 177 *Brouette* (North western regions).

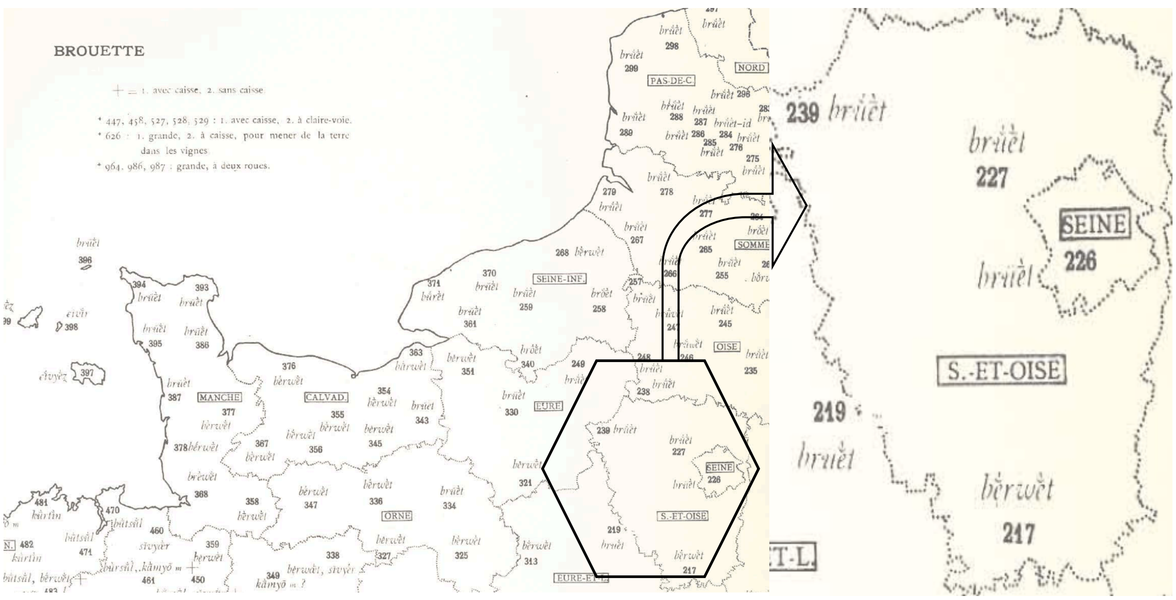


Figure 1: ALF map No 177 Seine et Oise - *Brouette*

Shown in Figure 1 is a portion (Northwestern France) of ALF map 177 *Brouette*. The *Seine et Oise* region is outlined and blown up to the right. Realisations of the word *brouette* ‘wheelbarrow’, an example of a lexicalised form once realised as a monosyllabic with an OLG sequence [bʀwɛt] illustrate variable relaxation methods as discussed in Morin (1976). Points 226, 227, and 239 are relaxed through glide vocalisation (1 according to Morin’s categorisation): [bruɛt], while point 217 is relaxed through schwa epenthesis [bəʀwɛt] (4 according to Morin). Each of these is realised as dieresis though monosyllabic forms can be found elsewhere (ALF points 106 and 8).

In Table 2 below, we show the various realisations for the lexicalised forms *néflier* ‘medlar tree’, *truie* ‘sow’, *ouvrier;ouvrière* ‘worker’, *du bruit* ‘some noise’, *la pluie* ‘rain’, *sanglier* ‘wild boar’ and *tablier* ‘apron/overalls’ all of which contain potential OLG sequences. The transcriptions of ALF data follow from the IPA equivalents set out in Jochnowitz (1973: 17-18. exceptions for /r/ and the *l* mouillé /ʎ/).

Table 2 Lexicalised OLG forms from ALF-Seine et Oise region

	O*l*G				**O*r*G**			
	N° 902 *néflier* ‘Medlar tree’	N° 1039 *la pluie* ‘rain’	N° 1189 *sanglier* ‘wild boar’	N° 1274 *tablier* ‘apron/ overall’	N° 180 *du bruit* ‘some noise’	N° 1342 *truie* ‘sow’	N° 177 *brouette* ‘wheelbarrow’	N° 962 *ouvrier; ouvrière* ‘worker’
Seine								
226	nefje	la plɥi	sãglie	tablie	dy brɥi	tryi	bruɛt	uvrie; uvriɛr
S. et Oise								
217	nefje	la pɥi	sãglie	tabje	dy bri	other	bərwɛt	uvrie; uvriɛr
227	--	la plɥi	sãglie	tablie; tabje	dy bri	other	bruɛt	uvrie; uvriɛr
239	--	la pli	sãgliji	tabliji	dy bri	tʁyi	bruɛt	uvriji; uvriɛr
Strategy Morin (1976)	3	2, 3	1 w/ glide ep	1, 3 w/ glide ep	2	1	1, 4	1 w/ glide ep

Our sample of ALF lexicalised forms, though geographically small, illustrates Morin’s findings:

> All these changes are observed in various dialects with various results. OLy clusters are always relaxed unless the yod belongs to the verbal endings *–ions*, *–iez,* which in some dialects remain nonsyllabic. Relaxation of OLẅ- and OLw-clusters, where *ẅ* and *w* are historic, is at most marginal in SF but frequent in other dialects.
>
> (Morin 1976: 42)

As observed in Morin (1976) relaxation may be obtained by way of various strategies as is indicated across this region (noted at the bottom of Table 2); responses at point 226 Seine (Paris), for example, show that even lexicalised forms are relaxed but not always in the same manner. Historic instances of [ɥ] are maintained in *du bruit* [dy brɥi] and *la pluie* [la plɥi] while *truie* [tryi] undergoes glide vocalisation. Across this region we observe a full array of relaxation strategies; OLG sequences are relaxed through glide vocalisation (1), loss of glide (2), loss of liquid (3) and finally insertion of an epenthetic

schwa (4). As regards choice of relaxation, strategies appear to be conditioned by the particular environments presentes by a word (allowing that some variability may be conditioned by individual speaker). We note that forms in *–ier* (*sanglier*, *tablier*, and *ouvrier*) are all relaxed through glide vocalisation however the speaker at point 239 also employs glide epenthesis. The form *neflier* with only two responses is relaxed by loss of liquid (as is also a possible realisation of *tablier* at points 217 and 227). We note also that a comparison of the O*r*G and O*l*G forms reveals a potential pattern; loss of liquid as a relaxation strategy never employed (by these speakers) for Obstruent+*r* sequences while /l/ is deleted in 3/4 instances (*neflier* points 226, 217; *la pluie* point 217; *tablier* points 217, 227). A cursory examination of all other northern points for *truie, brouette*, *ouvrier/ouvrière,* and *du bruit*, indicates that loss of /r/ rarely if ever occurs.

Morin proposes these strategies are employed specifically to relax tensions caused by existing OLG sequences. In table 3 below, we examine some ALF data for lexicalised forms where the GV sequence is preceded by a single /l/ or /r/.

Table 3: Lexicalised LG forms from ALF-*Seine et Oise* region

	/l/				/r/			
	Nº 124 *bélier* 'ram'	Nº 480 *escalier* 'stair case'	Nº 309 *collier* 'necklace'	Nº 769 *lièvre* 'hare'	Nº 815 *mariage* 'mariage'	Nº 1049 *poirier* 'pear tree'	Nº 1158 *rien* 'nothing'	Nº 1225 *serrurier* 'locksmith'
Seine								
226	beλe	ɛskaje	kɔye	jœv	marjaʒ	pwarje	rjɛ̃	sɛryrje
S. et Oise								
217	beje	ɛskaje	kɔye	jɛv	marjaʒ	pwɛrje	rjɛ̃	sɛryrje
227	beλe	ɛskaλe	kɔye	λɛv	marjaʒ	pwarje	rjɛ̃	sɛryrje
239	beλe	ɛskaji	koyi	λœv	marjaʒ	pwɛrje	rjɛ̃	sɛryrji
	l̯ mouillé = λ							

Table 3 shows the lexicalised forms *bélier*, *escalier*, *collier*, *mariage*, *poirier*, *rien*, and *serrurier* each of which are comprised of a simple onset /l/ or /r/ followed by GV sequence. All of these forms are consistantly realised as syneresis, however, *l*G sequences show loss of liquid [beje] as well as loss of glide [beλe]. The *r*GV sequences on the other hand show no modifications. None of the instances in Table 3 should trigger relaxation of an OLG tension as in Table 2, however, each does involve a liquid followed

by yod. In this environment, it is not unusual to observe palatalisation of a single onset consonant due to coarticulation effects with high vowels or glides. Based on the evidence of the ALF data, /l/ is more susceptible to palatalisation than is /r/. This coarticulatory pressure evidenced by palatalisation of the preceding consonant is most evident for yod in these lexicalised forms. Susceptibility to palatalisation appears to be conditioned by the consonant preceding: *panier* (map 965) points 217-239 [paɲe]; (map 911) *nièce* 'niece' points 217-239 [ɲɛs]; map 84_ *métier* point 226 [métȷ^{ʃ}e:mék̍je]; map 1302 *tiède* points 217, 226 [tȷ^{ʃ}ɛd], points 227, 239 [tjɛd].

According to the description of the ALF notations (Notice 1902), we have interpreted the diacritic symbol (ϵ) superimposed over yod to mean an articulation that is between yod and esh [ɛ] that could be either voiced or voiceless – we assume that in this context yod undergoes voicing assimilation to the preceding voiceless obstruent thus a voiceless articulation that is between a palatal approximant and palatal fricative.

Morin notes that "relaxation of OLẅ- and OLw-clusters, where *ẅ* and *w* are historic, is at most marginal in SF but frequent in other dialects" (1976: 42). Variability is not however limited to OLG contexts. Forms containing ɥV or wV, also show variability which is generally restricted to high vowel/glide alternations: map 928 *de nuages* shows 217 [nɥaʒ], 226 [nɥaʒ: nyaʒ], 227 [nyaʒ] 239 [nɥaʒ]. Carton notes this kind of variability stating that "[d]ans plusieurs régions de la francophonie, on est resté au stade voyelle + voyelle. À Paris, on est passé au stade semi-consonne + voyelle: *nuage* [nɥa:ʒ] au lieu de [nya:ʒ]" (1974 :163).

We also note that despite evidence to suggest that forms in *–iez* [je] and *–ions* [jɔ̃] are less susceptible to variability (even when preceded by an OL cluster), we find variable realisations: Map 400 *vous deviez* 227 [dəvje] but 217, 226, 239 [dvje]; Map 409 *vous disiez* 226 [dizje], 217 [dize].

The ALF data illustrate that variability is indeed an aspect of lexicalised glides even for those points that survey speakers of the Parisian standard. We have shown that such variability is not exclusive to the context of a preceding OL sequence. While less apparent for the two rounded glides /ɥ w/, at the turn of the last century variability of

lexicalised forms appears to be abundant for forms containing yod (Table 2). Lexicalised yod, whether preceded by OL sequence or by a simple C may alternate with its corresponding high vowel though the quality of the preceding C may be a conditioning factor.

We turn now to variability of lexicalised forms described for French in the twentieth century. Not only is variability in the realisation of these forms seen as a dichotomy between syneresis and dieresis but there is also much discussion of the variable status of glides (consonant-like versus vowel-like) based on the evidence of variable sandhi effects for liaison and elision.

1.4.4 Variability of modern lexicalised glides

As we have established previously, the association between glides and a corresponding high vowel is strong. Léon (1978) and Tranel (1987) both describe some variability involving the substitution of a glide with a corresponding high vowel. According to Léon "[d]ans une diction poétique on peut également prononcer [j] en lui restituant sa valeur de voyelle [i], selon les besoins du rythme (basé sur le nombre de syllabes en français); mais cela ne peut se produire que si [j] est précédé d'une consonne" (1978: 29, cf. *rien* in syneresis [rjɛ̃] versus *bohémien* in dieresis [bɔemiɛ̃] as employed by Rimbaud for the purpose of maintaining syllable count). On the subject of ɥ and w Léon offers that "[e]n poésie, par exemple, on peut substituer [y] à [ɥ], [u] à [w][…]" (1978: 32). Although glides may contrast among themselves (*rier-ruer-rouer)*, a glide is never found to contrast with its corresponding high vowel: "ce qui montre que si ces semi-consonnes peuvent s'opposer linguistiquement l'une à l'autre, elles ne s'opposent jamais aux voyelles correspondantes pour distinguer des mots de sens différents (1978: 32). Martinet (1945) reports considerable variation in the realisation of lexicalised forms such as *pied* amongst the officers who were interned with him in 1941. The ALF data we have just examined indicate that variability, defined here as the possibility of high vowel~glide alternation, is attested for yod more so than oué or ué. In many cases the high vowel associated with the glide in question may be produced presumably to serve a stylistic

puropose, be it the goal of giving emphasis, for syllable retention in poetics, or even as a demarcation strategy (Lyche and Girard 1995). In any case, variability where a lexicalised glide alternates with a corresponding high vowel is attested primarily for yod in words such as *lien* and *hier* (cf. Léon 1978), while similar alterability is not normally attested for true diphthongs such as are found in *boîte* *[buat] or *fuite* *[fuit], for example.

Another kind of variabilty observed for lexicalised glides involves variability in segmental status based on varying behaviour of glides observed in French sandhi effects in liaison and elision data. Glides are generally categorised as vowel-like consonants based on the facts of their distribution in margins yet, French liaison and elision data show that glide behaviour is not always predictable (Dell 1973, Tranel 1987, Heap et al. 1992). French liaison involves the appearance of an otherwise absent final consonant (with or without linking) when immediately followed by a vowel initial word (*les* [le] + *idées* [idé] → *les idées* [lezide]; cf. Encrevé 1988). French Elision involves the dropping or erasing of the unstressed vowel of articles when preceding vowel initial nouns (la + idée→*l'idée* 'the idea'; cf. Dell 1973). Some glides in word initial position behave both as consonants by blocking liaison and elision while other word initial glides behave as vowels allowing liaison and elision (32 and 33 below). This apparent ambiguity of segmental class (vowel versus consonant) occurs despite the identical orthography for the glide. Evidence for the ambiguity is illustrated by liaison and elision data (32 and 33).

(32) **Sandhi effects** – elision and liaison blocked

a.	[j] le yahourt [ləjauʀt]	*l'yaourt [ljauʀt]	'the yogurt'
	[w] le watt [ləwat]	*l'watt [lwat]	'the watt'
	[ɥ] le huis [ləɥi]	*l'huis [lɥi]	'the door'
b.	[j] les yahourts [lejauʀt]	* [lezjauʀt]	'yogurts'
	[w] les watts [lewat]	* [lezwat]	'watts'
	[ɥ] les huis [leɥi]	* [lezɥi]	'doors'

(Dell 1973; Encrevé 1988)

In (32), the glides behave in a consonant-like manner since elision of the preceding article is blocked (32a.) and liaison of the plural *-s* [z] is blocked (32b.). However, the very same glides in initial position but in different words behave vowel-like in that they

permit elision (33a.) and liaison (33b.).

(33) Sandhi effects - elision and liaison permitted

a.	[j] l'iode [ljɔd]	*le iode [ləjɔd]	'iodine'
	[w] l'ouïe [lwi]	*la ouïe [lawi]	'hearing (sense of)'
	[ɥ] l'huître [lɥitʀ]	*le huître [ləɥitʀ]	'oyster'
b.	[j] les iodes [lezjɔd]	* [lejɔd]	'the iodines'
	[w] les oiseaux [lezwazo]	* [lewazo]	'the birds'
	[ɥ] les huîtres [lezɥitʀ]	* [leɥitʀ]	'the oysters'

(Dell 1973; Encrevé 1988)

Still others may behave as both vowel and consonant *ouate* [wat] and *ouate* [ˈwat]. Despite evidence to suggest a clear dichotomy of glide behaviour, (Heap et. al. 1992), in a study that examines the inherent variability of glides in elision data, show that variability of glide treatments among speakers is much greater than is usually described.

> Il se trouve que les semi-voyelles (comme nous dirons ici par commodité) se comportent de façon peu prévisible pour ce qui est de l'élision. Une même semi-voyelle en position initiale, par exemple /w/, peut tantôt exiger l'élision (*l'oiseau*), tantôt ne pas la permettre (*le watt*).
>
> (Heap et al. 1992:165)

These authors conclude "[…] qu'en réalité l'élision ne divise pas les semi-voyelles en deux categories étanches. La variation interindividuelle est trop élévée pour nous permettre de prédire à partir des groupe [sic] glide semi-voyelle, si l'élision aura lieu ou non" (Heap et al. 1992:185).

In this chapter we have discussed various aspects of the Modern French lexicalised glides. We have shown that the modern French three-glide inventory is the result of a number of diachronic processes (primarily diphthongisation) that are active at different times throughout the historical evolution of French from Latin. These glides have evolved each at their own pace and resulting from different phonetic pressures and as a result they are quite distinct from one another in their distribution (both in their rate of occurrence as well as their access to the various positions within the syllable and word). Yod is most prevalent followed by oué then ué. This is congruent with the order in which

each glide emerges in the historical evolution (yod first, then oué followed by ué) and the number of processes in which each is implicated.

Through our survey of derived and lexicalised glides we have shown that all glides whether derived or lexicalised are very strongly linked to the presence of a high vowel in hiatus HVV. By alternating with a corresponding glide a bysyllabic sequence HVV (dieresis) may be reduced to a monosyllabic GV (syneresis) presumably in order to satisfy a grammatical dispreference for hiatus. Although all three modern glides have origins in processes of diphthongisation, [je] is often excluded from descriptions of diphthongs. We can perhaps attribute this to the distribution of [j] as compared to the other diphthongs in French; [j] never occurs after OL sequences, unlike [wa], [wɛ̃] and [ɥi]. Despite the possibility for a strong grammatical dispreference, hiatus is attested in French thus allowing for a considerable amount of variability. Variability defined as high vowel/glide alternation is most common in derived forms but we have shown that lexicalised forms are also susceptible. Finally from liaison and elision data we observe that a single glide can behave both consonant-like as well as vowel-like.

In the next section we survey phonetic descriptions of glides from the classical tradition (Passy 1913) to modern descriptions (Ladefoged 1993). In this tradition, phonetic descriptions based on contextual distinctions prove more difficult than would be expected.

Chapter 2

2 Phonetic and Phonological Traditions

This chapter begins with a brief discussion of the origins of the classical tradition in phonetics and its relationship with the normative tradition in French descriptions. Certainly establishing the proper French pronunciation to teach to non-native speakers is influenced by this tradition. It is perhaps as a result of this tradition that variability of glide phenomena remains so underdescribed for French.

We begin with the classical tradition wherein the International Phonetic association (IPA) is born (Passy 1913). The classical tradition, the IPA, and the normative tradition are inherently linked thus influencing all later descriptions until Pike (1943) and the inception of the Prague School of phonology.

By classical tradition, we include the work of several prominent European phoneticians: Passy, Sweet, Storm, and Sievers, whose work in descriptive phonetics of various European languages intended for teaching correct pronunciation to second language learners culminates in the creation of the International Phonetic Association (IPA) and the creation of the international phonetic alphabet (IPA) as a notational tool in this endeavor. Both the association and the alphabet it created continue to influence modern phonetic treatments.

> Quant à la phonétique élaborée dans les années 1870-1880, par l'école dite classique, avec Paul Passy en France, Henry Sweet en Angleterre, Johan Storm en Norvège, Eduard Sievers en Allemagne et un peu plus tard, J. A. Lundell en Suède et Otto Jespersen au Danemark, elle s'intéressait surtout aux applications pratiques (principalement à la correction de la prononciation dans les écoles, à cette époque où, pour la première fois, on attachait de l'importance à la langue parlée, et non seulement aux formes écrites de l'expression).
>
> (Malmberg 1971: 34)

Out of this tradition of language teaching and corrective phonetics, the *Association Phonétique Internationale* (API) is born in Paris (1886) led by these most notable pioneers in the phonetic tradition. According to the association website, the API first met

with a view to advance “the scientific study of phonetics and the various practical applications of that science”[6]. From their regular meetings, the International Phonetic Alphabet (IPA) is first devised and put into service in the description and teaching of phonetics for a variety of languages.

French phoneticians who embrace the classical tradition are thus concerned with teaching the proper pronunciation of French to their students. Glides (semi-vowels) present considerable difficulty to phoneticians who must rely on articulatory distinctions in order to differentiate glides from their corresponding high vowels and vowels from consonants.

> Les semi-voyelles sont encore éminemment des spirantes et aussi bien des fricatives et des constrictives. On leur a donné ce nom de semi-voyelles parce qu’on y a vu des phonèmes intermédiares entre les voyelles et les consonnes. On entend en effet dans le son qu’elles produisent à la fois le timbre d’une voyelle et le frottement d’une consonne spirante.
>
> (Grammont 1963:77)

Since the earliest phonetic descriptions phoneticians have sought, for better or for worse, to classify sounds of human language into one or the other of the two natural classes: consonant and vowel. As Abercrombie points out however the concept of vowel and consonant is based on the phonetic system of Greek where “the categories of ‘general phonetic vowel’ and ‘phonological representative of V element’ (and similarly the categories of ‘general phonetic consonant’ and ‘phonological representative of C element’) coincided more or less” (Abercrombie 1967: 79). Such a system suited the Greek and even Classical Latin descriptions quite well, however, since this time many languages under study do not benefit from a strictly one to one correspondence, as was once the case in ancient Greece.

In view of this Vowel/Consonant dichotomy, what are we to make of the group of sounds called glides (just one of many names) that on one hand appear inextricably linked to a corresponding high vowel while on the other hand behave much more like a consonant (depending on phonological environments)?

---

[6] http://www.langsci.ucl.ac.uk/ipa/index.html

In French, for example, it could be said that the graphemes <w> and <y> in word initial position and followed by a vowel must represent consonants: *whisky* ‘whiskey’ or *yéti* ‘yeti’ (much like the English words wood or yard), for example. Evidence from French liaison and elision however would not always support that position. We turn now to the work of the classical school of phonetics and the difficulties inherent to any description of glides.

We begin with a detailed look at the work of Paul Passy for French. Passy, first president of the International Phonetics Association (IPA), focuses primarily on phonetic description of French for the purpose of teaching to foreign language students the best possible pronunciation of French.

2.1 Passy (1913)

Members of the school of Classical Phonetics, motivated by the practical applications of their phonetic descriptions of language, provide descriptions which serve *principalement à la correction de la prononciation dans les écoles* according to Malmberg (1971). According to Passy himself, his is but an elementary treatment of phonetics: “Il est destiné en première ligne aux professeurs chargés d’enseigner notre langue à des étrangers : Ils y trouveront des indications de première utilité pour corriger les défauts de prononciation de leurs élèves” (Passy 1913: 8). Today, in the context of the language classroom, and our understanding of regional and individual variation the natural question might be to wonder which variety to model. Passy concludes that it should be his own Parisian variety: “Mais nous pouvons affirmer sans crainte qu’elle sera partout reconnue comme acceptable, comme correcte – partout du moins où on a tant soit peu étudié les faits du langage” (1913: 9).

In the next section we look at Passy’s descriptions as they pertain to his observations of the spontaneous speech of native speakers.

2.1.1 Spontaneous Speech phenomena

Distinguishing vowels (*les sons musicaux*) from consonants (*les bruits*): fricatives (*frottements*) from stops (*frappements*), Passy notes that when grouped together in syllables vowels are marked by their higher intensity among sounds:

> Le son le plus intense d'une syllabe s'appelle *son syllabique* ; les autres portent le nom de *sons consonants*. C'est presque toujours le son le plus sonore de la syllabe qui est le son syllabique. Les voyelles se prêtent particulièrement bien à ce rôle [...]. (Passy 1913: 58)

Passy describes articulatory phenomena observed in spontaneous speech involving the reduction of a VV sequence. These are not always specifically high vowel-vowel (HVV) sequences but where these do involve a high vowel followed by another vowel a glide may occur. The first of these contexts involves two vowels in juxtaposition that, according to Passy, may form a single syllable (diphthong).

> Par contre, toutes les voyelles sont syllabiques, ou à peu près. À la vérité, dans un parler rapide, il y a des cas où, deux voyelles étant juxtaposées, la moins sonore ou la moins accentuée perd son rôle syllabique, devient consonante, et forme avec l'autre une *diphtongue* ou réunion de deux voyelles en une syllabe ; ainsi dans certaines rencontres de mots comme *à outrance* (aŭtrɑ̃ːs), *il a écrit* (ilaĕkri), […] *il est si oublieux* (ilɛsiŭbliø) ; et quelquefois dans des mots comme *paysan* (peĭzɑ̃), *chaotique* (kaŏtik), *caoutchouc* (kaŭtʃu), *réuni* (rey̆ni), prononcés rapidement. Mais il est toujours possible de prononcer tout au long (-a-uˈtrɑ̃ːs), etc.
>
> (Passy 1913: 59)

Passy stipulates, however, that on one hand, glides (consonants) unite with the other vowel to form a diphthong while elsewhere, he observes that the glide remains a solitary consonant and does not unite with the following vowel.

> En dehors de ces cas, nous n'avons pas de diphtongues: si deux voyelles se rencontrent, ou bien elles forment deux syllabes comme dans *pays (pei), chaos* (kao), ou bien l'une d'elles se change en consonne, comme dans *où est-il* (wɛti), *ça y est* (sajɛ).
>
> (Passy 1913: 59)

The example *où est-il* ([uɛti]→ [wɛti]), though it is not a derivation shows glide formation across word boundaries. In the case of *ça y est* (sajɛ), modern accounts would suggest that the adverb *y* is lexicalised and will generally resyllabify to a following empty onset (*il y a* [ija] or [ja]). The distinction that Passy makes between the two environments is not phonetically justifiable. It seems to us that these observations would benefit from a phonological analysis. Passy's description (consonantisation resulting in either a *diphthong* as in *à outrance* [aŭtrɑ̃ːs] 'to excess' or an isolated consonant (not a diphthong) as in *où est-il* [wɛti] 'where is it/him') actually reflects the phonological environment. Modern analyses may refer to what Passy describes in (*chaotique* (kaŏ̠tik), *caoutchouc* (kaŭ̠tʃu), *réuni* (rey̠̆ni), for example) as falling diphthongues whereas *où est-il* ([uɛtil]→ [wɛti]) is not commonly described and more likely to be treated as phonologically derived glide formation. Perhaps most significantly, Passy notes the role of sonority and accent in these glide contexts.

The examples from Passy *chaotique* (kaŏtik) 'chaotic', *caoutchouc* (kaŭtʃu) 'rubber', *paysan* (peĭzɑ̃) 'peasant' and *réuni* (rey̆ni) 'united' resemble a typical falling diphthong. These are phonologically derived contexts. In all but one of the examples cited ([aŏ] in *chaotique* (kaŏtik) 'chaotic'), the environement observed involves a /V+HV/ sequence. As we have shown for historical glide data (cf. diphthongisation section 1.3), consonantisation (also semi-vocalisation) can be related to the lack or loss of tonic accent. In these examples, the second less sonorous vowel, and by implication unstressed vowel undergoes consonantisation and forms a diphthong with the preceding vowel ([aŏ], [eĭ], [aŭ], [ey̆]). With respect to sonority scale each of these examples shows a /V+HV/ sequence in which the first is a mid-low or low vowel followed by a second less sonorous vowel (a mid-high or high vowel). It should not be surprising then that the high vowels, least sonorous of all vowels, are universally involved in glide phenomena (cf.section 4.2 Maddieson 1987, section 4.5.1 Rosenthall 1997).

Passy sets apart from the contexts described above the examples *où est-il* (wɛtil) and *ça y est* (sajɛ) evident in his choice of transcription: [ŭ ĭ] versus [w j]. These two contexts are only marginally different from what we have previously described for glide formation in Modern French and from what we saw in the historical evolution of glides since none of these examples involves the primary stress accent (though secondary stress is likely). Given no phonetic basis to justify this distinction, we assume that, without explicitly saying so, Passy makes a distinction based entirely on position. The glides [w j] of *où est-il* [wɛti] and *ça y est* [sajɛ] would serve to fill the otherwise empty syllable onsets whereas the glides of *caoutchouc* [kaŭʧu] or *paysan* [peĭzɑ̃] occur in syllable final position with onsets in the following syllable being occupied. The only difference that distinguishes the two is that in the former this is the first vowel in the series to undergo glide formation whereas in the latter this is the second vowel in the series to become a glide. These vowels like the former vowels become consonants yet they do not form a diphthong with the other vowel in the series. Instead they remain independent consonants. In any case, each of the examples have in common an alternating high or even mid-high vowel with a corresponding glide (glide formation involving mid-high vowels is not uncommon cf. Rosenthall 1997 also Chapter 6), which, in the absence of a morphological operation we describe as phonologically derived glide formation (GF).

Passy observes that the vowel to undergo consonantisation is necessarily the least sonorous of a VV series; the implication being that GF is sensitive to primary and possibly even secondary stress accents. The first of vowel of the VV series is unstressed thereby rendering it least sonorous, so that it may be realised as a corresponding glide.

Data from Passy (1913) supports the position that glide formation occurs not simply word internally but also across word boundaries. Passy makes a number of important observations concerning the various outcomes when two vowels are in juxtaposition, curiously however, he does not discuss the two vowels in juxtaposition in the final syllable of *oublieux* in the example of *il est si oublieux* (ilɛsiŭbliø). These two vowels in spite of the juxtaposition do not undergo glide formation. Passy acknowledges that two

vowels in juxtaposition may remain unchanged maintaining hiatus as in *pays* (*pei*) and *chaos* (kao). Passy may simply consider this context to be one such instance. We take this as a testament to the strength of the blocking effect of a preceding OL sequence in operation in Passy's Northern French variety, at that time.

These observations that focus on phonetic features of spontaneous speech phenomena specifically cross word-boundary (#) glide formation bring us to a discussion of sandhi effects and in particular French liaison and elision as discussed next.

2.1.2 Sandhi effects: liaison and elision

In a chapter called *Combinaison des sons* Passy discusses the nature of phonetic change or combinatory effects on sound as he observes it for spontaneous speech. Included among such combinatory effects are the *sandhi* effects of liaison and elision. More than simple combinatory effects, Passy describes the *rules* of elision (given its regularity) as constituting part of the official grammar (Passy 1913:115). Elision he describes as "la chute complète d'un son, dans certaines combinaisons [...] (lami) pour (lə ami), (lekɔl) pour (la ekɔl) (Passy 1913:115). As for glides (here called *semi-voyelles*), he reports that elision may occur but with some variability.

> *Remarque* – La voyelle (ə) de *le, me, te, de, je,* et la voyelle (a) de *la*, s'élident, non seulement devant une voyelle, mais souvent devant les semi-voyelles (j, w, ɥ). On dit *le yak* (lə jak) mais *l'oiseau* (l wazo); *la yole* (la jɔl), mais *l'huître* (l ɥitr); (la jɛːn) ou (l jɛːn), (la wat) ou (l wat), (lə jɔd) ou (l jɔd).
>
> (Passy 1913:116)

Liaison, however is described as much more stylistic; "[l]'emploi des liaisons varie considérablement selon le style et selon les personnes" (Passy 1913:117). Passy reports variable liaison for words beginning with a glide.

> Pour la liaison comme pour l'élision, les semi-voyelles (j, w, ɥ), sont traitées tantôt comme voyelles et tantôt comme consonnes. On dit *les yeux* (lez jø), *deux oies* (døz wa), mais *les Wallons* (le walɔ̃), *les Yakutes* (le jakut); *les hyènes* (le jɛːn) ou (lez jɛːn), *avant-hier* (avɑ̃jɛːr) ou (avɑ̃tjɛːr).
>
> (Passy 1913:116)

The data in (34) show the extreme variability possible in the sandhi effects described by Passy. Recall that in (32-33, above), we showed that the application or non application of liaison and elision may occur for all three French glides. Using Determiner + Noun groups, Passy illustrates the dual nature of glides in elision and liaison contexts. In (34a.) we observe that at least some words allow both possibilities of elision or no elision and in (34b.) we observe the same duality for liaison.

(34) Variable liaison and elision (for a single word)

a. Elision
   i) *ouate* 'cotton batting' [la wat] or [lwat]
   ii) *iode* 'iodine' [lə jɔd] or [ljɔd]

b. Liaison
   i) *les hyènes* 'hyenas' [le jɛːn] or [lez jɛːn],
   ii) *avant-hier* 'day before yesterday' [avɑ̃jɛːr] or [avɑ̃tjɛːr]

(Passy 1913:115-117)

In the previous section, we discussed a distinction Passy makes so that a VHV sequence results in a diphthong while HVV sequences result in an isolated consonant for which he uses different transcriptions; [ŭ ĭ] as in *caoutchouc* [kaŭʧu] and *paysan* [peĭzɑ̃], for example and [w j] in *où est-il* [wɛti] and *ça y est* [sajɛ]. Although we find no phonetic reason (but for a possible difference in saliency) to distinguish [ŭ] from [w] or [ĭ] from [j] in these contexts, as we stipulated previously for these examples, it seems, at least possible if not probable, that Passy makes this distinction based on syllable structure. It should be implicit that the glides in *où est-il* [wɛti] and *ça y est* [sajɛ] occupy an otherwise empty onset whereas in *caoutchouc* [kaŭʧu] and *paysan* [peĭzɑ̃] the following onset definitely is not available. Thus the glide would be considered a constituent of the nucleus (or coda) of the first syllable. According to Passy, the VG sequence of the latter unites to form a diphthong while the former remain isolated. We really have no way to verify the syllable constituency when the glide is word internal whereas the French liaison and elision data do provide us with some very valuable diagnostic insights into the status of the glide based on the presence or absence of liaison or elision. Although we know that status may vary for a given glide, Heap et al. (1992) report that even for the

same word and for the same speaker glide treatment may in fact be variable.

2.1.3 Classifications of glides

In the preceding sections, we have illustrated the duality of glides and even their ambiguous nature. Glides in the classical tradition of phonetics are almost enigmatic. By their various names we observe the difficulty inherent in their classification in a phonetic description, particularly when adhering to the vowel~consonant dichotomy that presents. Or, is it that we are to understand the nature of the glide in a specific context by its name (semi-vowel versus semi-consonant, for example)? In this description, Passy often refers to glides as consonants either in isolation or as part of a diphthong. He also refers to glides as semi-vowels that are treated sometimes like vowels sometimes like consonants. These names are sometimes accompanied by a different transcription (as we noted above [ŭ ĭ] as in *caoutchouc* [kaŭʧu] and *paysan* [peĭzɑ̃], for example), and [w j] in *où est-il* [wɛti], and *ça y est* [sajɛ]. We might expect that distinctions as in [ŭ ĭ] versus [w j] have to do with noise-like qualities (either *frottement* or *frappement*) that would distinguish the semi-vowel from the consonant (constriction in modern treatments, perhaps).

Most often Passy's observations pertain to a single variety of French (his own, which he considers to represent the standard Parisian French variety emblematic of a proper (most desirable) pronunciation. He acknowledges, however, glide phenomena for regional varieties of French focussing on the ways in which these varieties vary with respect to their phoneme inventories, phoneme distribution and phoneme variants specifically in instances of two vowels juxtaposed.

> Il s'agit, ne l'oublions pas, de la prononciation courante de la région parisienne. Dans le Midi, on emploie souvent des *voyelles consonantes* à la place de *nos consonnes* (j, w, ɥ) ; on dit (mĭɛ̃, bŭɑ, ny̆aːʒ) où nous disons (mjɛ̃, bwɑ, nɥaːʒ). Dans la Suisse romande, certaines voyelles longues (ou autrefois longues) peuvent être diphtonguées : *année* (aneĭ). Ailleurs on dit (paĭ) où nous disons (pɑːj). (Passy 1913: 59-60 emphasis added)

Thus in Passy's variety of French, we should expect to hear the forms *mien* [mjɛ̃] 'mine', *bois* [bwɑ] 'wood' and *nuage* [nɥa:ʒ] 'cloud', which we take to be a monosyllabic GV pronunciation. In the Midi French, however, we will hear instead vowel consonants plus vowel [mĭɛ̃], [bŭɑ], [ny̆a:ʒ]. It is unclear as to what phonetic distinction is made in these classifications. We look to Passy's descriptions of the different subgroups of vowels and consonants (cf. §§145-225), however he does not explicitly describe what is meant by vowel consonants (ĭ y̆ ŭ) or how they may be distinguished from *nos consonnes* [j, w, ɥ]. Neither does he explicitly describe a subgroup of vowels that can become consonants (i, u, y, for example). What he does describe are three fricatives among the consonant inventory of French (cf. §§203, 204, 210) that correspond to his description *nos consonnes* [j, w, ɥ] above. Here, Passy distinguishes the bi-labial fricatives /ɥ / and /w/ from other non-French bi-labial fricatives /F/ and /ʋ/ by their combined articulation, "[...] au rétrécissement bilabial se joint un rétrécissement dans une autre partie de la bouche: ce sont à proprement parler des *consonnes composées,* mais dans lesquelles l'action des lèvres domine" (Passy 1913:100). Further to this distinction among bi-labial fricatives sets apart a palatal fricative [j] while grouping these with liquids:

> La première de ces consonnes, (ɥ), est le *u* de *huile* (ɥil), *huit* (ɥit), *buis* (bɥi), *luire* (lɥi:r), *nuée* (nɥe), *nuage* (nɥa:ʒ) […] La seconde fricative bilabiale (w), se trouve dans *oui* (wi), *ouest* (wɛst), *whist* (wist), *loin* (lwɛ̃), *doit* (dwa)[...] La *fricative palatale* (j) se forme en approchant le milieu de la langue du palais dur, et en chassant l'air : *yak* (jak), *hyène* (jɛ:n), *bien* (bjɛ̃), *rien* (rjɛ̃), *paille* (pɑ:j), *médaille* (meda:j). Comme (ɥ), (w), et les liquides, (j) peut être dévoisé sous l'influence d'une consonne soufflée précédente ou suivante, comme dans *pied* (pj̊e), *feuilleter* (fœj̊te). (Passy 1913: 100-103)

Passy stipulates that French has no diphthongs but for those arising from two vowels in juxtaposition, we must assume then that *paille* (paĭ) as in the Midi French pronunciation with a vowel consonant is a diphthong, which contrasts with (pɑ:j), which is not a diphthong but instead a long vowel /ɑ:/ followed by the consonant /j/ (a closed syllable). This would be consistent with what we have already discussed with respect to the

distribution of /j/ in syllable-final or in word-final position whereas the other two glides /w/ and /ɥ/ are excluded from this position.

From a strictly phonetic viewpoint, it is not clear what distinction is being made here between the vowel consonants /ĭ ŭ y̆/ and the fricatives /j w ɥ/. We must assume then that the glide sounds of Parisian French are that much more noise-like than those in the same context in Midi French. According to Passy the fricatives in *mien* [mjɛ̃], *bois* [bwɑ], and *nuage* [nɥaːʒ] are consonants, with no mention of a relationship to a corresponding vowel and not necessarily variable as is mentioned for the derived forms; "il est toujours possible de prononcer tout au long (-a-uˈtrɑ̃ːs)" (1913: 59).

Passy is, nonetheless, aware of the ambiguity that glides present for a description and adds these remarks:

> *Remarque.* – Les trois consonnes (ɥ), (w), (j), qui n'existent régulièrement chez nous qu'à l'état de sons voisées, se distinguent des autres fricatives par la faiblesse du frottement ; si bien qu'on peut se demander parfois si on n'a pas affaire à des voyelles consonantes [§109]. Toutefois quand on les prononce soufflées le frottement est encore assez sensible [du moins dans le Nord de la France] pour leur mériter le nom de consonnes. On leur donne parfois le nom de *semi-voyelles* [Cp §§ 241, 243]. (Passy 1913: 103)

In his earlier description of those high vowels that would correspond to these consonants Passy also attributes to these segments weak frication.

> Dans le langage aussi il y a des sons mixtes, pour lesquels le son de la voix s'unit à la présence de la voix; c'est aussi le cas des voyelles telles que (i, u, y), pour lesquelles le son de la voix s'unit à un frottement plus ou moins prononcé. (Passy 1913: 24)

In the various distinctions that Passy makes we have as many as three different names for glides that are observed in various contexts with very little phonetic criteria to distinguish one from the other. Glides in French that are either consonants (bi-labial or palatal fricatives) also called *consonnes composées* (compound consonants) or they are called *semi-voyelles* (semi-vowels) /j w ɥ/ as in *mien* [mjɛ̃], *bois* [bwɑ], and *nuage* [nɥaːʒ].

We turn now to Jones (1932) whose description of glides in English further illustrates the issues for glide descriptions in the phonetic tradition.

2.2 Jones (1932)

According to Fischer-Jørgensen (1975), Jones discovered he had a 'particular talent' for language; he studied phonetics in Germany and afterwards with Paul Passy developing his own interest in the 'practical applications' of phonetics. While Jones would undoubtedly have been influenced by the work of Passy, he advances many of his own notions about the phonetics of the speech signal while building on the earlier work of his mentor.

Jones' treatment of glides in English includes a three-way distinction in the naming of glide-like sounds: semi-vowels, consonantal vowel and frictionless continuant. Distinctions between sound classes for English are based on notions of prominence and sonority.

> It is a consequence of this principle of relative prominence that certain short vowel-glides must be regarded as consonants. Such are the English j (as in *yard* jɑːd) and w (as in *wait* weit). In making these sounds the speech-organs start in the position of i and u respectively and without remaining there any appreciable time proceed very quickly to the position of another vowel (ɑ in the case of *yard* and e in the case of *wait*). Such vowel-glides are often called *semi-vowels*. It must be remembered that such sounds have to be regarded as consonants on account of (1) their gliding nature, (2) their shortness and (3) their lack of stress as compared with the succeeding vowel. (Jones 1932:25 original emphasis)

Jones also uses the term consonantal vowel to denote the glide portion of a diphthong: "[t]he term consonantal vowel is sometimes used to denote the less prominent part of a diphthong. Thus it is said that the English diphthong **ai** consists of 'an **a** followed by a consonantal **i**.' This manner of regarding a diphthong, though not quite accurate, is sometimes convenient in practical teaching." (1932: 59).

Different from Passy's descriptions, Jones does not refer to friction (noise) in the articulation of glides but rather introduces a subgroup of consonants called frictionless continuants.

> There exist voiced consonants which have the same or very nearly the same articulatory positions as fricatives, but in which no friction is audible; the absence of audible friction is due either to the fact that less exhaling force is used than for corresponding fricative, or to the fact that the aperture at the place of articulation is somewhat wider, or to a combination of both these features. Many English people pronounce **r** as a frictionless continuant instead of as a fricative.
>
> (Jones 1932: 190)

Based on the criteria of sonority and relative prominence, Jones' classification of glides groups the subgroup of consonants frictionless continuants with English /r/.

> Frictionless continuant variants of w and j are also sometimes heard in English in place of the ordinary semi-vowels (§§ 802, 813). These variants may be indicated phonetically by writing a length-mark after the letters w and j. Thus the interjections *well* and *yes* are occasionally pronounced w:el, j:es. These are single syllables, and distinct from the groups u:el, i:es which would be disyllabic. It is weakness of exhaling force which causes continuant w: and j: to be consonantal.
>
> (Jones 1932: 190 original emphasis)

To this description of the group frictionless continuants, Jones adds the following:

> The palatal and velar frictionless continuants have the organic positions of close vowels. In fact, they are vowels, but they are uttered with very little breath-force as compared with the normally pronounced vowels which adjoin them in connected speech. These frictionless continuants are to be considered consonants on account of their consequent lack of prominence as compared with the adjoining vowels.
>
> (Jones 1932:46, note 5)

The palatal [j] and velar [w] frictionless continuants <u>are</u> vowels that, due to their 'weakness of exhaling force', lack prominence. We turn now to Jones' description of the subgroup semi-vowels where we see that they are defined by that very same initial position.

> Semi-vowels are defined as independent vowel-glides in which the speech-organs start by forming a close or fairly close vowel and immediately move to another vowel of greater prominence; the initial vowel-position is not held on for any appreciable time. It is the rapid gliding nature of these sounds, combined with the use of rather weak force of exhalation that renders them consonantal.
>
> (Jones 1932: 191)

Both the frictionless continuants and the independent semi-vowel are articulated by attaining the position of a close vowel and are rendered consonantal by their inherently weak force of exhalation. The distinction that Jones makes between these two segment

groups (frictionless continuant versus semi-vowel or vowel-glide) is not clear. The frictionless continuants have the 'organic' positions of close vowels whereas for *semi-vowels* the speech-organs start by forming a close or fairly close vowel.

To summarise, to our list of names for glides, such as, *voyelles consonnes*, *consonnes composées* (bi-labial and palatal fricatives) and *semi-voyelles* of Paul Passy, we can add Jones' semi-vowel even consonantal vowel and frictionless continuant. As for their classification, inherent lack of prominence means that glides are judged to be consonants. The phonetic criteria evolve from Passy's discussion of weak friction to Jones discussion of 'weakness of breath force' that lends to weakened sonority and prominence.

In the next sections, we continue our survey with modern phonetic treatments. In the following section we examine a small survey of the phonetic descriptions to show how the analysis of glides evolve in the discipline of phonetics over the next century.

2.3 A survey of the classifications of glides in modern phonetic treatments

The goal of our survey is to illustrate the difficulty that glides present for phonetic descriptions. We examine more broadly the phonetic treatments of glides as a subgroup of sounds that occur in many languages other than French. Throughout the phonetic tradition that categorises sounds as one of either consonant or vowel, glides remain a peculier entity that is neither and both. As a result, we observe an evolving list of terms used in their description.

The field of phonetics takes a marked turn of direction in the 20th century leading the focus of the phonetic tradition away from the 'practical applications' of second language teaching and towards the phonetic description of as yet described languages. Exposure to a much broader field of languages puts great emphasis on recording in detail these languages of the world. The International Phonetics Association continues to meet and regularly updates the IPA transcriptions as needed (most recently 2005).

We have arrived at the era in which the study of phonology is emerging and the two disciplines evolve into two separate and distinct fields of modern linguistics. They are not however mutually exclusive. According to Saussure, "[p]ar rapport aux autres voyelles, ces sons [i y u] supposent une fermeture encore considérable, assez voisine de celle des consonnes. Il en résulte certaines conséquences qui apparaîtront plus tard, et qui justifient le nom de *semi-voyelles* donné généralement à ces phonèmes." (1982: 75).

The divergent yet symbiotic relationship becomes particularly evident with the Prague School phonological tradition. This relationship is made most evident by Pike (1943). As we transition from the classical tradition of phonetics to the modern tradition we examine the glide treatments for English simply because they reflect the general application of phonetic tools to all languages being described in our modern tradition.

2.3.1 Modern Phonetic Treatments of Glides

In the following section, we examine the continuously changing descriptions of glides with a goal to illustrate the evolving treatment of glides informed by the emerging phonological tradition in the 20th century. We survey Gleason (1955), Abercrombie (1967), and Ladefoged (1993). We will use their transcriptions when we cite from their work. In such cases the palatal glide /j/ is transcribed as /y/. In these later descriptions we see how the field of phonetics is adapting to the observations made in the emerging field of phonology and to the recognition of the separate role of phonemicity.

In a description of the phonetic system of American English, Gleason describes /r y w/ as members in a group called median resonants, “[i]n / r y w/ the passage through the nose is open at the mid-line. The term resonants also includes all English vowels. Because of the close similarity of /r y w/ to the vowels, they are sometimes called “semi-vowels” (Gleason 1955: 21).

We note that Gleason retains the /r/ as part of this group of semi-vowels, as was first proposed in Jones' (1932) group called frictionless continuants. Pike (1943) later adapts this grouping of /r/ with the glides /j w/ and names them central resonant continuants. The inclusion of the English /r/ amongst a classification whose members are predominantly

glides may seem at first controversial but for many phonetic systems of English at least it is well grounded to group these sounds together.

With regards to the Modern French phonetic system it would depend on the variety but in most cases the retroflex /r/ is not among the inventory of sounds or if it is included it is allophonic. In any case the /r/ of Hexagonal French is predominantly realised as a uvular fricative [ʁ]. In the modern phonetic tradition, glides are now included in the category of median resonants. According to Gleeson, "[m]edian resonants include most vowels and various vowel-like consonants. The latter are commonly called semi-vowels. Whether median resonants are to be considered as consonants or as vowels is a matter of phonemic function in the particular language not of phonetic nature." (1955: 250).

Gleason's description of American English reflects the change in practise in phonetics with regards to adhering to an absolute dichotomy of membership for consonant or vowel status. This is a change that reflects the emerging phonological tradition and the separation of phonetic category from phonemic function. In this way, phonetic descriptions are concerned with phonetic category whereas phonemic function (phonemicity) is a matter for phonological analysis.

We turn now to Abercrombie (1967) who describes the classification of approximant. As we learned from Gleason, the resonant will have a passage of air that is open at the mid-line. Abercrombie proposes the name approximant.

> Finally there are segments made with central passage of the air-stream and open approximation of the articulators, so that no noise of friction is produced. All vowel segments are made by this type of stricture (sometimes without any actual contact of the articulators at all), and so are various segments often called consonants, [...] The rather unsatisfactory names semivowel and frictionless continuant have usually in the past been given to such segments, but, following a recent suggestion by P. Ladefoged, we shall here adopt the term approximant for them.
>
> (Abercrombie 1967:50)

Abercrombie acknowledges the longstanding problem of confusing phonetic form and phonological function inherent in the use of the terms vowel and consonant and provides his own inventory of the various terms suggested in their descriptions: "The confusion which has resulted is demonstrated by the emergence of compromise terms such as semi-

vowel, demi-vowel, semi-consonant, vocalic consonant, consonantal vowel, vowel-like, consonantoid, vocaloid, and others" (1967:80).

We turn now to Ladefoged (1993) who is a student of Abercromie but whose considered opinion is respected very early on and whose expertise is unmatched in the field of phonetics in the 20th century.

Peter Ladefoged whose work on the languages of West Africa and in the field of acoustic phonetics, as well as in association with the International Phonetics Association has had enormous influence on how we perceive phonetics and the role of phonetics inside of the field of phonology. We can attribute the most recent name for glides, that of approximant, to Ladefoged, who defines the approximant this way:

> A semivowel is a kind of approximant consisting of a nonsyllabic vowel occurring at the beginning or end of a syllable. When at the beginning of a syllable, it usually consists of a rapid glide from a high vowel position to that of the following vowel. The semivowels in English are [j] and [w], which are like nonsyllabic versions of the English high vowels [i] and [u] respectively.
>
> (Ladefoged 1993: 229)

The second half of the 20th century is dominated by phonological theory yet Ladefoged's work has brought about renewed appreciation for general phonetic description. While his work in acoustic phonetics provides precision of phonetic detail, he has also focused on comparative analysis of phonetic description of various languages which allows us to examine what is common to phonetic systems of different languages as well as what sets them apart. Ladefoged offers his insights on glides in languages like French: "[i]n other languages, there are the three high vowels [i, u, y]. In some of these languages (for example, French), there is also a semivowel corresponding to the high front rounded vowel [y]. The symbol for this sound is [ɥ], an inverted letter h" (Ladefoged 1993: 229).

Phonetic classifications have changed over time and we can now refer to a glide as one particular type of approximant as was proposed by Ladefoged. Approximant implies "a particular place of articulation" (Ladefoged 1996: 229) which is a quality more common to the classification of consonants than to the classification of vowels. Progress in the description of glides reflects our better understanding of not only glides but also the role

of phonetics in linguistic description, informed by phonological analysis. However, the term ‘semi-vowel’ continues to be prevalent. As such, the ambiguity presented by the duality of glides persists.

2.4 Summary of the naming of glides in the phonetic tradition

We have shown the extent to which understanding the nature of glides has been problematic for the field of phonetics. The naming of glides in phonetics has evolved with a deeper understanding of this group of sounds we call glides. The vowel/consonant dichotomy on which all earlier description is based, only serves to constrain descriptions since it requires that all speech sounds fall (sometimes forcibly) into one or the other of two main classes. In the early part of the 20th century with the emergence of the phonological tradition with its notion of phonemicity, we see, at last, the separation of the description of the surface form (phonetics) and the description of the function of the phoneme (phonology). Within this timeframe glides have been given a variety of names that include: *voyelles consonnes* (vowel consonants); bi-labial or palatal fricatives also called *consonnes composées* (compound consonants); *semi-voyelles* (semi-vowels) of Paul Passy, to which we can add Jones' glide; independent vowel-glide; semi-vowel; consonantal vowel as well as frictionless continuant; from Gleason; median resonants and finally from Abercrombie and Ladefoged the term approximant.

Our survey of the phonetic tradition shows that the description of glides has been problematic throughout the history of the discipline. As such, the phonetic description of glides has been in constant evolution. At least some of this difficulty is the result of the dichotomous structure on which the classification of all segments is based. Segments are classed either as vowels or consonants and, while glides share a common distribution with consonants, in phonetic quality they are more closely related to vowels. The search for that particular phonetic quality inherent to a glide, which would distinguish it from its high vowel counterpart, proves to be elusive.

We turn now to a discussion of the emerging phonological framework.

2.5 Early Phonological Tradition

Pike is neither part of the classical tradition of phonetics nor is he part of the Prague School structrualism that dominates the emergence of the phonological discipline. Nevertheless, his work in the description of language, syllable theory, and then the implications of syllable theory for a description of glides, is invaluable. Pike proposes a reorganised system for the descriptions of sound inventories informed by Prague School phonology addressing perceived problems of notation in the phonetic tradition.

2.5.1 Pike (1943)

Pike's *Phonetics* 1943, the first of the highly regarded series including *Phonemics* 1947, serves to bring together the two disciplines that are to this point distinctly separate sciences. From *Phonemics,* Pike offers this: “Phonetics gathers the raw material. Phonemics cooks it” (1947:57).

Pike recognises a tendency to confuse the description of a sound with the description of the phonological context in which that sound may appear. Given a system based on the vowel/consonant dichotomy it is not surprising thus he sets about inventorying some earlier phonetic descriptions with a view to offering an alternative. Previous descriptions are too often framed by a normative view (particularly when discussing a single language) that is then biased by a perspective that privileges the accepted norm versus the unaccepted non-norms (for that single language).

In some cases he expresses a very clear disdain for the descriptive methodologies adopted in earlier work including that of Passy and Jones which is based on a limited set of (Eurocentric) empirical data.

> Normative prejudice causes in part the selection of classification criteria which cannot be strictly applied, and leaves contradictions in the material. The most important instance is that of vowel and consonant differentiation (see Chapter V, “Classification Criteria”). [...] Rarity and frequency of occurrence of phonetic sound types as phonemes or variants of phonemes should be investigated by the phonemecist, but the phonetician is doing him no favor when he prejudges the conclusions by giving normalized classifications and suppresses the very data that would provide the phonemic choice of sounds with a contrasting setting.

(Pike 1943:11-12)

Pike disagrees with ubckuding superfluous phonetic detail that does not lead to our better understanding of the phonemes of a language (as in the case of glides as the transitions between speech-sounds, for example). He takes exception to the description of any sound as incidental (transitory or otherwise): "If one cannot have consistent criteria for determining whether or not a glide is a separate sound, or part of a preceding or a following one, there remains no chance for consistent segmentation. If phonemic criteria are used, the division is valid only for phonemic purposes" (1943: 49-50).

Pike's observation of the need for consistent criteria for distinguishing a sound marks the beginning of the discussion of the inherent relationship between phonetics and phonology and the danger of confusing the two. As we noted earlier, it appears that some of the criteria used in the description of glides need clarification. Often phonetic descriptions are confused by the function a sound might serve in syllable structure but in this case the difficulty goes in the direction of yet another extreme. The description of glides as the incidental sounds in an utterance verges on ineffectual. Of particular concern for Pike are the imprecise practises in the use of terminology.

> Passy says that "Glides are accessory sounds produced involuntarily in the articulation of given speech sounds." "Involuntarily" with him obviously does not mean "inevitably," since he discusses errors of certain foreigners who fail to pronounce some of the glides properly. The implication, then, is simply that by glides he means nondistinctive parts of some phonemic unit. Noël-Armfield conveys a similar implication when he says that, strictly, the term should "be used to express nothing more than a necessary intermediate sound between two regular speech-sounds [i. e. phonemes]."
>
> (Pike 1943:49)

We have mentioned in the previous section that some of the difficulty in describing the nature of glides seems to come from confusing their phonetic form with their phonological function. With respect to the strict adherence to the vowel/consonant dichotomy (and with respect to the work by Jones, in particular), Pike suggests that these descriptions often reflect an adherence to a phonetic dogma without principled descriptions or justifications for classifications.

> The most basic, characteristic, and universal division made in phonetic classification is that of consonant and vowel. Its delineation is one of the least satisfactory— a difficulty passed on to the instrumentalist. Frequently for descriptions of single languages the division is assumed, with no attempt to define it. The distinction is often presented as if it were clear-cut, with every sound belonging to one or the other of the groups. [...] Later, however, various sounds are mentioned [by Jones] which have to be discussed separately, under different rules, or with various kinds of reservation, because they do not neatly catalogue themselves.
>
> (Pike 1943: 66)

Simply because the dichotomy consonant/vowel works in a particular fashion in one language does not mean it necessarily holds for another. Often, however, this dichotomous relationship as it was described for the facts of a single language (namely Greek) is held up as the model for all other languages much the way the normative view would dismiss 'marginal' sounds (sounds that are more rare like whispered sounds or clicks) since they are insignificant in language (cf. Pike 1943 § Marginal Sounds). Pike then addresses the articulatory criteria used to support the distinction between vowel and consonant: "The most important reason for the difficulty at the border line between consonants and vowels is that many criteria, and criteria of conflicting status, are used.[...] Obstruction of the air stream at times constitutes one articulatory criterion for consonants. This by itself is insufficient" (1943: 66).

Let us recall that Passy's description of consonants as 'noises' is largely based on the criterion of obstruction of the air stream and thus he justifies a classification of glides as consonants based on the (untenable) observation of audible frication of glides. Pike points out that the obstruction criterion can prove to be a red herring.

> What does constitute obstruction of the kind to make consonants? If contact at the sides of the tongue, or small height of the opening (so that a small hole is left for air to escape), is such, then [i] would be a consonant, since Kenyon notes that it may have more tongue contact than [l]. There seems to be no articulatory measuring rod for degree of constriction or obstruction which marks the consonant-vowel border.
>
> (Pike 1943: 67)

Jones had of course revised this distinction in his description adding that it is sonority combined with relative prominence that ultimately distinguishes vowels and consonants and in this way he reclassifies glides as 'frictionless continuants' (when they are not

semi-vowels, of course). Pike points out that Jones will amend this statement with a footnote “in which he admits that cardinal [i] is apparently less sonorous than some other speech sounds” (Pike 1943: 69).

> If a syllabic is established acoustically by relative loudness or prominence it cannot be any innate natural sonority which always forces certain sounds to be syllabic, or else there could never be any variation between a sound which, in the same phonetic context, is sometimes syllabic and sometimes not, like English sonants [l], [r]. [m], and [n], or like nonsyllabic vowels [...]. When Jones says that “It is a consequence of this principle of relative prominence that certain short vowel-glides must be regarded as consonants,” we are immediately aware that he is using contextual function as his criterion.
>
> (Pike 1943:75)

With respect to the glide-vowel distinction Pike offers a summary of some of the very issues we have already noted.

> Jones and Kenyon use movement as a criterion for certain consonants (glides, [j], [w]) in contrast to vowels, which are said to have fixed, position; but Jones also uses movement as a criterion for certain types of vowels (diphthongs). The criterion is vitiated when he shows variants of both the glides and the diphthongs, written with the same letters (or the same letters plus a length sign), which are not quick glides but have at least part of the sound relatively fixed.
>
> (Pike 1943: 68)

Pike argues that all three of Jones’ criteria for distinguishing vowels from consonants (relative force, steady-state, and duration) are quite unreliable for this purpose and he demonstrates that Jones’ observations are really contextual and as such require reference to that context and to the qualities of other sounds that would occur in the same context. This is not good enough says Pike: [i]f the phonetician first delimits supposed articulatory classes by the phonemic features, how can he then describe the phonemes with articulatory methods?” (1943: 78).

In an effort to facilitate an unencumbered phonetic description, Pike offers alternate terminology that allows the phonetician to discuss the vowel-like and consonant-like groups of sounds without necessarily having to evoke the role that each plays in syllable structure: "[n]o other phonetic dichotomy entails so many difficulties as consonant-vowel division; articulatory and acoustic criteria are there so thoroughly entwined with

contextual and structural function and problems of segmentation that only a rigid descriptive order will separate them." (Pike 1943: 78-79).

Pike treats glides as members of the group of continuants (which includes both consonants and vowels). This is a classification that he divides into two subgroups (frictionals and non frictionals) which are themselves divided into two more sub-groups (orals and nasals). Thus, glides are part of the second classification (non frictionals) of oral continuants (cf. Pike 1943: 142):

> The sounds which as a group function most frequently as syllabics are *vocoids*. Phonetically they comprise the central resonant orals as already defined [central escape of air, [o] and [h], for example]. Vocoids include practically all sounds which are usually called "vowels" (whether voiced, voiceless, or whispered), except that "fricative vowels" are excluded while "vowel glides" such as [r], [w], and [y] are included.
>
> (Pike 1943: 143)

According to Pike, the terms consonant and vowel appear dichotomous when in fact they are not since the naming of a segment depends on the position it takes within the syllable (context) whereas the terms vocoid and contoid are absolutely dichotomous "All vocoids are nonsyllabic vocoids while functioning as nonsyllabics (e.g. [y] in 'young'; [w] in 'woo'; [r] in 'rich'). All contoids while functioning as nonsyllabics are consonants" (Pike 1943:145). To the terms vocoid and contoid Pike adds the term syllabic to denote that a segment may constitute a syllable nucleus. These terms add clarity to a treatment of glides in phonetic or phonological accounts. Glides are vocoids, above all else, nonsyllabic vocoids when in a syllable margins.

Pike and Pike elaborate further the structure of the syllable in their article on the syllable constituents of Mazateco: "[i]t is well known that sentences have internal structure which can be analyzed in terms of successive layers of immediate constituents [...] It is convenient to describe syllables of Mazateco in a similar fashion. The different layers in the syllable tend to have different phonetic and grammatical characteristics." (Pike and Pike 1947: 78).

In a detailed discription of Mazateco, Pike and Pike define a syllable structure that can be seen as a precursor to hierarchical models that follow. In Mazateco (a tone language)

syllable structure includes margins (consonant strings composed of principal and subordinate consonants) that precede nuclei (sequences of one to three vowels that, according to Pike and Pike, do not constitute units of length cf. 1947: 82). These authors refer to syllable constituents (margins and nuclei) as layered where layers constitute principal members and subordinate members. According to these authours, the relationship between members is different depending on the particular constituency. Pike and Pike elaborate further: "Mazateco is tonal, with significant lexical or grammatical pitch on each of its syllables. Many words differ in meaning with concomitant differences of pitch, but such pitch contrasts are limited exclusively to the segments here called the vowels, and the vowels are limited to the nuclei. " (Pike and Pike 1947: 79).

Pike and Pike do not specifically address glides, however, their model for syllable structure and discussions dealing with contrastivity in the language is an example of the value inherent to a description that is informed by knowledge of the phonological system the grammar.

We turn now to a discussion of the emergence of the early phonological tradition. In this section we survey in a chronological order developments in the early phonological tradition. Our survey begins from its emergence in the work of Saussure to the key concepts put forth by Prague School phonologists. Some key concepts elaborated in Prague School (distinctive features, for example) have implications for the analysis of glide phenomena in French that are made evident in the generative framework introduced later by Chomsky and Halle in their formative work *Sound Patterns of English* (SPE hereafter).

2.6 Phonological Tradition

The particular difficulty for a phonetic description of glides in French does not go unnoticed by one of the foremost pioneers in phonological theory. In the preceding sections we have discussed the extent to which the phonetic description of glides (as distinct from their corresponding high vowels) is problematic when based entirely on the acoustic qualities of a sound. The most prevalent opinion is that a glide is consonantal but

defining the acoustic cue for consonantal status is not obvious. In phonological treatments, the absence of an acoustic cue to point to the consonantal status of a glide is less a problem since phonologists look instead to other, more structural, factors.

In the next section, we examine some of Saussure's observations concerning the duality of glides.

2.6.1 Saussure and Structuralism

Saussure makes several observations about the nature of high vowels and their corresponding glides in French:

> Enfin les voyelles de quatrième aperture [voyelles hautes] donnent lieu à certaines observations. Nous avons vu p. 81 que, contrairement à ce que l'on constate pour d'autres sons, l'usage a consacré pour ceux-là une double graphie ($w = \overset{<}{u}$, $u = u^{>}$; $y = {}^{<}i$, $i = i^{>}$). C'est que dans des groupes tels que *aiya*, *auwa* on perçoit, mieux que partout ailleurs, la distinction marquée par < et >; $i^{>}$ et $u^{>}$ donnent nettement l'impression de voyelles, $\overset{<}{i}$ et $\overset{<}{u}$ celle de consonnes. Sans prétendre expliquer ce fait, nous observons que ce *i* consonne n'existe jamais sous l'aspect fermant. Ainsi on peut avoir un *ai* dont l'*i*ʾ fasse le même effet que le *y* dans *aiya* (comparez l'anglais *boy* avec le français *pied*) ; c'est donc par position que *y* est consonne et *i* voyelle, puisque ces variétés de l'espèce *I* ne peuvent pas se manifester partout également. Les mêmes remarques s'appliqueraient à *u* et *w*, *ü* et *ẅ*. [7]
>
> Saussure (1982: 91)

Saussure proposes that a clear distinction between high vowels and their corresponding glides be made on the basis of their distribution. In this way, we can consider that the segments [i] and [j] together comprise *une espèce I* (similarly *l'espèce W* and *l'espèce Ü,* 1982: 66) thus it is not on the basis of acoustic quality alone but by their different distribution in groups of sounds (syllables) that a glide is distinguished from its corresponding high vowel. This position can be understood today as allophones of a single phoneme, a concept that is central to any phonological model and well adapted to the glides of French as we will see in our later discussions (cf. Chapter 3-4).

---

[7] Saussure uses the following conventions: < = consonant-like and > = vowel-like.

Rather than resolve any longstanding debate about phonetic quality of a glide, the discussion shifts to the phonological description (distribution and function). The most common position for French is the one wherein all glides are considered the non syllabic realisation of underlying high vowels from which they derive their identical phonetic composition.

Prague School phonologists, undoubtedly influenced by Saussure's theory of structuralism in linguistics, elaborate many of his notions within the Prague School theories:

> Saussure's achievement may be characterized best by enumerating a number of dichotomies which he established between different aspects of language, viz. langue/parole, signifié/signifiant [...], form/substance, syntagmatic/associative (later termed paradigmatic), synchrony/diachrony. Taken separately these dichotomies are not new. They can be traced back to various of Saussure's predecessors, but Saussure combined them into a coherent theory, and within these distinctions he often stressed aspects which had until then been neglected: i.e. langue, as opposed to parole; form, as opposed to substance; synchrony as opposed to diachrony.
>
> (Fischer-Jørgensen 1975:11)

After a general discussion of the Prague School we turn our focus to the work of Martinet whose work in French phonology and glides, in particular, is relevant here. Martinet brings a very interesting perspective informed by diachronic as well as synchronic studies of sound change to the analysis of French phonology. His studies of the regular relationships (generalisations) between sounds are enhanced by his empirical studies. His empirical studies are motivated by the desire to offer a more reliable description of the phonological system of French in the twentieth century.

2.7 Prague School

> The Prague School did not want simple [sic] to introduce new points of view in phonetics but to create an entirely new discipline – phonology – which was to be independent of phonetics.
>
> (Fischer-Jørgensen 1975:22)

As the name suggests, Prague School phonologists come out of the European linguistic tradition. Influenced by the work of Saussure, Prague School phonologists take the study

of language beyond the phonetic tradition of describing the acoustic qualities and correct pronunciation of the inventory of sounds in a particular language to the function of these sounds.

2.7.1 Nikolai Trubetzkoy and Roman Jakobson

Since a principal goal of the Prague School is to develop phonology as a discipline of linguistics independent of phonetics, both aspects of language (concrete and abstract) are central to the discipline of phonology that it proposes. The phoneme, the abstract representation of a distinctive speech sound, and not the phone is the core principle to their analyses. Ironically, the further we go towards a more defined phonological theory, the more it appears that phonetics and phonology are not so easily separated. Martinet, in particular, understood this, describing phonology as "functional phonetics" (below). Perhaps most notable is Trubetzkoy's theory of the phoneme as the minimal distinctive unit. His work on the phoneme and the archiphoneme leads to the elaboration of the notion of Markedness (Greenberg 1966a). Further to the definition of the phoneme as the smallest phonological unit, Jakobson proposes a theory of distinctive features of phonemes (Jakobson, Fant and Halle 1952, Jakobson and Halle 1956).

> This formulation is due to Jakobson, who as early as 1932 in an article in a Czech encyclopedia defined the phoneme as "a set of those concurrent sound properties which are used in a given language to distinguish word meanings". It is only a very short step from this formulation to the conception of the phoneme as consisting of components such as voicing, labiality, etc., whereby these components, and not the phonemes, become the smallest phonological units"
> (Fischer-Jørgensen 1975:145)

Jakobson's distinctive features (cf. Jakobson and Halle 1956: 29–31) reflect the fine detail of the phonetic observations that preceded. Jakobson's formulation of distinctive features internal to segments (cf. Jakobson and Halle 1956: 29-31) plays an important role in the categorisation of segments particularly in the generative framework (Dell 1973, Schane 1968), for example).

With respect to the analysis of French phonology, and glides specifically, we consider the work of André Martinet whose observations in phonetics and phonology (both diachronic and synchronic) focus on but are not strictly limited to French. His contributions,

particularly in matters of language variation and change, not only build on the foundations laid by the other Prague phonologists but "have also influenced considerably the Prague School in the post-war period and have been integrated in their own view of sound change" (Fischer-Jørgensen 1975:47). According to Fischer-Jørgensen, Martinet's contributions to Prague school phonology can be summarised by these four main ideas: the role of functional load; the tendency for minimal phonetic distance; the notion of harmonious systems as economy; and the role of physiological factors in asymmetry versus symmetry.

2.7.2 André Martinet

Martinet advocates the study of phonetic change (diachronic and synchronic) and variability with a goal to understand its significance in phonological systems; "[t]he aim of phonological analysis is to identify the phonic elements of a language and to classify them according to their function in that language," Martinet (1960: 53). He points to the example of sudies of French in particular. While in the first part of the twentieth century the discipline of phonology is still quite young, Martinet observes that such a well-documented language as French should provide an ideal situation for elaborate phonological descriptions. Still, very few descriptions of French existed at that time and of those few offer more than "un exposé dogmatique" (Martinet 1945:1) perpetuating the myth of French as a homogenous language.

Martinet's passion for such a description (an accurate description of linguistic variation and change in Modern French) is ignited in the prospects of a Phonological Atlas of European languages as is proposed by Trubetzkoy and Jakobson to the members of the l'*Association internationale pour l'études phonologiques* (cf. Martinet 1971:5). However, the political turmoil of WWII would present a less than obvious opportunity for Martinet to correct this situation. Martinet, while detained in a German prisoner of war camp in 1940-41, conducts what would be the first of several empirical studies. The survey used for this study is based on the work that was done in preparation for the French contribution to the Phonological Atlas Project.

> [...] fait prisonnier le 12 juin 1940, le secrétaire de la Société française de phonologie a passé quinze mois de sa captivité dans un camp d'officiers, dans un milieu où, en général, la langue commune n'est pas l'idiome second, comme elle le serait le plus souvent parmi des ruraux ; mais la langue unique apprise dès l'enfance de la bouche des parents ou des camarades de jeu. On a donc pu tenter une enquête visant à déterminer les particularités phonologiques des différents sujets de langue française qu'on pouvait approcher.
>
> (Martinet 1945: 6)

While imprisoned, Martinet collects 409 completed questionnaires. The results show a much greater degree of variability than even he might have expected given the homogeneity of this group of respondents who share a similar education and social class but hail from a variety of geographical regions:

> Parmi les officiers, il n'en est certainement aucun à ranger ni dans le prolétariat urbain, ni parmi le petit peuple campagnard [...] De façon générale, du fait du mode de recrutement des officiers, on peut estimer que la presque totalité de nos sujets a poursuivi des études sanctionnées soit par le baccalauréat, soit par le brevet supérieur [some exceptions cf. note 1]. Nous avons donc affaire à un milieu socialement assez homogène.
>
> (Martinet 1945:19)

The survey consists of 45 questions that focus on some of the most recognised aspects of French phonology. Many of the questions are designed to treat various classical phonological oppositions. Respondents are asked to indicate their usual pronunciation for a given opposition (as observed in the pair *brin~brun* [ɛ̃]~[œ̃], for example). Question 41, "Prononcez-vous en une ou deux syllabes : a) *pied* ? ... b) *lion* ?... c) *Riom* ? d) *bouée* ? ... e) *buée* ? ... f) *louer* ? ..." (Martinet 1971:13), examines potential variation (syneresis versus dieresis) in realisations of HVV contexts, such as in *bouée* pronounced [bu.e] or [bɥe]. Although, the realisation of glide versus high vowel not distinctive, it is of interest nonetheless, as Martinet notes "[l]a question des semi-voyelles est une des plus débattues de la phonétique générale" (1945:175).

While this quasi-opposition ([i] ~ [j]) may not be a phonologically significant opposition, as Martinet stipulates "un [i] non syllabique ou un [y] constrictif ne sont pas nécessairement des réalisations d'un autre phonème que [i]" (1945:175, note that [y] is [j]). The distinction (consonantal versus vocalic) however is not likely related to any inherent phonetic quality like closure or constriction.

The results from question 41 are surprising even for Martinet:

> On ne doutait pas, en effet, que la prononciation monosyllabique fût seul possible pour *pied*. Or, 23 sujets ont déclaré prononcer le mot en deux syllabes (12 pour Fnm. [français non-méridional] et 11 pour le Midi), c'est-à-dire plus de 5%, ce qui, à la lumière de ce que nous avons vu jusqu'ici, paraît bien considérable comme pourcentage d'erreur. Ce qui est surtout troublant est que, dans le Nord par exemple, 4 sujets sur 19 déclarent prononcer *pied* en deux syllabes, ce qui donne un pourcentage de 21 qui n'est pas négligeable. [...] Nous sommes donc tenté d'admettre qu'il existe dans le Nord, et probablement aussi dans le Midi (11 % avec bon échantillonnage), une tendance, sinon de faire de *pied* deux syllabes avec une prononciation du type *pi-yé*, du moins à donner à *i* une valeur assez vocalique pour que plusieurs sujets soient tentés d'y voir un sommet syllabique.
>
> (Martinet 1945: 177)

The degree of variability in a word like *pied* (assumed to be lexicalised as a monosyllabic) is significant if for no other reason than to document the real state of variability in the pronunciation of HV+V sequences. "Il s'agit simplement d'étudier l'extension et la répartition de deux types de réalisation, vocalique et consonantique, des phonèmes i, u et ü lorsqu'ils se trouvent devant voyelle" (Martinet 1945: 176). Results from question 41 show that this variability is not isolated to the word *pied*; the other words examined demonstrate a comparable degree of variability:

> [E]nfin les chiffres tout à fait analogues pour *bouée* et *louer* : pour l'ensemble de la France, les pourcentages respectifs sont 63,4 et 64. La suture morphologique entre *lou-* et *-er* ne semble donc guère influencer la prononciation, ce qui tend à prouver l'instabilité des oppositions du type *loua/loi* proposées par M. Gougenheim.
>
> (Martinet 1945: 177)

Some thirty years after this first study and in collaboration with Henriette Walter, Martinet conducts another survey (Martinet and Walter 1973). This survey involves 17 participants and an extensive number of newly surveyed words (10, 000). The survey by Martinet and Walter results in a dictionary of the pronunciation of Modern French that reflects actual pronunciation as well as the degree of variation observed in actual practice for a subset of the French lexicon, setting this dictionary apart from other classic dictionaries of French pronunciation. The results of their survey (cf chapter 1) challenge "[...] la conviction qu'à de rares exceptions près, chaque mot a une prononciation correcte [...]" (Martinet and Walter 1973: 16).

Martinet and Walter summarise their findings for the high vowels and their corresponding glides as follows:

> Il existe un phonème /i/, un phonème /y/ et un phonème /u/. En finale de syllabe, on distingue /j/ de /i/ (/pɛj/~ /pɛi/, *paye ~ pays*), mais [ɥ] et [w] peuvent être considérés respectivement comme la forme qui peuvent, devant une autre voyelle, les phonèmes /y/ et /u/ : /bye/ *buée* se réalise comme [bɥe] ou comme [by-e], *bouée* comme [bwe] ou comme [bu-e].
>
> (Martinet and Walter 1973: 32)

These findings show that the strongest tendency is for high vowels in hiatus to be realised as a corresponding glide. Sometimes hiatus prevails, particularly in the case of rare words such as *Riom*.

In view of the role of the phoneme and distinctive features, and yet acutely aware of the variability attested, Martinet has this to say about the phonetic characteristics and phonemic status of glides in French:

> Ces caractères, consonant ou constrictif, ou l'ensemble de ces deux caractères, ne prendront la valeur de traits pertinents que pour autant qu'ils permettront dans un état de langue donné de distinguer entre des quasi-homonymes : s'il existe en français y distinct de i, c'est que *paye* [pèy] ne se confond pas avec *pays* [pèi].
>
> (Martinet 1971:175)

Martinet (1945, 1971) and Martinet and Walter (1973) capture, with empirical evidence, the variable nature of glide phenomena in French. GF proves variable even for an individual speaker and this variation may also be conditioned by the particular high vowel involved (i » u » ɥ). The characteristic that distinguishes a glide from a corresponding high vowel (constriction versus distribution) cannot alone account for the application or non-application of GF.

We turn now to the next era in phonological analysis and glide treatments: the generative framework. The Generative model is built solidly on a foundation of notions that are of primary importance to the Prague School of phonology: the phoneme, distinctive features and the work of Martinet with respect to phoneme inventory and in particular the notion of harmonious systems as economy. We focus our discussion on those aspects of Generative phonology germane to an account of the French Glide data.

Chapter 3

3 Linear Generative treatments of glides

The work of Prague School linguists lays the foundation for the generative framework which dominated the later half of the 20th century but it is specifically the notion of Distinctive Features (Jakobson, Fant and Halle 1952, Jakobson and Halle 1956) that is at the core of the phonological level of representation in this model that will be central to the treatment of glides.

In this chapter, we discuss briefly the Generative model as presented by Chomsky and Halle in their work Sound Pattern of English (hereafter SPE). For a detailed look at French glide data we turn to Schane (1968) who, as a contemporary of Chomsky and Halle, provides a detailed analysis for French data in the generative model. Our discussion of treatments in the generative model for French data concludes with Dell (1973) who demonstrates the transformational cycle of this model for glide data. Finally, to round out the discussion of French glide treatments in the Generative model, we look at two articles by Morin (1976) and Lyche (1979), since these two analyses bring new data that prompt some important questions for the discussion of glides in French.

In the sections that follow, we examine French glide data in much more detail in the work of Schane (1968) and Dell (1973) who use the generative framework to treat other French glide data.

3.1 The Sound Pattern of English (SPE): the framework

> The transformational cycle operates within word boundaries in a much more far reaching and extensive way [...]. In complex derivational forms, for example, it seems quite natural to suppose that the phonetic shape of the full form is determined by general rule from the ideal representation of its parts in much the same way as in syntactic constructions.
>
> (Chomsky and Halle 1968: 27)

While Chomsky and Halle (1968) is recognised as the first complete study in the generative framework, Schane's (1965) thesis, unpublished at that time, offers a complete

study of French phonology in the generative phonological framework. Chomsky and Halle (1968) and Schane build on the work of the Prague School, notably Jakobson's theory of distinctive features, and bring together the notions defined by Prague School phonologists into a single framework. In their evaluation of the framework proposed (cf. Chapter 8), these authors apply their model to various phonological phenomena in other language data and in doing so account for a limited sample of French data from Schane (1965).

We begin with the phonetic representation of English glides. Chomsky and Halle adopt distinctive features as "the minimal elements of which phonetic, lexical, and phonological transcriptions are composed, by combination and concatenation" (1968: 64). In this framework Chomsky and Halle must stipulate that "[a]s classificatory devices features are binary" (1968: 65) thus features are marked as + to denote the presence or – to denote the absence of a given feature though they also note that this is "[a]s a first approximation [...] since phonetic features are generally multivalued" (Chomsky and Halle 1968: 65). By multivalued we understand scalar (for more cf. Chomsky and Halle 1968: 164).

The glides of English are distinguished as a major class distinguished by their feature value (– cons, – voc). The other major classes are described as follows: vowels (– cons, + voc), nonnasal consonants (+ cons, – voc) and liquids (+ cons, + voc) as in (35) below. As we have seen in several of the phonetic descriptions that preceded, the voiceless glottal fricative h and the voiceless glottal stop ʔ are treated as glides (separately below but later all are grouped together). This grouping will be debated. However, it can be considered to be inconsequential to a discussion of French since glottal stops exist only as a result of coarticulations and the *h aspiré* (with no concrete phonetic manifestation) often patterns with consonants in the liaison and elision data presented below.

(35) Glides as a Major Class[8]

	sonorant	consonantal	vocalic
voiced vowels	+	–	+
glides (I): *w, j*	+	–	–
glides (II): *h, ʔ*	+	–	–
liquids	+	+	+
nasal consonants	+	+	–
nonnasal consonants	–	+	–

(adapted from Table 1, Chomsky and Halle 1968: 303)

As a major class glides (of English) are distinguishable from one another by their place of articulation features (36).

(36) Glide features in SPE

	high	low	back
j	+	–	–
w	+	–	+
h, ʔ	–	+	–

(Chomsky and Halle 1968: 307)

Given our survey of the various phonetic descriptions for glides (cf. Chapter 2 § 2.1-2.4), we might wonder what (if any) is the phonetic basis on which the features (– cons, –voc, as in (36) above) are attributed to glides. Chomsky and Halle describe these criteria as follows:

> Vocalic sounds are produced with an *oral* cavity in which the most radical constriction does not exceed that found in the high vowels [i] and [u] and with vocal cords that are positioned so as to allow spontaneous voicing; in producing nonvocalic sounds one or both of these conditions are not satisfied. Vocalic sounds, therefore, are the voiced vowels and liquids, whereas glides, nasal consonants, and obstruents, as well as voiceless vowels and liquids, are nonvocalic. Consonantal sounds are produced with a radical obstruction in the midsagittal region of the vocal tract; nonconsonantal sounds are produced without such an obstruction. [note 7: Recent work indicates that in place of "vocalicness" the phonetic framework should contain a feature of "syllabicity" —see Chapter Eight, pages 353-55.]
>
> (Chomsky and Halle 1968:302)

[8] Note: throughout this thesis transcriptions are given in Americanist phonetic notation in the original: [y], [ü], [ẅ], are given here in IPA [j], [ɥ] and [w].

By their own admission in note 7 above, these feature classifications may be insufficient. From the feature description above, glides (–voc) must exceed the most radical constriction found in the high vowels since there is no question of their spontaneous voicing. Yet, as non consonantal sounds (– cons), they must be produced without “a radical obstruction”. Based on the articulation described, we could just as easily be describing liquids, yet, liquids are assigned the features (+ voc, + cons). In fact, but for their different distributions in syllable constituencies (glides are never found in syllable peaks), glides and liquids are much more similar than they are distinct (as observed in Chapter two several phonetic descriptions for English include /r/ as part of the glide inventory). As for the issue of constriction, acoustic studies have found that there is no significant difference in the constriction of glides as compared to their corresponding high vowels (Straka 1964, Ladefoged 1993, Levi 2004).

Chomsky and Halle stipulate the following features for the high front vowel [i] and the high front glide [j]. [i] is vocalic while [j] is neither vocalic nor consonantal.

(37) Feature Matrices in SPE

$$\begin{array}{cc} [\text{i}] & [\text{j}] \\ \begin{bmatrix} +\text{voc} \\ -\text{cons} \\ +\text{high} \\ -\text{back} \end{bmatrix} & \begin{bmatrix} -\text{voc} \\ -\text{cons} \\ +\text{high} \\ -\text{back} \end{bmatrix} \end{array}$$

(Chomsky and Halle 1968: 176)

The status [– voc, – cons] for glides proves problematic for French liaison and elision data since glides can be shown to participate sometimes as a consonant (*les yaourts* ‘yogurts’) and sometimes as as a vowel (*les oiseaux* ‘birds’). Chomsky and Halle demonstrate the adaptability of the SPE framework through a demonstration of the accesibility of individual features.

Based on data from Schane (1965), these authors propose that transformation by rule can account for French liaison and elision data: /peti#livr/ *petit livre* but /pətit#ami/ *petit ami,* or /lə#livr/ *le livre* but *l(e)’enfant* for elision (for the complete data set used cf. Chomsky and Halle (1968: 353)).

Following Schane, their rules reflect the observation that "vowels are truncated before vowels and glides (38a.) and consonants are truncated before consonants and liquids (38b.); glides and liquids, on the other hand are not truncated" (1968: 352). These data may be accounted for using a simple transformational rule (38):

(38) Elision rule

a.
$$\begin{bmatrix} +\text{voc} \\ -\text{cons} \end{bmatrix} \longrightarrow \varnothing \;/\; \_\_\_\_\#\,[-\text{cons.}]$$

b.
$$\begin{bmatrix} -\text{voc} \\ +\text{cons} \end{bmatrix} \longrightarrow \varnothing \;/\; \_\_\_\_\#\,[+\text{cons.}]$$

(Chomsky and Halle 1968:353, cf. (63))

Chomsky and Halle adopt the use of variables as coefficients of features based on Schane's analysis. Using alpha notation for variable coefficients of features as shown in Schane (1965), the rule in (38) above can be made more efficient through the reformulation in (39).

(39) Elision rule

$$\begin{bmatrix} -\alpha\ \text{voc} \\ \alpha\ \text{cons} \end{bmatrix} \longrightarrow \varnothing \;/\; \_\_\_\_\#\,[\alpha\ \text{cons}]$$

(Chomsky and Halle 1968:353 cf. (63))

A single rule such as in (39) to account for the French liaison and elision data is very elegant indeed, however, glide data are more variable than this rule allows (as in Heap et al. 1992). In an excursus, Chomsky and Halle note that Milner and Bailey observe that liaison and elision does not occur before borrowings such as *le yogi* and *les yogi* without liaison, for example. Chomsky and Halle propose "a diacritic category 'foreign' and restrict the rule to occur before words which are [– foreign]" (1968:353). This ad hoc solution is not entirely satisfactory since it requires two new rules (cf. (66) p. 354) that cannot be reduced. In order to capture the symmetry of the two processes, Chomsky and Halle opt for the addition of the feature ± syllabic, as proposed by Milner and Bailey, "[w]hen vowels become nonsyllabic, they turn into glides: high vowels turn into the high glides [w] and [y]; nonhigh vowels into nonhigh glides symbolized by [h]" Chomsky and Halle (1968:354).

Resorting to the feature [± syll] in the SPE signals a first step in acknowledging the role of the syllable in a phonological representation. Until this time, while rules have included word boundaries, there is no syllable structure mechanism in place in this model.

Given the addition of the feature ± syllabic, Chomsky and Halle propose the following amendment to the major class features to reflect the adoption of the syllabic feature (40).

(40) Major Class features (integration of the feature syllabic)

	Sonorant	syllabic	consonantal
vowels	+	+	–
syllabic liquids	+	+	+
syllabic nasals	+	+	+
nonsyllabic liquids	+	–	+
syllabic nasals	+	–	+
glides: *w, j, h, ʔ*	+	–	–
obstruents	–	–	+

(Chomsky and Halle 1968: 354)

We note that two classes of glides are merged as one. Thus the class now includes both high [w, j] and [*h, ʔ*] glides. The feature ± syll. results in the revised elision rule 2 shown in (41) which serves to replace those shown in (38-39) above.

(41) Elision rule 2

$$\begin{bmatrix} +\text{son} \\ +\text{syll} \\ -\text{cons} \end{bmatrix} \longrightarrow \varnothing \;/\; \_\_\_\_ \,\#\, \left\{ \begin{matrix} \begin{bmatrix} +\text{son} \\ +\text{syll} \\ -\text{cons} \end{bmatrix} \\ \begin{bmatrix} +\text{son} \\ -\text{syll} \\ -\text{cons} \end{bmatrix} \end{matrix} \right\}$$

$$\begin{bmatrix} -\text{son} \\ -\text{syll} \\ +\text{cons} \end{bmatrix} \longrightarrow \varnothing \;/\; \_\_\_\_ \,\#\, \left\{ \begin{matrix} \begin{bmatrix} -\text{son} \\ -\text{syll} \\ +\text{cons} \end{bmatrix} \\ \begin{bmatrix} +\text{son} \\ -\text{syll} \\ +\text{cons} \end{bmatrix} \end{matrix} \right\}$$

(Chomsky and Halle 1968: 354, cf. (68))

Note however that these rules do not reduce because of the variable feature values for sonorant consonants (± son). Chomsky and Halle opt for the ad hoc ±foreign diacritic. Thus with the alpha coefficients of features notation adopted from Schane (1965), the rules reduce to the two rules in (42) in order to account for variable data involving foreign words in French.

(42)

a. French words

$$\begin{bmatrix} \alpha\ \text{syll} \\ \alpha\ \text{cons} \end{bmatrix} \longrightarrow \varnothing \ / \ \_\_\_\_\ \# \begin{bmatrix} \alpha\ \text{cons} \\ -\ \text{foreign} \end{bmatrix}$$

b. Foreign words

$$\begin{bmatrix} \alpha\ \text{syll} \\ \alpha\ \text{cons} \end{bmatrix} \longrightarrow \varnothing \ / \ \_\_\_\_\ \# \begin{bmatrix} \alpha\ \text{cons} \\ +\ \text{foreign} \end{bmatrix}$$

(Chomsky and Halle 1968:355)

This is the extent of the discussion of French glide data in SPE but there is a discussion of glide phenomena in the plural forms of Kasem that demonstrate Glide Formation in a particular context such that we observe the singular and plural pair [buga]~[bwi] river(s): "To account for the appearance of the glide [w] in [bwi], we need, in addition to a rule of Velar Elision, the Glide Rule (80), which turns a high vowel into its cognate glide before another vowel" (Chomsky and Halle 1968: 359). We have reproduced the glide rule in (43) below.

(43) Glide Rule

$$\begin{bmatrix} -\ \text{cons} \\ +\ \text{high} \end{bmatrix} \longrightarrow [-\ \text{voc}\ ] \ / \ \_\_\_\_ \left\{ \begin{matrix} \# \\ + \end{matrix} \right\} \begin{bmatrix} +\ \text{voc} \\ -\ \text{cons} \end{bmatrix}$$

(adapted from Chomsky and Halle 1968:359)

Though not stipulated for Kasem, we may also want to include the context across morpheme boundary and even across word boundary with the inclusion of the symbols (+) and (#) shown in (43) above.

This rule enables us to account for data for word-medial Glide Formation from Schane (1968) shown in (44) and discussed in section 1.2 from Martinet and Walter.

(44) Glide Formation (GF) in French verb inflections

a. HV+V→ GV

scie	[si]	'she saws'	[si] + [ɔ̃] → [sjɔ̃]	*scions*	'we saw'
tue	[ty]	'she kills'	[ty] + [ɔ̃] → [tɥɔ̃]	*tuons*	'we kill'
joue	[ʒu]	'she plays'	[ʒu] + [ɔ̃] → [ʒwɔ̃]	*jouons*	'we play'

b. HV+V→ VGV

troue	[tru]	'she punctures'	[tru] + [ɔ̃] → [tru(w)ɔ̃]	*trouons*	'we puncture'
crie	[kri]	'she cries'	[kri] + [ɔ̃] → [krijɔ̃]	*crions*	'we cry'

(Schane 1968: 56)

Chomsky and Halle have only touched on the implications for glides in French phonology. This is certainly not the whole story for glides as we will see in the next section but it should be noted that the discussion for the inclusion of the feature ± syllabic will be crucial to the representation of French glide data with the added benefit of eliminating the ad hoc feature ± foreign. We turn now to Schane 1968, the first account of the phonology of French in the generative framework.

3.2 Schane (1968)

Schane (1968), the publication of his 1965 doctoral thesis, is an extensive study of French phonology presented in the generative framework. Discussions of French liaison and elision data in Chomsky and Halle (1968) come from Schane's (1965) thesis. Given issues surrounding the variability of French glide data, we prefer to begin with Chomsky and Halle (1968) as an initial presentation of these data, introduction of the feature ± syllabic and the apparent need for the ad hoc diacritic ± foreign in order to account for glide variability.

In the sections to follow, we discuss implications of the generative framework and the transformational cycle for French glide data as discussed in Schane (1968) and Dell (1973).

3.2.1 Vowel/glide alternations and Glide Formation (GF)

Beyond the interplay observed at the segmental level for an account of liaison and elision data, Schane turns the discussion to the implications for stress in the process of vowel deletion:

> Stress at the word level is predictable and the rule for its assignment is well known: *the final syllable bears the stress unless its vowel is a schwa, in which case the stress falls on the penultimate*. When the stress rule is applied to the example, the word *joli* will receive stress on the final vowel; and since the vowel bears a stress, it will not undergo deletion. The rules for stress assignment and truncation must apply in this order; otherwise, the final *i* of *joli* would be elided.
>
> (Schane 1968: 10, original emphasis)

Schane proposes a revised truncation rule to account for the effect of the primary stress in the nonapplication of truncation on word-final vowels. A slightly modified version of this rule is given in (45).

(45) Revised Truncation Rule

$$\begin{bmatrix} \alpha \text{ cons} \\ -\alpha \text{ voc} \\ -\text{stress} \end{bmatrix} \longrightarrow \varnothing \;/\; \_\_\_\_\# \,[\alpha \text{ cons}]$$

(adapted from Schane 1968: 10)

The position of primary stress plays a role in historical vowel changes in French (cf. Chapter 1.3), and in a classic generative approach, Schane demonstrates that many historical changes to the French vowel system can be accounted for through synchronic transformations by rule operations. One of those proposed is a rule for diphthongization which operates on the resulting segments after operation of the rule for pretonic vowel raising (1968: 23). Schane points to the many vowel alternations observed in the French lexicon. Of particular interest, he points to the alternations such as pretonic /e/ which alternates with /wa/ in tonic position as in *espérance*: *espoir*, or /wa/ in tonic position alternates with /ə/ in pretonic position *doivent*: *devons* and pretonic /ɛ/ alternates with /jɛ/ in tonic position as in *bénir*: *bien*, tonic /jɛ/ alternates /ə/ in pretonic position in *viennent*: *venons* since this analysis proposes to account for the historical diphthongs of the French vowel system through crucially ordered rules to adjust (raise) then diphthongise the underlying vowel /ɛ/. Schane shows that although the pretonic vowels are identical, the

resulting alternations are not. We require a different underlying vowel from which we can derive the other diphthong:

> This vowel should be similar to /e/ in pretonic position, and it is desirable for simplicity that the conversion entail the least number of feature changes. The obvious vowel in this case is /ɛ/. A rule is therefore required which raises /ɛ/ to /e/ in pretonic position.
>
> (Schane 1968:23)

Once /ɛ/ is raised (an operation that changes the feature +low to –low) "underlying /e/ and /ɛ/ undergo diphthongization: the former will become /wa/ [after an additional rule that would adjust e→ a], the latter /jɛ/. We shall consider diphthongization to be the insertion of a glide before the vowel" Schane (1968:23).

(46) Rule for Diphthongization
Insert a glide before a stressed nonhigh front vowel
The [–low] vowel /e/ takes on a /w/ glide, whereas the [+low] vowel /ɛ/ takes on a /j/ glide.

(Schane 1968:23)

After establishing the minimal underlying vowel inventory from which he derives all other vowels, nasal vowels and vowel alternations, Schane sets out to account for all other instances of glides.

3.2.2 Glide Formation in French Verb Inflection

As we discussed in the preceding chapter, we observe GF in some French verb inflections. Verbs with a stem final high vowel may undergo GF when a vowel initial inflectional morpheme is added as in the data of (cf. discussion in section 3.1 of the glide rule for Kasem (44) reproduced here with primary stress indicated (47).

(47) Glide Formation in French verb inflections

scie	[sí]	'she saws'	/sí + õ /→ [sjố]	*scions*	'we saw'
tue	[tý]	'she kills'	/tý + õ / → [tɥố]	*tuons*	'we kill'
joue	[ʒú]	'she plays'	/ʒú + õ / → [ʒwố]	*jouons*	'we play'

(Schane 1968:56)

GF, sensitive to primary stress placement, shows that only unstressed high vowels will undergo the transformation. We take the singular form, which would bear the primary stress on the high vowel, as the root form. With the addition of the vowel initial inflectional morpheme – ɔ̃, the primary stress shifts to the vowel of the inflectional morpheme /sˈi + ɔ̃ / → /sjˈɔ̃ / since it is that vowel which is now the word-final syllable. We stipulate that, while it is the case for the verb morphology data of (47), GF does not require that the high vowel, which undergoes transformation, be pretonic. Schane proposes the following rule for GF.

(48) Rule for Glide Formation

$$\begin{bmatrix} -\text{cons} \\ +\text{voc} \\ -\text{stress} \\ +\text{high} \end{bmatrix} \longrightarrow [-\text{voc}] \;\Big/\; \_\_\_\_ \begin{bmatrix} -\text{cons} \\ +\text{voc} \end{bmatrix}$$

(Schane 1968:57)

All other instances of prevocalic glides are accounted for in this way:

> What appears phonetically as a prevocalic glide has as its origin either a lax front vowel which diphthongizes (/e/ → /we/, /ɛ/ → /jɛ/), or else a high vowel (/I/ in the case of /j/, /O/ in the case of /w/, and /U/ in the case of /ɥ/) which precedes another vowel.
>
> (Schane 1968:59)

Finally, we discuss the postvocalic /j/ and the /j/ of the imperfect tense and the subjunctive. We will not go into the details of the yod of the imperfect and subjunctive verb forms except to say that the yod is derived from the pretonic vowel: /ɛ/→ /j/ (for more details cf. Schane 1968:76).

In consideration of the postvocalic yod, this is the glide that is most often considered an underlying segment because of minimal pairs like *abbaye* [abei] ~ *abeille* [abɛj] and *pays* [pei] ~ *paye* [pɛj] (Martinet 1960, Léon 1966). Recall that according to Martinet "If it is necessary, in French, to distinguish a phoneme /j/ from a phoneme /i/, this is because *paye* and *abeille* are not identical to *pays* and *abbaye*" (1960:72). Schane elaborates further:

> The problem is, of course, that if the postvocalic glide is also derived from a vowel, then, for example *oui* and *houille* would be indistinguishable since both would be represented as /OI/ (the /O/ is raised to /u/). One must conclude either that *i* and *j* are distinct underlying segments, or that postvocalic /j/ has a different origin from prevocalic /j/.
>
> (Schane 1968:57-58)

Schane builds a case for the latter option. Based on alternating forms such as veulent/ veuillent, he proposes an l/j alternation. To avoid indiscriminate yod formation this particular /l/ must be distinct from all other instances of /l/. Since the resulting segment is a palatal glide /j/, Schane postulates that the logical intermediate step is to derive a palatal /l/ from the existing dental /l/ through the rule for l-palatalisation: /l/ → /ʎ/, which will then undergo the rule for palatal glide conversion: /ʎ / → /j/ (Schane 1968: 58-59). The intermediary stage that Schane proposes is, in fact, a variable realisation that remains in some regional varieties. Schane has demonstrated that French glides need not be part of the phonemic inventory since all glides can be derived by transformational rules.

We turn now to Dell 1973, whose work on the generative model provides some important insights for its application to French, in particular, the significance of features (±syllabic) and ordered rule operations in the derivation of glides in French.

3.3 Dell (1973)

> *Sont linguistiquement pertinentes toutes les propriétés d'un énoncé qui sont gouvernées par des règles.*
>
> (Dell 1973: 46)

Dell provides an elaborate discussion of the generative transformational framework with a view to providing a complete and accessible explanation of the model for the specialist and non specialist alike. After a detailed intoduction to the generative model, Dell discusses implications for features in phonological processes for French.

> Si toutes les obstruantes sont nécessairement consonantiques, il n'en va pas de même des sonantes. Outre les consonnes nasales et liquides, il existe aussi des sonantes non-consonantiques. On range les sonantes non-consonantiques en deux catégories selon qu'elles sont ou non *syllabiques*. Sont syllabiques ([+ syll]) les sons qui peuvent à eux-seuls constituer une syllabe. Les autres sont *non-syllabiques* ([– syll]).
>
> (Dell 1973:63)

We should now be able to put to rest the issue of a phonetic quality that distinguishes a high vowel from a glide. In the generative model, glides are phonetically identical to their corresponding high vowel. The two segments are instead distinguished by the status of syllabic versus non syllabic. In fact as we have seen in Schane (1968) glides are taken to be the non syllabic realisation of a corresponding high vowel which can then be derived by transformational rules that represent the operation of a particular process on the high vowel when in the presence of the conditioning factors to model the phonetic alternations so often observed. In the spirit of this positional distinction we will adopt the term vocoid from Pike and use the term high vocoid to refer to the three high vowel segments that are distinguished as the [+syll] series /i y u/ and their glide counterparts distinguished as [-syll], that is the series /j ɥ w/. The distinction between a high vowel and its corresponding glide is now firmly based in syllable constituency. No longer is their distinction a matter for phonetics. As we observed for some liaison data, syllable constituency can be seen to play a very significant role in the realisation of non syllabic vocoids in French. We will see more on the implications of syllable constituencies for glides in Chapter 4.

Dell demonstrates in great detail the rule operation or transformational component of the generative framework. We will begin with the discussion of rule derivations that operate in the phonological component resulting in the phonetic outputs of the speech signal:

> La composante phonologique est un dispositif qui associe à chaque représentation phonologique une ou plusieurs représentations phonétiques. On peut concevoir ce dispositif comme une sorte de machine à fabriquer des représentations phonétiques à partir des représentations phonologiques qu'on lui fournit. À l'entrée (input) de cette machine on introduit des représentations phonologiques et à la sortie (output) on obtient des représentations phonétiques. Pour comprendre comment ce mécanisme fonctionne, il faut d'abord expliciter ce que sont exactement les représentations phonologiques, et de quelle façon les règles phonologiques s'y appliquent.
>
> (Dell 1973: 70)

Dell illustrates the basic notion behind the phonological component of a transformational grammar with classic French data involving the processes of Glide Formation (semivocalisation) and voice assimilation (devoicing) while demonstrating the implications for strict rule ordering. It should be understood that in the very short period between the SPE and this work, the feature ± syllabic is adopted in phonological representations. Evident from his earliest illustrations of the rule derivation component, Dell discusses the basic rule operations for the French phenomena to derive from phonetic forms processes of voice assimilation and Glide Formation (cf. also rules 8, 9 and 10, 1973: 72).

(49) Rule 2: Les segments [+ son, + haut] se prononcent [- syll] lorsqu'ils précèdent immédiatement une voyelle qui appartient au même mot.

(Dell 1973: 67)

Phonological transformations begin with the phonemic representation as input to the machine that through ordered rule operations then turns out the phonetic representation of the speech signal for a particular utterance. Glides do not exist in the phoneme inventory but are instead the [-syll] representation of a [± syll] high vocoid. As phonemic input to the machine, Dell takes the high vowels /i y u/ to represent the underlying form from which all glides will be derived.

> Voici enfin un dernier exemple. Le radical du verbe *manier* se prononce [mani] dans *il le maniera* [manira], *j'en manie un* [mani], *j'en connais le maniement* [manimã]. Il se prononce [many] dans *vous le maniez* [manye], *c'est très maniable* [manyabl]. Seul varie le dernier segment, et encore sa variation est-elle minime, puisque [i] et [y] ne se distinguent que par la spécification du trait [syll] : [i] est [+ syll] et [y] est [– syll]. L'alternance entre [+ syll] et [– syll] dans la prononciation du dernier segment de *mani-* est un cas particulier de la règle (2).
>
> (Dell 1973: 67)

Dell's rule (2) in (49) above is explained further in a footnote: "*La formulation de cette règle a été grandement simplifiée pour les besoins de l'exposition. Nous ne tenons pas compte des cas où le segment [+ son, + haut] est précédé d'une obstruante et d'une liquide, comme dans refluer, trouer etc., cf. Morin (1971) et Dell (1972)*" (Dell 1973: 67, note 15).

The rule for GF in French as given in (49) might also be represented in feature matrices as in (50) below.

(50) Rule governed operations in French Glide Formation

$$\begin{bmatrix} +\text{ syll} \\ +\text{ haut} \end{bmatrix} \longrightarrow \begin{bmatrix} -\text{ syll} \end{bmatrix} \;/\; \underline{\qquad\quad} \; \begin{bmatrix} +\text{ syll} \end{bmatrix}$$

Thus to demonstrate the transformational component or machine of the generative model, Dell illustrates the transformations by rule in the example of the verb phrase *vous écriviez*. The phonemic input for *vous écriviez* is thus /# # vuz # ekriv+i+ez # #/ where /i/ in the underlying form represents a [± syll] high vocoid. During the operation of this transformational component we will see the application of several rules: devoicing of /r/ when following another voiceless segment; Glide Formation when a high vowel is followed by another vowel; lastly, final consonant deletion, which transforms the utterance to its final phonetic representation [vuzekr̥ivye] in (51).

(51) Operation of rule governed transformations: *vous écriviez*

Représentation phonologique: INPUT		/##vuz#ekriv+i+ez##/
	Rule 1 [r] → [r̥]	/##vuz#ekr̥iv+y+ez##/
	Rule 2 [i] → [j]	/##vuz#ekriv+y+ez##/
	Rule 3 [z] → ø	/##vuz#ekr̥iv+y+e##/
Représentation phonétique: OUTPUT		[vuzekr̥ivye]

In this particular example, if we change the order of operation of these rules, the phonetic representation would remain unchanged. However, this is not always the case. Dell illustrates the crucial nature of rule ordering in the example of devoicing or DEV in combination with the rule of GF which he refers to as semi vocalisation or SEM. We give the rule for devocalisation or DEV in (52).

(52) DEV

$$\begin{bmatrix} +\text{ son} \\ -\text{ syll} \end{bmatrix} \longrightarrow [-\text{ voix}] \;/\; [-\text{ voix}] \; \underline{\qquad\quad}$$

(Dell 1973:78)

Rule (52) for devocalisation (devoicing) states that a non syllabic liquid, nasal or glide is devoiced when immediately preceded by a voiceless segment.[9]

Consider the data in (53) below.

(53) *dévie* [devi] ~ *dévié* ‘deflect/ed [devye]; *défie* [defi] ~ *défié* ‘defeat/ed’[defy̥e]
avoue [avu] ~ *avoué* ‘admit/ted’ [avwe]; *bafoue* [bafu] ~ *bafoué* ‘flout/ed’[bafw̥e]
gradu [grady] ~ *gradué* ‘graduate/ed’ [gradɥe]; *situe* [sity] ~ *situé* [sitɥ̥e]
‘situate/ed’

(Dell 1973:86)

These data show that glides effectively undergo devoicing when immediately preceded by a voiceless segment. As the rule states, it will operate on non syllabic sonorants, thus ruling out its operation on vowels. A rule SEM (semivocalisation) must operate before DEV to change the vowel /i/ into a glide /j/ and arrive at the output forms in (53).

In (54), we look at three derivations operating the rules SEM before DEV for *défié*, *dévié*, and *chiffré*.

(54)	/defi+e/	/devi+e/	/ʃifr+e/
	-	-	-
SEM	y	y	-
	-		-
	-		-
DEV	y̥		r̥
	[defy̥e]	[devje]	[ʃifr̥e]

(Dell 1973: 87)

By operating the rules in the order SEM before DEV for the three forms /defi+e/, /devi+e/ and /ʃifr+e/, we observe that although a change in ordering would not change the output for *dévié* or *chiffré*, this order is crucial to rendering the desired output for *défié*. The rule

[9] Dell stipulates that the devoicing rule does not apply to vowels. However, in light of the work of Cedergren and Simoneau (1985) on the varying degrees of high vowel devoicing and syncope, we would point out that for Canadian French, at least, this is not accurate. The rule would have to be modified if it were intended to account for other varieties of French. Nevertheless, for the purposes of this demonstration of crucially ordered rule operations these data serve their purpose.

SEM has triggered the change from [+ syll] to [– syll] rendering the resulting [– syll] high vocoid susceptible to DEV.

Chomsky and Halle (1968) claim the generative framework proposed is adaptable to all languages and indeed they demonstrate the transformational component using French glide data in liaison and elision processes requires further revision to the feature component. Schane (1968) and Dell (1973) account for a much broader sample of French glide data.

3.3.1 Summary of Observations

From Saussure to Trubetzkoy and Jakobson, the basic tenets of phonological analysis as elaborated by the Prague School: phoneme, distinctive features, and markedness, serve to advance the discipline of phonology on a broad scale but the role of syllable structure as was discussed in Pike (1943), and in Pike and Pike (1947) for glide phenomena is not taken up until later.

Martinet reminds us that contrary to the popular image of French (Standard or Northern Variety) as a homogenous language of the classic descriptions. In his survey of the WWII combattants with whom he was detained, results show much greater variability than is usually described for SF speakers and indeed glide variability proves a challenging issue for the generative model proposed by Chomsky and Halle (1968).

Chomsky and Halle's (1968) transformational generative model places the theory of distinctive features at the heart of the transformational component of their framework. Treatments of French glide liaison and elision data shows that the duality of glides that so plagues phonetic descriptions proves to be as challenging for the generative model: are glides consonants [-voc], or vowels [- cons]? Or are French liaison data best treated using [± foreign] or [± syll]? The questions about the featural composition of glides are persistent as a broader set of glide data is considered.

Schane's (1965) thesis (one of the earliest in this framework) focuses on transformational rule operations for French in particular to account for glide phenomena and brings to light

various aspects of glide data. Dell illustrates the generative model focusing on the aspect of rule ordering in transformational operations for French glide data. These accounts reveal areas of issue for an account of glide data in French (the featural decription of glides, for example).

In the next sections, we discuss later analyses of of French glide data that consider some as yet unexamined issues.

3.4 Later linear treatments of Glides

In the preceding sections, we have examined various aspects of French glide phenomena in a survey of early generative treatments (Chomsky and Halle 1968, Schane 1968, Dell 1973). In this section, we examine later work that specifically addresses aspects of French glide data and treatments in the generative model. Morin's analysis is based on a discussion of formal versus functional explanations. Using Glide Formation and schwa deletion data, Morin (1976) demonstrates that what he calls an OLG tension (a language specific constraint against Obstruant+Liquid+Glide (OLG) sequences) that may constrain derivations by rule. Lyche (1979) examines generative treatments for French glide data involving the noun derivational morpheme and verb inflection. She points to some failings of the transformational model (abstractness as well as inconsistencies of complete neutralization rules) and proposes a Natural Generative Phonology (NGP) analysis following Hooper (1975, 1976) and Rudes (1976). The specific theoretical implications are of less interest to us than are the particular glide data that Morin and Lyche bring to light.

3.4.1 Morin (1976)

Morin (1976) brings to light some important observations concerning the more recent development of a tension in the presence of an OLG sequence in the historical evolution French. This tension can be seen to act as a derivational constraint against the operation of GF and can be linked to properties of schwa deletion. According to Morin, "[i]t is argued that there must have been a phonetic change in the nature of glides, in this period,

making them more constricted and therefore less capable of appearing after an OL-cluster" (Morin 1976: 41 our emphasis). According to Morin, schwa deletion data (or the prevention of schwa deletion) points to a conspiracy under the tension to avoid the creation of an OLG cluster:

> When the liquid consonant is followed by a glide, the deletion of schwa is impossible in SF [Standard Fench], e.g. *chapelier* [ʃapəlje: *ʃaplje] 'hatter', *appeliez* [apəlje:* aplje] 'you called', *monteriez* [mɔ̃tərje:* mɔ̃trje] '(you) would climb. It must be observed here that glide-formation must take place before schwa-deletion; otherwise one could derive the deviant forms *[ʃaplije], *[aplije], * [mɔ̃trije], the underlying schwa being deletable in the intermediate forms /šapəlie/, /apəlie/, and /mɔ̃t + ə + r + ie/.[10]
>
> (Morin 1976: 40)

The OLG tension should therefore be understood not merely as "the result of some universal tendencies to avoid some complex configurations" (1976: 41), but as a language specific development that not only constrains the rule of glide formation but also has implications for schwa deletion (cf. Morin 1974) in the grammar of French. Evidence that at least for a period of time glide formation was unrestricted, occurring even after OL clusters, comes from three main sources: early diphongisation (historical glides): *brief*, *ouvrier*, *trois*, *croix*, *fruit*; 12th century reduction of sequences *ijV* to *jV* in verb endings *–ions* and *–iez* (*montrions* and *souffriez*, for example) and finally glide formation (Morin 1976:41).

Morin observes various strategies to relax the OLG tension (55).

[10] Morin notes that "[i]n the phonetic representations given here [below] schwas are represented as *œ* as this is their phonetic realization in most dialects" (1976: 39). We give transcription in IPA

(55) Strategies to relieve OLG tensions

a. Glide vocalisation: yod in Standard French (SF) (j→i: *ouvrier* [uvrie], *brièvement* briɛvəmã) in non-standard varieties vocalisation is primarily observed for ɥ→y and w→u (*fruit* [fʀɥi], *pluie* [plɥi]);

b. Loss of glide: *bref*, *grève* in (SF), *fruit* [fri], and *pluie* [pli] in non-standard varieties;

c. Loss of liquid (examples from non standard dialects not exclusively following OL *sanglier* [sãgje] but also *soulier* [suje]): *froid* [fwa], *crois* [kwa], *pluie* [pɥi];

d. Epenthetic schwa (examples from non standard dialects): *ouvrier* [uvərje: uverje], *sanglier* [sãgəlje]

(Morin 1976: 42 transcriptions added)

These strategies are employed in varying degree depending on the region. Morin distinguishes three dialects differentiated by the degree to which the OLG tension is relaxed: 1–tense dialects (with little to no exceptions); 2– lax dialects (always exceptions for *–ions* and *–iez*); and 3–semi-tense dialects where sequences of Oly are always possible exceptions but where Ory clusters are always relaxed.

> We observe in some dialects words where OLG-clusters are relaxed through schwa epenthesis, even though the glide is historically the initial high vowel of an hiatus, e.g. *prier* [pœrye: perje], *oublier* [ubelye], *trouer* [terwe], *truelle* [tœrẅɛl] (data from Gilliéron, Landreau, and Bourulot). In these forms the schwa cannot be accounted for unless the glide-formation applied first, and the resulting OLG-cluster was subject to tension-relaxation. In most cases (and in particular in SF), however, the initial high vowel of a hiatus after an OL-cluster is now syllabic. This was to be expected: when the effect of the tension began, glide-formation was optional, and therefore every OLG-cluster where the glide is a reflex of an earlier high vowel had a non tense variant with a high vowel instead of the glide. It is only normal that this variant should survive, and the other relaxed versions be exceptional.
>
> (Morin 1976:43)

Morin brings to light some important historical aspects of the OLG constraint that explain the presence of surface OLG sequences in the lexicon while OLG appears to be blocked from resulting in synchronic derivation.

> We have been able to show that at least in some dialects (and possibly in all dialects) both glide-formation and schwa-deletion […] were not restricted by the development of a tension in OLG-clusters […]. It has also been observed that relaxation was incomplete in SF and did not apply to historical [ɥ] or *w*, although always to derived [ɥ] or *w*. This is why in SF glide-formation is blocked in the synchronic derivation of words such as *trouait* /tru+ɛ/ > [truɛ:*trwɛ] and *troua* /tru+a/ > [trua:*trwa], even though OLw-clusters are found in the language, when *w* is historic, e. g. *trois* [trwa].
>
> (Morin 1976: 44)

This observation supports earlier observations about the significance of recognising syllable structure in glide treatments and theoretical models of French. We will see more on this in Chapter 4. Some historical forms (*-ions, -iez*, for example) will require special treatment in the transformational generative model (cf. Lyche 3.4.2), to account for their occurrence.

In a diachronic survey of developements in glide formation, Morin illustrates the effects of this tension on glide formation across regions and time. On the basis of which, this establishes his scheme for variant dialects (1976: 45 cf. Table 1): tense (forms adhere to OLG relaxations, semitense (forms adhere while others do not (conditioned by the particular high vocoid involved), lax (forms show no signs of tension).

In the next section we discuss Lyche (1979) who examines some problems observed for treatments put forth in the generative model that take the position all glides are derived from underlying high vowels.

3.4.2 Lyche (1979)

According to Lyche, since Schane (1968) and Milner (1969) "[i]t has been accepted that, whenever possible, glides should be derived" (1979: 315). Lyche reviews the standard generative treatment of glides, to demonstrate the inadequacy of the transformational generative model as compares with the Natural Generative Phonology (NGP hereafter). Lyche explains, "the extant transformational generative analyses of French glides have to be rejected not only because of their abstractness (complete neutralization rules) but also because they are inconsistent with the very principles of transformational generative phonology" (1979: 315). Lyche proposes instead an NGP analysis for a particular set of

French glide data (noun-forming suffix /ie/ and imperfect first and second person plural endings /iɔ̃/ and /ie/). We revisit the issue of rule ordering, first examined in Dell (1972), inherent in the transformational generative model. It is our goal to highlight issues for some problematic data that Lyche brings to light in her analysis and that are germaine to our discussion of French glide treatments.

Lyche (1979) calls into question the position held in the standard generative treatment (hereafter SG) that all glides should be derived from underlying high vowels. According to Lyche this position requires two separate (but similar) glide formation rules in SG treatments that must be extrinsically ordered to prevent any eventuality of illicit forms. Lyche focusses on rule operations involving the derivational morpheme *–ier* /ie/ (*sablier, encrier*, for example) as well as the imperfect 1st and 2nd person plural verb inflections *–ions* /iɔ̃/, *–iez* /ie/ (*vous guardiez,* for example). Note that these glide formation environments are preceded by an OL sequence. The rules used to derive glides for the SG analysis are referred to as SV1 (semi-vocalisation), SV2 (semi-vocalisation) and J-INS (epenthesis) reproduced here in (56-58).

(56) Glide Formation Rule

i. SV1 operates within a morpheme
ii. SV2 operates on a root final high vowel

$$\begin{pmatrix} +\text{syll} \\ +\text{high} \end{pmatrix} \longrightarrow [-\text{syll}] \;/\; X \_\_\_\_ [+\text{syll}]$$

a. $X \neq$ OL

b. $X \neq [-\text{syll}] \begin{pmatrix} -\text{syll} \\ -\text{cons} \end{pmatrix}$

(57) Glide Insertion Rule
J-INS
$\varnothing \rightarrow j /\ i \_\_\_ [+\text{syll}]$ (Lyche 1979: 316-317)

In (56) SV1 operates within a morpheme (first and second person plural verb inflections – ions /iɔ̃/ and –iez /ie/ as well as the noun-forming suffix /ie/ whereas SV2 applies to the final vowel in a root-morpheme i.e. at the edge of a morpheme (/li+e/, /ty+e/, /nu+e/). Only the domain of operation distinguishes one rule from the other: "[t]he two rules

present such striking similarities that one may wonder whether they are not in fact the same rule" (Lyche 1979:318). The glide insertion rule J-INS in (57) shows that yod insertion (epenthesis) operates only after the conditions (a. must not be precede by OL and b. must not be followed by G) are not met.

According to Lyche these rules require the strict ordering: SV1 before SV2 and SV2 before J-INS. The implications for the order SV1 before SV2 are illustrated in (58) below.

(58) *vous tuiez* 'you kill'

	a. /ty+ie/		b. /ty+ie/		c. /ty+ie/
SV1	j	SV2	ɥ	J-INS	j
SV2	—	SV1	—	SV2	ɥ
	[tyje]		*[tɥie]		*[tɥije]

(Lyche 1979: 318)

If J-INS applied before either SV1 or SV2 the wrong output is derived (58c.). While it may seem logical to collapse the two, Lyche (1979: 318) points out that SV1 must apply before SV2; SV1, obligatory dieresis in verb endings, is never allowed whereas SV2 is optional: "Even within the framework of transformational generative grammar, evidence against this particular order is not difficult to find," Lyche (1979: 318).

Lyche proposes that a Natural Generative Analysis (NGA) based on the following conditions offers a superior analysis:

(59) Conditions of NGA
a. No extrinsic ordering
b. Restrictions of Underlying Forms: the underlying representation of a morpheme must be identical to one of its phonetic realizations.

(Lyche 1979: 319-320)

In view of Hooper's True Generalisation Condition (TGC) which states "the rules speakers formulate are based directly on surface forms and that these rules relate one surface form to another, rather than relating underlying to surface forms" (Lyche 1979: 316), and since there is only one possible phonetic realisation for the verbal and noun-forming suffixes /je/ and /jɔ̃/ (dieresis is never possible):

SV1 is therefore a complete neutralization rule, a type of rule which must be highly motivated in order to be maintained in a particular grammar. Neither Morin [1971] nor Dell [1972] offer any such motivation. It is, moreover, likely that in order to maintain the generality of the analysis nonattested morphemes will have to be posited. It is clear that forms like *février* 'February' [fevrije], *ouvrier* 'worker' [uvrije], *sanglier* 'boar' [sãglije] are affected by the same process as *encrier* [ãkrije]. Thus, *février* and *sanglier*, for example, would have to be derived from /fevr + ie/, /sãgl + ie/. Thus, we conclude that SV1 is a highly suspect rule and should be dispensed with.

(Lyche 1979: 319)

Given the conditions of NGA (59), "[...] since the nouns considered above never alternate, they will be entered without any modification in the lexicon: *banquier* /bãkje/, *sablier* /sablije/. No special treatment is now needed to handle *février* and *sanglier* which have straightforward representations: /fevrije/ and /sãglije/" (Lyche 1979: 319-320). And by extension then the imperfect plural morphemes with only one possible phonetic realisation [jɔ̃], [je] must have the underlying representations /jɔ̃ / and /je/.

Lyche then turns her analysis to variable forms in French in an NGP analysis. Following Rudes (1976), the variation concerns alternating verb forms: *tuer* /ty+e/ [tɥe:tye], *lier* /li+e/[lje:lie] and *louer* /lu+e/ [lwe:lue]. According to Lyche alternation can be attributed to differing speech rates in two styles described here in terms of tempo (*andante*-slow formal speech and *allegro*-fast casual speech). Thus it is tempo in conjunction with the operation of two rules: the first J-INS (insert yod between iV) and the second I-INS (insert i between OLj). I-INS "will function as a redundancy rule in the lexicon and will account for forms such as *sanglier*, *ouvrier*, and *encrier*, etc." (Lyche 1979: 321). While Rudes (1976) stipulates that the output of *Andante* must be the input of *Allegro*, Lyche shows that this is not possible. "In our analysis the underlying forms to each tempo are identical: the input to Allegro is not the output of Andante" (Lyche 1979: 323). To strictly adhere to Rudes' framework would mean eliminating the two rules of insertion which would actually weaken the analysis since many of the generalisations observed would be lost. Lyche observes:

> But, if the output of Andante is the input to Allegro, we would start our derivations with base forms where a yod is already present: *sembliez* [sɑ̃blije], *priez* [prije], *oubliez* [ublije] [...]. The [i] in *sembliez* must now be erased, but not the [i] in *oubliez* (*[ublje]). Since we are dealing with surface forms, morpheme boundaries have been deleted. It is then no longer possible to differentiate between the two sequences. We are therefore forced to reject the idea that the output of Andante is the input to Allegro.
>
> (Lyche 1979: 324).

Lyche explains that the Strong Naturalness Condition (SNC) has been "widely criticized" prompting such fixes as to propose underlying representations in archisegmental forms. She opts to examine alternative analyses such as a Surface Phonetic Constraint to deal with this level of variability in the Glide formation data. We will see more of the archiphoneme in the next section when we discuss Kaye and Lowenstamm *De la syllabicité* (cf. Chapter 4.1).

To this point in the analysis, all Glide Formation, operating in conjunction with glide insertion rules, is subject to the condition X ≠ OL. In the analyses that follows, Lyche illustrates that "it cannot, however, be maintained that *GG or *OLG are true surface phonetic constraints (SPC) in French." (1979: 325).

Lyche discusses the data in (60) to show that OLG (60a.) and GG (60b.) do occur across word boundaries. She shows that the outcome after resyllabification would violate any surface phonetic constraint banning these sequences:

(60) Glide formation resulting in restricted OLG or GG forms

a. OLG <u>*Jacques riait*</u> *beaucoup* [ʒakrjɛ...]
'Jacques laughed a lot'

b. GG <u>La fille y est</u> bien arrive [fijje...]
'The girl managed well'

BUT

c. *Elle* <u>*envie Annie*</u> [ɑ̃viani]; *[ɑ̃vjani]; *[ɑ̃vijani]
'she envies Annie'

d. Il est si admirable [siadmirabl]; *[sjadmirable]; *[sijadmirable]
'He's so wonderful'

(Lyche 1979: 325)

Data in (60cd.) show that glide insertion is also not a requisite surface constraint to satisfy a preference for CV syllble structure. Lyche proposes that though OLG sequences are not banned from occurring in neighbouring contexts in spontaneous speech as in (60a.) *Jacques riait*, they are banned from occurring within the same syllable in the presence of a morpheme boundary (61).

(61) Restricted Glide formation *$OLG+V$
il cloua ‘he nailed’ /klu+a/ [klua] *[klwa]

Still there are contexts that require special consideration: "There exist, however, two syllable structure conditions which do not require any reference to a morpheme boundary: *GGV, *OrjV$. The function of *$GGV$ is to reject ungrammatical forms derived by (6): /apɥi + ɔ̃/, *[apɥjɔ̃] (*appuyons*). Rule (8) (i/j insertion) is entirely motivated by *OrjV$." (Lyche 1979: 322).

Lyche also mentions the possibility of environmental as well as dialectal variation such that these surface forms may in fact be permitted: "Our analysis clearly indicates that an OLj sequence can appear in French because of the restriction imposed on I-INS in Allegro (cf. I-INS in Andante). Moreover there exist dialects which allow Orj sequences in verb forms. For those dialects there is no rule of I-INS in Allegro." (1979: 327).

According to Lyche, “[w]e can now see that the rule of J-INS [yod insertion] is part of a well-known phenomenon in French: resyllabification in order to favour a CV structure. A similar process occurs in liaison where /lez#ãfã/ (*les enfants* 'the children') is syllabified /lezã$fã$/” (1979: 327).

Lyche brings to our attention the role of syntactic categories (as we saw in Schane’s analysis for the classic French phenomenon of liaison. This is a subject that gets much more attention later (cf. Encrevé 1988, section 4.3). Lyche examines here the implications for strong and weak syntactic boundaries in glide insertion data (we take glide insertion to be equivalent to glide epenthesis). While there is a great deal of similarity between liaison and glide insertion phenomena, maintaining the more desirable

CV syllable structure cannot be the only motivation for glide insertion. While liaison may occur across a weak syntactic boundary, “[i]n those contexts however, glide insertion [epenthesis] is not always possible" (Lyche 1979: 328). Below, we can see that glide epenthesis is not possible in (62ab.), whereas in (62c.), the almost identical context, we may see any of hiatus (c.i), glide (c.ii) epenthesis or glide formation (c.iii).

(62) Variable glide formation data

a. *Une joli enfant* ‘a handsome boy’
i. [ynžoliãfã] ii. *[žolijãfã] iii. *[ynžoljãfã]

b. *Il est si amusant* ‘he is so funny’
i. [siamyzã] ii. *[sijamyzã] iii. *[sjamyzã]

c. *Si Arthur veut venir* ‘If Arthur wants to come/to go’
i. [siartyr] ii. [sijartyr] iii. [sjartyr]

(Lyche 1979: 328)

Lyche proposes and we concur that the only difference that could explain the varying outcomes in (62b.) as compared with (62c.) is the possibility of greater stress given that the monosyllabic nonlexical item *si* in (62b.) “is a degree modifier and bears a certain amount of stress which pevents the application of the glide insertion rule” (1979: 328).

Lyche concludes that glide insertion (epenthesis) can occur after a monosyllabic non lexical item and across a weak syntactic boundary but only when unstressed. Therefore, glide epenthesis cannot be motiated entirely by a grammatical preference for CV syllable structure.

In this chapter, we show that even as theoretical models advance, the representation of glides in French continues to challenge the limits of the model. From feature representations of glides in French and their variability in liaison and elision data, to glide formation in processes of derivation and inflection, and OLG tension, French glide data remain central concern for phonological models.

In Chapter 4, we discuss the emergence of autosegmental phonology in the generative model. In this multi-level approach to phonology and morphophonology, the role of syllable structure is of primary concern.

Chapter 4

4 Autosegmental, Typological, and OT treatments of Glides

Although the role of the syllable in phonology is recognised early (Pike and Pike 1943; Pike 1947), the syllable and its constituencies do not play a central role until the advent of multi-linear representations and the autosegmental framework, wherein the linearity of segments and their distinctive features are integrated into a hierarchical model.

Due in part to the increasing recognition of the role of the syllable in phonological representations but also the emerging empirical study of non Indo-European languages (African and Asian tone languages and morphology in Semetic languages), phonological theoretical models move away from the rule-ordered linear approach of the standard transformational generative model and towards this multi-level metrical approach of the autosegmental model of a generative grammar. As Dell and Vergnaud observe:

> Selon certains travaux en phonologie métrique, il s'agirait d'une structure hiérarchique ayant les mêmes propriétés formelles que la structure en constituants immédiats familière en syntaxe, mais où les unités comprendraient par exemple : le segment, la rime, la syllabe, le pied, le mot phonologique, etc. [...] La phonologie autosegmentale s'est dévelopée autour de la remise en cause d'un autre aspect des représentations proposées dans SPE. La disposition en matrice implique une synchronisation uniforme des lignes entre elles, avec pour unité de temps le segment : chaque fois que, sur une ligne donnée, on passe d'une spécification à la spécification suivante, on passe aussi forcément, sur toutes les autres lignes, d'une spécification à la suivante, et cette propriété est déterminante pour le choix du formalisme dans lequel écrire les règles.

(Dell and Vergnaud 1984: 3)

We begin our discussion of this autosegmental approach with the work of Kaye and Lowenstamm (1984). *De la syllabicité* is a pivotal to a discussion of the modern treatments of glides since this is the first autosegmental treatment of glides in French.

4.1 Kaye and Lowenstamm (1984)

Kaye and Lowenstamm (1984) examine the implications of syllabicity for glide phenomena in French: "Dans cet article, nous montrons, à notre connaissance pour la première fois, que les 'effets de syllabicité' généralement conçus comme découlant de l'action de règles phonologiques, relèvent en fait de principes très généraux et ne requièrent aucune stipulation particulière au niveau des grammaires individuelles." (Kaye and Lowenstamm 1984:123).

It is the goal of these authors to demonstrate through glide phenomena the implications of syllabicity in French and Universal Grammar (UG) . They credit as inspiration the work of Pike (1947), Pike and Pike (1947) and the work in prosodic phonology of Vergnaud and Halle (1978).

First, these authors describe the formal representation of the syllable as well as other mechanisms within this model before examining the representation of glide data from French. Kaye and Lowenstamm (1984) begin by stipulating that syllable constituents are maximally binary branching. Cross-linguistic syllable inventories can be seen as the result of an implicational hierarchy: "[…] si une langue présente un certain type de syllabe, elle présentera également tous les autres types de syllabe de moindre complexité. Ainsi, si une langue présente des syllabes de type CVCC, elle présente aussi des syllabes de type CVC et CV, mais non pas vice-versa." (Kaye and Lowenstamm 1984: 124)

Kaye and Lowenstamm (1984) adopt Vergnaud's (1980) percolation convention that states: "[l]es noeuds d'une structure métrique peuvent être étiquetés. Lorsqu'un noeud est ainsi étiqueté, tous les nœuds qu'il domine doivent eux aussi porter l'étiquette en question*"* (Kaye and Lowenstamm 1984: 125). The convention of percolation is of primary significance in this model of the syllable since all metrical levels (including the syllable) are present in the lexicon where segmental material is assigned in accordance with all syllabic and formal constraints. Any rule that would modify the syllabic structure does so for the preservation of syllable structure. For this reason Kaye and Lowenstamm (1984) propose that null positions [-segment] must also be present yet constrained: "[l]es éléments nuls ne peuvent apparaître dans des constituents syllabiques branchants" (Kaye

and Lowenstamm 1984:127). Lexical representations of morphemes, for example, are subjected to reanalysis during a morphological derivation which requires the satisfaction of constraints at all metrical levels: "L'application de telles règles sera bloquée si la structure (l'analyse) syllabique qui en résulte, n'existe pas de façon indépendante, ou, de façon équivalente, si le type syllabique résultant n'est pas une élément de l'inventaire syllabique de la langue. (Kaye and Lowenstamm 1984: 127)"

After a derivation, syllable reanalysis may take place according to the principle of resyllabification:

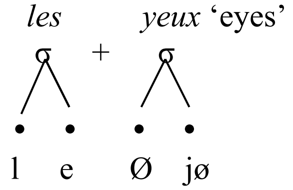
(63) The principle of resyllabification
Resyllabifier en éliminant les éléments nuls lorsque cela est possible, c'est-à-dire sans violer les contraintes syllabiques.
(Kaye and Lowenstamm 1984: 127)

If a resulting syllable structure does not exist in the syllable inventory of the language in question, rules for the preservation of syllable structure (epenthesis, for example) will apply.

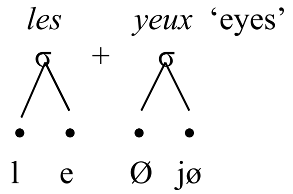
In (64) below, we illustrate the phonological derivation for the sequence *les yeux* 'eyes' /le+jø/ intended to demonstrate structure preservation by rule. There are a number of formalisations that are of great importance to our discussion of glide treatments yet to be discussed.

(64) A possible autosegmental derivation based on Kaye and Lowenstamm (1984)

a. Each lexical entry is associated with syllable structure and empty elements

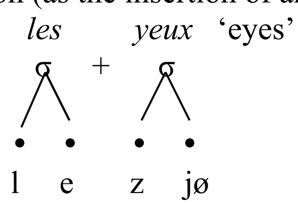


b. resyllabification removing all empty elements iff possible otherwise rules for the preservation of structure



Here the form *[lejø] would be deemed ungrammatical therefore resyllabification does not occur. The phonological derivation is then subjected to rule operations that will preserve structure (64c.).

c. liaison (as the insertion of an epenthetic segment)



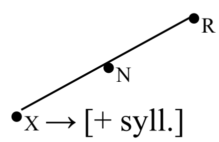
In (64c.) we give the example *les yeux* where the process of liaison in French is modelled as a process of epenthesis in order to demonstrate the resyllabification and structure restoration processes described by Kaye and Lowenstamm (1984).

Thus far the model as described applies to all languages varying only by their particular syllable inventory. There are however a number stipulations that relate specifically to glide phenomena and in particular to syllabicity. Recall that based on French liaison data Chomsky and Halle (1968) add the feature [± syllabic] to the distinctive features with which they define a minor class of segments (liquids nasals and glides) that behave as both [+ syllabic] and [- syllabic]. Kaye and Lowenstamm (1984) propose that in a framework such as theirs where the syllable is defined formally this feature must be

reconsidered: "En effet, dans un tel cadre, la notion ‘noyau’ ou ‘sommet’ syllabique est définie de manière structurelle: il s’agit des noeuds terminaux du constituant. La syllabicité peut être attribuée aux segments comme une propriété dérivée de leur position dans la structure métrique." (Kaye and Lowenstamm 1984:130).

As we have mentioned in the preceding chapter, most twentieth century accounts (Martinet 1945, Pike 1947, Schane 1968, Dell 1973), consider the French high vowels and their corresponding glides to be allophones of a single archiphoneme /I Y U/. According to this framework the feature ± syllabic is determined by the position in which a segment occurs in a syllable. The high vowel occurs in vowel nuclei and therefore is assigned the feature [+ syll.], while the glide occurs elsewhere and is assigned the feature by its non nuclear position [- syll.].

(65)



(Kaye and Lowenstamm 1984:130)

In view of this stipulation, French glides are considered inextricably linked to their corresponding high vowel such that all glides and all high vowels are underlyingly underspecified as the corresponding archiphoneme /I Y U/. Surface phonetic realisations depend entirely on syllabification (resyllabification) to satisfy constraints of syllable structure. With these conventions defined, the authors propose to resolve the issue of featural specifications for glides that presented such difficulty in the generative model of SPE.

> Nous remédions à cet etat de choses en proposant que les semi-voyelles et les nasales soient spécifiées, exactement comme les liquides, comme [+voc, +cons]. L’ hypothèse formulée ici est que l’important est que ces trois séries: nasales, semi-voyelles et liquides soient spécifiées de manière à ce que leur soit garanti l’accès aux trois positions A, N, C [Attaque, Noyau, Coda].
>
> (Kaye and Lowenstamm 1984: 131)

The authors set about illustrating the various mechanisms of their theory while accounting for a number of phonological issues (liaison, elision, and implications for true versus apparent diphthongs), for glides in Modern French.

Kaye and Lowenstamm (1984) note that the GV sequences in word initial position appear on the surface to be all alike, however, they do not behave alike with respect to processes of liaison and elision. The authors show that sandhi effects (liaison and elison) allow us to distinguish between apparently identical GV sequences involving a single glide. On the one hand a GV sequence may correspond to a nuclear diphthong preceded by an empty onset (permitting liaison and elision); on the other hand an apparently identical GV sequence may correspond to a nuclear V preceded by a non nuclear G associated with the onset, which blocks liaison and elision. This same distinction allows Kaye and Lowenstamm (1984) to account for the *OLG constraint which blocks the creation of new OLG sequences where a branching onset would be followed by a non nuclear glide while allowing apparently similar OLG sequences where the branching onset is followed by a nuclear diphthong as in *truite* [tʀɥit], and *trois* [tʀwa].

According to Kaye and Lowenstamm (1984), all derived glides are syllabified to an onset [-syll] position. OL sequences block glide syllabification to the onset since all syllable constituents are maximally binary. Under this account, branching onsets can only be followed by a glide which is part of a nuclear diphthong. However, the only branching onsets Kaye and Lowenstamm (1984) address are OL sequences and not for example sC onsets[11].

In the following section, we examine implications for glides in an autosegmental phonology model as proposed by Encrevé (1988). We begin the discussion with an account of homophonous yet distinct glide forms in Berber (Geurssel 1986). We then examine Encrevé (1988) who in an autosegmental account models French glide phenomena. These autosegmental accounts are followed by discussions of the implications for glide data in prosodic phonology. Hannahs (1995) discusses the prosodic domaine of GF in French, while Lyche and Girard (1995) discuss some of the implications for stress in metrical phonology on glide formation.

---

[11] The presence of derived sCG sequences like *skier* [skje] leads to some problematic implications: either onsets are not maximally binary thus permitting sCG sequences, or initial s is not parsed into the onset, or the glide is nuclear rather than associated to the onset. In any case these issues are beyond the scope of the present study.

4.2 Guerssel (1986)

Berber data shows high vowel glide contrast that cannot be accounted for using syllable structure associations. Berber verbs that appear to share segmentally identical stem forms surface differently. These apparently identical stems show the high vowel /u/ in final position, however, when a vowel initial suffix such as -ax is added surface forms show significant differences.

(66) High vowel glide alternations in Berber

a. /turi + ax/ → /turiyax/ 'she wrote us' not *turyax
/tessu + ax/ → /tessuyax/ 'she made us a bed' not *tesswax

b. /tusi + ax/ → tusyax 'she carried us' not *tusiyax
/tessu + ax/ → tesswax 'she made us drink' not *tessuyax

(adapted from Guerssel 1986: 4)

The forms in (66a.) show yod insertion (epenthesis involving yod only, Guerssel uses [y]) between two vowels in hiatus whereas the forms in (66b.) show alternation of the root final high vowel to a corresponding glide (GF, i→j, u→w). These data are most striking for the verb form *tessu* which shows yod epenthesis where forms with GF are not permitted in (66a.) but we observe GF in (66b.) where yod epenthesis is not permitted. Geurssel points out that these data are problematic for the model proposed by Kaye and Lowenstamm:

> If Kaye and Lowenstamm's hypothesis were adhered to, the facts described in [67] could not be handled adequately, because the alternating segments could not be distinguished from the nonalternating ones. Since Kaye and Lowenstamm do not distinguish the two sets, one would expect to find either one of the patterns illustrated in (13). But as indicated, only one set consists of acceptable strings.
>
> (Guerssel 1986: 8)

According to Guerssel neither syllabicity nor feature content alone can explain these facts for Berber. He proposes, then, that if some high vocoids are lexically preassociated with rime heads while others remain unassociated, the difference that they present can be viewed as a structural one (67).

(67) Lexical preassociation of rime heads in Berber

```
a.  t u r i           b. t u s i
    | | | |              | |  | |
    x x x x              x x  x x
      |   |                |
      R   R                R
```

(based on Guerssel 1986: 5)

Varying forms in Berber are accounted for through rules of syllabification that will determine their surface form.

> I will assume that in terms of feature content “high vowels” and “glides” have identical specifications and belong to the class [- cons]. Furtermore I propose that some members of this set should be lexically attached to rime heads, whereas all other segments should be unassociated.
>
> (Guerssel 1986: 5)

The rules of syllabification employed in Geurssel’s analysis are shown in (68) below:

(68) Rules of syllabification in Berber

a. Where X’, an unassociated [- cons] segment not adjacent to a rime head, is associated with a rime head.

```
X’ → X
     |
     R
```

b. All unassociated X’ not adjacent to a rime head are associated with an onset.

```
X’ X → X  X
   |   |  |
   R   O  R
```

c. All unassociated X’ adjacent to a rime head are associated with a branching rime.

```
X  X’ →  X  X
|        |/
R        R
```

d.

```
          i
          |
∅  →  X / X    X  (yod epenthesis)
      |   |    |
      O   R___R
```

e. All assigned segments are associated with a syllable head

```
(X)  X → X   X
 |   |   |   |
 R   O   O   R
         |/
         σ
```

Guerssel’s analysis shows merit for instances of exceptional glide realisations in French. We will discuss the implications of this model for variable glide data in Chapter 7.

4.3 Encrevé (1988)

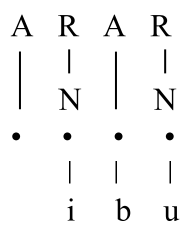
In an empirically driven formal analysis, Encrevé (1988) applies the autosegmental model to the representation of autosegments and segments in interaction with skeletal

positions and syllable constituencies to elaborate a description of variable liaison in French as can be observed in modern usage. In doing so he proposes:

> [...] une distinction claire entre deux types d'épenthèses (introduction d'un segment) qui sont généralement confondus: les « *vraies* » *épenthèses* pour lesquelles il n'y a pas de « structure d'attente » dans le squelette de la représentation lexicale tridimensionnelle et les « *fausses* » *épenthèses* qui sont structuralement attendues ; avoir posé une hypothèse solide sur l'autosegment flottant comme jonction entre la variation et la structure : comme locus structural des potentialités de variations grammaticalisées.
>
> (Encrevé 1988: 18, original emphasis)

Following the formalism presented in Vergnaud (1982) and by extension Kaye and Lowenstamm (1984), Encrevé offers the representation of *h-aspiré* words in the autosegmental model:

> Dans le cadre tridimensionnelle, nous avons proposé (Encrevé (1983) de suivre Vergnaud (1982) qui intérprète le *h aspiré* comme un élément A [attaque] *associé* à la position initiale du squelette du mot, mais sans segment correspondant sur la ligne segmentale: une *attaque vide non nulle ancrée* dans le squelette (cf. (21c)). Ainsi *hibou* sera représenté sous la forme (70):
> (70) *hibou*

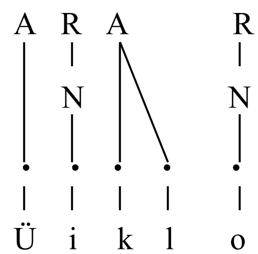


(Encrevé 1988: 197, original emphasis)

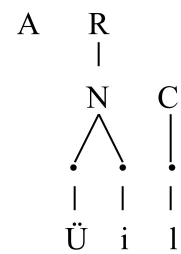
While the *h aspiré* receives no phonetic interpretation, the empty position is associated (anchored) to a skeletal position (•). As a result, no *consonne de liaison* (CL) may be associated with this position. The discussion of anchored versus non anchored positions has implications for glide representations. Encrevé notes that variable liaison has led some to propose that these words could be treated as words beginning with an *h aspiré* (cf. Cornulier 1978). Encrevé, however, does not agree:

Il n’y a pas de raison pour traiter ces mots [à glide initial] comme des mots à *h aspiré*. Nous adoptons pour eux une representation reprenant, sur ce point, dans notre cadre, les propositions de Kaye and Lowenstamm (1984): dans *huis clos* and *yaourt*, le glide est associé à l’attaque *(68a)*, ce qui interdit la liaison puisque la Condition d’ancrage *(45)* n’est pas remplie, mais non l’enchaînement des CF dans la mesure où il peut se réaliser avec une attaque pleine: dans *huile* ou dans *yeux*, au contraire, le glide est associé au noyau (noyau complex) et précédé d’une attaque nulle, qui permet la liaison (73b):

(73) a) huis clos



b) huile



(Encrevé 1988: 200, original emphasis)

Encrevé uses the transcription Ü for [ɥ]. The model proposed by Encrevé works much as that proposed by Kaye and Lowenstamm (1984). In a multilinear model, Encrevé proposes different syllable structures to represent vowel initial forms that show variable liaison. Glide initial words as in *huis clos (73a)*, for example, may have the structure wherein the glide [ɥ] is syllabified to the syllable onset (Attaque) or it may have the structure as in the word *huile (73b.)* wherein [ɥ] is syllabified as part of diphthong in the nucleus (N), essentially, variation by lexicalisation. In (69) we show Encrevé’s representation for the variable form *ouate* which shows both realisations with liaison and without liaison.

(69) *ouate*

We turn now to discussion of French glide data in the prosodic phonology model following Nespor and Vogel (1986). In these various discussions, we observe that both

the domain of an operation as well as the level of primary stress assignment may be factors in accounts of glide phenomena in French.

4.4 Prosodic Phonology

Nespor and Vogel's (1986) model of the prosodic hierarchy can allow us to account for the interaction between the various components (domains) of the phonological grammar:

> The model presented here is a theory of phonological domains, that is, a theory that organizes a given string of language into a series of hierarchically arranged phonological constituents that in turn form the contexts within which phonological rules apply.
>
> (Nespor and Vogel 1986: 6)

Nespor and Vogel stipulate "[a]lthough we will not be concerned here with the principles of syllabification per se, it is nevertheless necessary to consider the question of the domain within which syllabification applies […]" (1986: 62). Using data from French and Spanish, they show that while all languages exhibit word-domain syllabification wherein one set of rules governing syllable well formedness may apply early in the grammar, it is also necessary to stipulate that for languages like French and Spanish a grammar may allow cross word-boundary resyllabification to apply later in the grammar:

> Whether or not resyllabification is allowed is a parameter of the phonology that must be fixed for each language: it is probably not a coincidence that resyllabification tends to be present in Romance languages but absent in Germanic languages. What still remains to be determined is the larger domain within which resyllabification takes place in those languages in which the phenomenon occurs. Above the word level there are four prosodic constituents, each of which is potentially a domain for resyllabification. In French it seems that the domain of obligatory Liaison is the phonological phrase (see chapter 5 below and Selkirk, 1978b), while in Spanish the domain appears to be larger –the intonational phrase, or perhaps even the phonological utterance.
>
> (Nespor and Vogel 1986:72)

According to Hannahs' account of the application versus non application of phonological rules involved in the phonological processes of several unrelated languages (Italian /s/ voicing, Hungarian vowel harmony and compounding, and in Korean the neutralisation of syllable final consonant contrast), prefixes often behave as an independant

phonological word in that they are seen to block the application of the phonological rule in question. Hannahs proposes to test this hypothesis for French:

> In fact, certain observations in the literature lend initial support to this hypothesis: in discussing glide formation (see section 3.1 below), Johnson (1987) notes that "certain prefixes behave as 'separate words.'" Tranel (1976) observes that the productive prefix *in-* acts as though there were a word boundary (#) following it (see section 3.2 below).
>
> (Hannahs 1995:1130)

To test this hypothesis against the facts for French, Hannahs posits the French phonological word as a stem plus any suffix or a prefix (70).

(70) French phonological word

a. stem plus any suffix(es)
b. prefix (Hannahs 1995: 1130)

Hannahs examines the domain of rule applications for glide formation and prefix nasalisation data. We restrict our discussion to arguments made for GF. Hannahs identifies two sources of glides in Modern French: underlying as in *trois* and *truite*, and derived as in *troua* and *monstrueux*. He limits his discussion to derived forms. "While GF applies word-internally [...] it does not typically apply across words" (Hannahs 1995:1131).

(71) Evidence against glide formation across word boundaries

j'envie Alain
[ʒɑ̃vialɛ̃] cf. *[ʒɑ̃vjalɛ̃] 'I envy Alain'
(je) joue au (football)
[ʒuo] cf. *[ʒwo] 'I play football'
(il a) dû attendre
[dyatɑ̃dʀ] cf. *[dɥatɑ̃dʀ] 'he had to wait'

(Hannahs 1995:1131 cf. also Johnson 1987: 893)

As noted in Chapter 3, Lyche suggests that glide formation across word boundaries is a possible outcome in spontaneous speech (variable surface forms for two apparently identical contexts (Lyche 1979:83)) which we have reproduced here in (72) below.

(72) Variable cross word-boundary glide formation

a. *Une joli enfant* ‘A handsome child’
i. [ynžoliãfã] ii. *[žolijãfã] iii. *[žoljãfã]

b. *Il est <u>si amusant</u>* ‘He is so funny’
i. [siamyzã] ii. *[sijamyzã] iii. *[sjamyzã]

c. *Si Arthur veut venir* ‘If Arthur wants to come’
i. [siartyr] ii. [sijartyr] iii. [sjartyr]

(Lyche 1979: 328)

With respect to cross word-boundary glide formation it would appear that not all cross-word-boundary contexts behave alike.

Hannahs claims that evidence from the prefix *anti*- as well as noun compound data support the position that the domaine of phonological word (PW) as defined here

is the domain of glide formation “[t]here are, however, two environments in which GF does not occur word-internally (distinct from the underlying glides discussed above); These are between prefixes and stems, such as *anti-, semi-* [73a.], and between members of a compound [73b.].”

(73) GF blocking environments

a. antialcoolique
[ɑ̃tialkɔlik] cf. *[ɑ̃tjalkɔlik] ‘anti-alcohol’
b. tissue-eponge
[tisyepɔ̃ʒ] cf. *[tisɥepɔ̃ʒ] ‘terrycloth’ (Hannahs 1995:1131)

Following Johnson (1987), who finds that some prefixes behave as separate words with respect to phonological processes, Hannahs concludes that “consideration of the prosodic hierarchy and its involvement in the application of phonological rules provides a coherent explanation for the application of GF and, more important, its blocking not only at the phrasal level between words but also in compounds and prefixed words.” (Hannahs 1995:1131-32).

Hannahs mentions that an alternate analysis which considers the role of stress in glide formation still supports the hypothesis that the PW is the domain for glide formation:

> That is, that a stressed high vowel does not become a glide even when followed by another vowel. This may be a correct characterization of why GF is blocked at the end of PW—that the PW defines a domain of stress assignment in French and that the vowel in such a domain is stressed. If this is the case, it provides further evidence for the proposed configuration of the PW: given that French has word-final stress, the occurrence of a stressed vowel at the end of a prefix would imply the presence of a word boundary there, as suggested by this analysis.
>
> (Hannahs 1995:1132)

Hannahs concludes, "GF is a domain span rule of French applying within the phonological word." (1995: 1133). This follows from the criteria motivating a phonological constituent set out by Nespor and Vogel "there are rules of the grammar that have need to refer to it in their formulation, [...]" (1995:59).

The observations in Hannahs (1995), as shown here, follow from only a small subset of data which cannot be considered representative of all GF environments in French. If, however, we were to consider even the slightest degree of variability in 60 above, such as discussed in Lyche (1979), and the possibility that a stress assignment rule operates higher in the prosodic hierarchy say, at the level of the intonational phrase, we would have to allow that either the definition of the phonological word or the domain of glide formation be reconsidered. From our discussion of variability in glide formation (Chapter 1), it should be evident that syllable structure alone does not condition GF. Rather glide formation is but one strategy by which syllable structure may be altered, constrained (or reduced) in an utterance, as need be.

Hannahs acknowledges that stress placement may also be a factor but he considers only stress placement at the level of the PW. Nespor and Vogel propose that in French and Spanish resyllabification may occur at a later level (the intonational phrase, for example). Resyllabification at a higher level than the PW could therefore account for glide formation occurring across word boundaries where prefixes like *anti*-, *semi*- and compounds may be accounted for in the PW domain.

We turn now to Lyche and Girard (1995) who provide evidence to suggest that stress placement is in fact in a state of flux leading to an increased significance in the role of the lexical word in French. Lyche and Girard cite Saussure (1975:154) "[l]e mot, disait

Saussure malgré la difficulté qu'on a à le définir, est une unite qui s'impose à l'esprit, quelque chose de central dans le méchanisme de la langue" (Lyche and Girard 1995:205).

> Si, en français, le mot lexical n'est pas perçu comme une unité phonétique, cela est dû essentiellement à deux facteurs: l'accent et la syllabation. Dans ces deux cas le mot lexical est subordonné au groupe et c'est l'unité que l'on nomme traditionellement 'groupe rhythmique' qui constitue le domaine d'accentuation privilégié.
>
> (Lyche and Girard 1995: 206)

Lyche and Girard propose that glide formation is one of several processes, each long-attested in French, which conspire to support the autonomy of the word in French phonology. As presented here, glide formation (and glide epenthesis) operates on the high vowels at a morpheme boundary (74a.), and not, at a word boundary (74b.).

(74) Glide formation

a. at morpheme boundaries

je tue [ʒøty] 'I kill' *nous tuons* [nutɥɔ̃] 'we kill'

il rit [ilri] 'he laughs' *vous riez* [vurje; vurije] 'you laugh'

b. blocked at word boundaries

joli enfant 'handsome child' [ʒɔliɑ̃fɑ̃] not *[ʒɔlijɑ̃fɑ̃]

(Lyche and Girard (1995: 210)

Lyche and Girard add in a note: "Cette même règle nous permet d'ailleurs d'affirmer que les prefixes *anti* et *bi* forment chacun un mot phonologique car la règle de semivocalisation ne s'applique pas dans *antiatomique*, *biannuel* (Basbøl, 1981)" (1995: 210, cf. note 9). In this light the word is not so much defined by its orthographic boundaries as by the domain of the phonological word which is itself defined by the limits of the various processes observed.

According to these authors, recently in French, an innovative initial stress is observed in use: "Cet accent diffère de l'accent final de mot ou de groupe qui est beaucoup plus faible. Fouché (1969: LXII) a reproché à la langue familière l'emploi systématique de cet accent [...] L'accent d'insistance devient fréquent, si fréquent qu'il peut finir par paraître normal, " (Lyche and Girard 1995: 212). Lyche and Girard note that both Lucci (1983) and Encrevé (1988) have commented on the use of an innovative accent preferring the term *initial* "[n]ous dirons aussi que l'on trouve cet accent dans le style *oratoire* (1995:

212 cf note 11, original emphasis).

As noted earlier, the traditional (fixed) stress pattern in French involves stress assignment to the last syllable in the intonational phrase (IP). The initial accent discussed here is assigned to the first syllable of the group at the level of the clitic group (CG). In some cases a CG can be identical to a phonological word (PW). According to Lyche and Girard "Toutes les modifications des règles phonologiques [...] contribuent à introduire dans la langue un nouveau signal de demarcation, car [...] la fonction primordiale de l'accent initiale est une fonction démarcative [...]", (1995: 218).

(75) Son imagination 'her imagination'

a. [sõnimaʒinaˈsjõ]

b. [sõˈnimaʒinasjõ] (Lyche and Girard 1995: 218 cf. (42))

Given the proposed availability of the clitic group to an additional stress assignment we wonder if there are not also implications for variable glide formation. Since glide formation occurs only when a high vowel in hiatus is not stressed the initial accent described by Lyche and Girard here could certainly block glide formation in monosyllabics, for example, where it might otherwise occur.

The initial accent could have serious implications for various processes of French phonology (as discussed by these authors). It is not however a serious risk to the rules of accentuation that are systematic.

> Cet accent, nous l'avons vu, est fort répandu dans les media, parmi les enseignants, chez tous ceux qui prennent la parole en public. Pourtant, ce n'est pas à notre avis une nouvelle norme à enseigner aux étudiants étrangers. En effet, s'il est très répandu, il est loin d'être généralisé.
>
> (Lyche and Girard 1995: 219

As discussed here, the inital accent can be seen rather as a form of as stylistic variation such as is observed in the media.

According to Hannahs, "certain phonological phenomena occur within a stem or stem + suffix but do not occur across a prefix and stem. " (1995: 15). These GF blocking environments can be taken, therefore as an indication that the domain of glide formation

is the phonological word. Tranel describes these environments as follows:

(76) GF blocking environments

a. across a prefix+stem juncture : anti-aerien [ãtiaeʀjɛ̃] ‘anti-aircraft’
b. across compound juncture : Marie-Anne [maʀian]‘Marie-Anne’
c. across words : si adorable [siadɔʀabl] ‘so cute’

(Tranel 1987: 119)

We would propose rather that the examples in (76) are, in fact, better accounted for by primary and secondary stress assignment within the accentual phrase. Glides are never nuclear and are therefore never assigned a primary stress accent. By corollary, a high vowel that is assigned stress will not undergo glide formation. We wonder if a demarcative stress such as described by Lyche and Girard (1995) may apply to the forms given in (76).

These various discussions of prosodic domains in a phonological grammar are quite compatible with the model of lexical phonology and post-lexical phonology (cf. Kiparsky 1982). In this model phonological rules are compartmentalised such that in the lexical component of the grammar some rules interact at various levels with word building processes such as suffixation and inflection. In the post-lexical component some of the same phonological rules (usually involving allophonic variation) interact between words. It is in this post-lexical component of the grammar that cross-word-boundary glide formation might be modelled. The mechanism of cyclicity in this model provides that a single phonological rule (glide formation, for example) may operate more than once and in more than one component. At least some aspects of the grammar that are seen to operate in the lexical component may remain active in the post-lexical component. In the next discussion we examine very briefly the implications for the displacement of a primary stress accent and the resulting glide formation in Spanish.

Carreira (1989) examines the effect of stress for glide formation in Spanish. She demonstrates that a high vowel in hiatus through derivation relinquishes stress to the benefit of the affixal nucleus (/*perú* + *ano*/ → /*peruáno*/ ‘Peruvian’, /*actú* +*amos*/ → /*actuámos*/ ‘we act’. As a result of the transfer of stress the formerly stressed high vowel

in hiatus undergoes glide formation and with the newly stressed affix vowel as the on-glide of a rising diphthong. A similar process can be observed in French. During suffixation that results in a HV+V sequence, the stress of a root morpheme is transferred to the inflectional suffix (77).

(77) Stress accent transfer in a process of inflection

i) /i/: *je lie* [lí] 'I tie'; *nous lions* /lí + ɔ̃/→ [ljɔ̃́] 'we tie'

ii) /w/: *je joue* [ʒú] 'I play'; *nous jouons* /ʒú + ɔ̃/ [ʒwɔ̃́] 'we play'

iii) /y/: *je soue* [sý] 'I sweat'; *nous suons* /sý + ɔ̃/ [sɥɔ̃́] 'we sweat

Unlike Spanish, the primary stress accent is not contrastive in French. This does not however change the analysis of stress transfer from a root to an affix during derivation or inflection. The stressed high vowel of the root morpheme must relinquish the stress accent in order to satisfy final stress placement of the grammar and as a result the formerly stressed high vowel undergoes glide formation.

We turn now to a discussion of the cross-linguistic typology of glides and their implications for markedness. In the following section we examine the inventory and distribution of French glide data in view of a cross-linguistic typology of glides. Finally, we examine French glide data in an optimality theoretical (OT) approach.

4.5 Linguistic Universals: Markedness, Typology, and Glides

> Underlying the endless and fascinating idiosyncrasies of the world's languages there are uniformities of universal scope. [...] Language universals are by their very nature summary statements about characteristics or tendencies shared by all human speakers. (Greenberg 1963: xv)

For a discussion of the implications of linguistic typology for a theoretical model of phonology it is imperative to acknowledge the work of Joseph Greenberg. Though his early training and work are firmly rooted in the tradition of anthropological linguistics, it is by his later association with Jakobson and Martinet that Greenberg's work comes to reflect influences of Prague School Structuralism. Jakobson acknowledges the impact of his work on linguistic universals and their implications for phonological theory:

> On the grammatical level, J. H. Greenberg's list of 45 implicational universals is an impressive achievement. Even if advancing research somewhat reduces the number of exceptionless universals and increase the sum of near-universals, these data will remain invaluable and indispensable preliminaries to a new typology of languages and to a systematic outline of the universal laws of grammatical stratification. (Jakobson 1966: 268)

Most pertinent to our discussion of glides, is the notion of the Implicational Universal: "[t]he logical form may vary. It is typically, though not invariably implicational. For all values of X, if X is a language, then, if it contains some feature α then it always contains some further feature β but not necessarily vice-versa." (Greenberg 1966: 61).

The size and shape of a phonemic inventory not only for a given language but also cross-linguistically has significance in understanding universals. With respect to glides in particular, cross-linguistic distributions for glides and any implications that may be gleaned therein serve to inform our analysis of glides in French. In the following sections we discuss implications of markedness, typology, lexical frequency, and child language acquisition for French glides.

4.5.1 Markedness

Through observations of the general tendencies in phonological grammars, Trubetzkoy (1939/1949) proposes a theory of markedness to explain the significance of the relationships contracted between the various segments, or classes of segments and their respective distribution within the grammar of a language. According to Hume "Trubetzkoy assumed that one member of a sound opposition bears some property or 'mark' that the other member of the opposition lacks" (Hume 2003: 295). Today markedness has become a significant tool for phonological models. Through the lens of markedness relationships in a phonological model, we gain some added perspective on sound classes within inventories which informs our understanding of the significance of class and featural distinctions and their role in the mechanisms of phonological processes. Rice points out however that, "[o]ver the years, the use of this term has grown and expanded in many ways so that today, while the notion of markedness is core to phonological theory, capturing exactly what it means is not straightforward" (2007:79).

Greenberg applies the notion of markedness to the study of language universals and typology. In his (1966) study of the phonological, grammatical and semantically marked/unmarked relationships, he focuses on the particular implications for textual frequency to determine marked versus unmarked forms. According to his analysis marked and unmarked status amongst segments can be a correlate of frequency; unmarked being the most frequently occurring forms in a grammar. He relates the correlation between frequency and markedness to Zipf's principle of least effort; "For, if the marked feature contains something which is absent in the unmarked, it is relatively more complex and by Zipf's well-known principle of least effort the more complex should be used less frequently" (1966: 64). However as Rice (2007) points out, the notion of complexity is hardly cut and dried. Correlation to frequency is but one of many factors which may significantly contribute to our understanding of markedness within the phonological grammar:

> Jakobson (1941/1968) proposes that markedness constrains phonological inventories, systems, and rules and plays a role in determining sound change and the order of acquisition of sounds; relative frequency, combinatorial capacity, and assimilatory power of features are determined by the priority relationships within the universal feature hierarchy that he proposed.
>
> (Rice 2007: 80)

Rice illustrates by way of the taxonomy of markedness relationships to what extent phonologists have embraced or expanded on its role in phonology: structural markedness, or frequency markedness attributed to Bybee (2001), for example.

Perhaps most pertinent to our discussion of glides, syllable structure is one of the most practical and least controversial examples with which we may also demonstrate the implications of structural markedness "[b]ased on implication, CV syllables are considered to be unmarked: the existence of, for instance, CVC syllables or of V syllables in a language implies the existence of CV syllables in that language" (Rice 2007: 81) but as Rice points out it is not always this straightforward. Carton provides some of the statistics as well as commentary on syllable structures in French:

> Pour le français, Delattre a établi, d'après un corpus parlé qu'il y a 54,9 % de structures CV (consonne + voyelle), alors qu'il n'y a que 17,1 % de structures CVC, 14,2 % pour CCV et 1,9 % pour VC. L'espagnol est assez proche du français avec 54 % de CV, mais l'allemand et l'anglais sont très différents. Les structures lourdes en consonnes sont rares en français. Le français, dit-on, est « riche en voyelles » ; mais il ne s'agit en fait que d'une impression due à *la grande proportion de syllabes libres*.
>
> (Carton 1974: 78, original emphasis)

Specifically, Rice addresses the difficulties of capturing exactly what featural markedness means. She points out several different diagnostics based on various phenomena observed and their outcomes which can be described not only as the emergence of the unmarked (the case of neutralisation and epenthesis) but also the submergence of the unmarked (coronal as target of assimilation, deletion or coalescence wherein unmarked features are lost), while among vowels:

> Similarly, high and low vowels are often proposed to be unmarked with respect to mid vowels. Phonologically, high vowels are common in epenthesis and often result from neutralization; in addition, high vowels are frequent in inventories and, generally, the presence of mid vowels in an inventory implies the presence of high vowels.
>
> (Rice 2007: 82)

If we consider the notion that the marked element is defined as containing something "which is absent in the unmarked" (consider CV versus CV<u>C</u> syllables, for example) this particular element could potentially be defined on more than one level in the phonology of French. Bullock and Tranel point to the implicational aspects of French glide distribution so that this element could be defined in terms of the defective distribution for /w/ and /ɥ/ that is not the case for the /j/. Still other aspects including featural markedness, as mentioned by Rice (2007), may inform our analysis of glides. We observe, for example, that the bi-labial velar /w/ and the bi-labial palatal /ɥ/ glides, in terms of their phonetic make-up, are described as doubly articulated segments and as such they can be said to be more complex than the palatal /j/. Certainly distribution data from French support this observation. Glides may alternate with a corresponding high vowel (in glide formation); glides may occur as an epenthetic segment when the high vowel occurs in hiatus, yet as Tranel (1987) and Bullock (2002) point out, not all glides participate equally. They characterise this variable participation in glide formation and

epenthesis as resulting from an implicational hierarchy j » w » ɥ such that if glides are going to emerge in processes such as epenthesis or glide formation, first, /j/ will occur before /w/ and /w/ before /ɥ/. Thus, if we observe the participation of /ɥ/ we must assume the presence of the other two glides. This hierarchy of segmental distribution might also be described in terms of place features so that the front unrounded occurs before the back rounded and the back rounded occurs before front rounded.

In the following section we look at some data for the cross-linguistic distribution of glides. We will see that this implicational hierarchy holds, not merely for French data, but also for the cross-linguistic typology of glides. All of these factors point to an unmarked status for yod in French which is in keeping with Greenberg's observation of a correlation between markedness, complexity, and frequency.

4.5.2 Cross-linguistic Typology of Glides

> Organization of some central source of data, something like a cross-cultural file for a large and representative sample of world languages, would vastly facilitate the establishment of well grounded universals and their continued study by scholars.
>
> (Greenberg 1963: xvi)

To categorise glides in French we propose the following scheme based on a triggering phonological environment /HV+V/ in the case of variable realisations in (81ab.), or on inalterability with a corresponding high vowel in the case of (81c.), an invariable realisation such as word-final yod in *baille* [baj] *[bai] or in the case of the diphthongs [wa], *[ua], [ɥi], *[yi], [jɛ̃], *[iɛ̃].

(78) Three environments for glides in French

a. Derived: variable high-vowel (HV)/glide (G) alternations (HV→G) observed in stem final high vowels when followed by a vowel initial morpheme: (C)HVV→ (C) GV as in *tuer* [tɥe].
b. Derived: variable epenthetic glide (spreading of the stem final HV to a following empty onset in the desinence): (C)HV_V→(C)HVGV as in *trouer* [tʀuwe];
c. Lexicalised: invariable glides: high vocoid initial morphemes *–ien(nne)*, *–ier* word medial GV in mono morphemes such as *rien*, *pied, bien,* for example ; word-final G paye [pɛj]; travail [tʀavaj]; word medial GV (true diphthong) proie [pʀwa], *truite* [tʀɥit), for example.

Before we look more closely at the case for glides in French specifically, let us examine the data for the distribution of glides cross-linguistically. According to the UPSID (*UCLA Phonological Segment Inventory Database*) database which examines the phonemic inventories of 317 languages, Maddieson (1984) reports that 90 percent of these languages have at least one vocoid approximant (glide or semi-vowel). Hexagonal French, with a three-glide inventory /j w ɥ/ (as well as an extensive vowel inventory), appears to be quite exceptional. Of the 317 languages surveyed, the vast majority (86.1 percent) have the palatal approximant /j/ and 75.7 percent have the velar bi-labial /w/. The labial palatal /ɥ/ however is rare; only four cases or 0.01 percent of languages count /ɥ/ in their inventory.

> Apart from those approximants which have a lateral escape or belong to the family of r-sounds, the only frequently-occurring approximants in the world's languages are those which have vocoid characteristics (Pike 1943). [...] In the UPSID file, the vocoid approximants have been coded as consonants if they don't alternate with syllabic vocoid pronunciations and share distributional properties with other consonants.
>
> (Maddieson 1984: 91)

If we consult the phoneme charts provided, we find that for French (number 010) all three vocoid approximants (palatal /j/, labial-velar /w/ and labial-palatal /ɥ/) are listed in the consonant inventory. Based on the stipulation above we take this to mean that the three glides of French are considered phonemic non-alternating consonants. While it is true that the glides in French are never syllabic, thus sharing distributional properties with consonants, their phoneme status remains unclear. Levi (2004) suggests that the phonemic status of the glides in phoneme inventories should be closely scrutinised: "It is

my belief that in most cases, the evidence needed to posit underlying glides is not found; thus the language learner instead assumes that the surface glide is the result of universal syllabification tendencies" (Levi 2004: 6).

Maddieson offers some explanation of the methodology used in determining phoneme status. He stipulates that determinations are based on distribution and contrastivity:

> For our purposes "contrasts" are sound differences capable of distinguishing lexemes or morphemes in the language involved, given that data on factors that principally concern phonological properties themselves, such as stress placement, syllabification, boundaries (word, morpheme, etc.), and so on, can be used to predict variants but that diacritical features, arbitrary rule types, morpheme classes etc. cannot be so used. These principles are applied to evaluate critically the information in the source, and consequently the resulting analysis may differ from the phonological inventory assumed in the source.
>
> (Maddieson 1984: 161)

In some cases where contrast is not clearly established for particular marginal segments, these segments are included in the phoneme inventory:

> The remaining issue concerning the size of the inventory of segments has to do with inclusion or exclusion of segments with more or less marginal status. [...] In other instances, segments which might from certain points of view be considered marginal are included without any indication of a distinctive status.
>
> (Maddieson 1984: 162)

On this basis we assume that glides in French are considered among these marginal phones and as such contrasts like those offered by Léon (1978) are evidence enough of contrastivity. In any case, glides are included in the phoneme inventory of French.

(79) Phonemic status of glides
/w/ s'oppose à /ɥ/ et à /j/ dans des mots comme:
souhait – suer – scier; roué – ruer – riez
/swɛ/ - /sɥe/ - /sje/ /rwe/ - /rɥe/ - /rje/ (Léon 1978:32)

Of course these same contrasts cannot be established between glides and their corresponding high vowels, suggesting that these segments are in fact in complementary rather than contrastive distribution. We would also like to note the absence of diphthongs (true diphthongs /wa/, /ɥi/, for example) from the inventory for French. Maddieson

explains that establishing the phonological status of diphthongs presents a number of difficulties:

> Relatively few languages are considered to have phonologically unitary diphthongs under the criteria used in UPSID [...] Obviously a much larger number of languages permit sequences of juxtaposed vocalic segments which might be considered phonetically to be diphthongs, or have diphthongal sounds which arise allophonically. Because diphthongs are so frequently derived in this way rather than being underlying segments, UPSID does not provide a good basis for analysis of phonetic patterns of analysis.
>
> (Maddieson 1984: 133)

The exclusion of historical diphthongs from the phonological inventory of French implies then that these diphthongs are considered to be phonetic (consisting of two independent vowels). Maddieson explains in more detail criteria for the inclusion of diphthongs in the phoneme inventories of the languages analysed (cf. Maddieson 1984 section 10.3 p.161), but it will suffice to say that, by excluding true diphthongs from the inventory of French, these have been treated as derived sequences of glide plus vowel and not as distinctive autonomous units.

Regardless of their phonemic status, we feel a study of the generalisations observed in glide distributions across languages is nonetheless merited. Glides by their inherent link to high vowels are most often treated simply as phonetic variants of the high vowel phonemes of a given language.

In a discussion of the most commonly occurring consonants and their correlation to the overall inventory Maddieson finds that “[l]anguages are most likely to have 2 liquids and 2 vocoid approximants (41% and 72 percent respectively)” (1984: 12). In a chart of the model inventory (containing the twenty most commonly occurring consonants), the two vocoid approximants included are the palatal /j/ and the velar bi-labial /w/. In fact, as defined by the UPSID these two glides are among the thirteen most common consonants.

Amongst the two most common glides /j/ and /w/, distribution data point to an important implicational relationship which can be characterised as a significant universal implication. For 240 languages with /w/ occurring in their inventory 226 (71.3 percent) also have /j/. Thus in only 14 languages (4.4 percent) does /w/ occur without /j/

(Maddieson 1984:92). However, the segment index lists only 238 languages with /w/, (Maddieson 1984: 246). Conversely in 273 languages (according to table 6.1 Maddieson 1984:92) with /j/, 47 (14.8 percent) do not have /w/. Maddieson discusses the statistical significance of this observation in the UPSID data as well as some earlier findings from the *Stanford Phonology Archive* (SPA):

> In the UPSID data, therefore, the occurrence of /w/ usually implies the occurrence of /j/ in the same language […] Nonetheless their claim [Stephens and Justeson 1979] that there is a statistically significant tendency for /w/ to occur only if /j/ also occurs is confirmed by our data (significance from χ^2 better than .001).
> (Maddieson 1984: 92)

It should not be surprising then that this implicational relationship exists for high vowel inventories as well. First with respect to the implicational correlations for high vowels and their corresponding glides, though not a categorical finding, the UPSID data show that if a language counts in its inventory a given glide it will almost certainly include the corresponding high vowel. According to Maddieson "[t]he greater frequency of /i/ is undoubtedly a predictor of the greater frequency of /j/. There are, however, for both /j/ and /w/ a few languages which have the approximant but lack the corresponding vowel" (1984: 94). Only eight languages in the UPSID database are said to have /j/ without /i/ and perhaps most surprising is that it is reported that for a full 23 languages with /w/ have no corresponding /u/ (three times as many as for /j/). Finally, of the four languages reporting to have /ɥ/, three also have /y/.

With respect to an implicational relationship in high vowel inventories we look at the general patterns of distribution within vowel inventories which show a general preference for high vowels. Concerning high vowels, the vast majority of languages count /i/ and /u/ among their vowels while /y/ is highly marked.

A closer look at the UPSID inventories shows that the statistics bear this out. The high vowels /i/ and /u/ are among the top three most common vowels in the vowel inventories. Perhaps not surprising since the smallest vowel inventories count only three vowels /i/, /a/ and /u/ which correspond to the "acoustic extreme" of the vowel triangle, still /u/ is dispreferred amongst the two high vowel extremes. According to Maddieson:

> The 3 vowels at the corners of the conventional vowel triangle, /i, a, u/, are the most widespread, but note that there are 24 fewer languages with /u/ than with /i/. These three vowels might be expected to be equally favored, because they each lie at an acoustic extreme. The low vowel /a/ has the highest first formant, /i/ and /u/ have the lowest first formant but /i/ has the highest second (and third) formant, whereas /u/ has the lowest second formant. However, a contributory factor to the relative disfavoring of /u/ may be the lower amplitude typical of /u/.
> (Maddieson 1984: 125)

The most common inventory in the UPSID is a five vowel inventory (21.5 percent; 68 languages) which by implication includes the three vowels of the acoustic extremes as well as two mid-range vowels /e/ and /o/. By mid-range Maddieson includes all vowels in the general area of the two mid-vowels /e/ and /o/. In fact, it must be said that over 55 percent of the languages surveyed have 7 or fewer vowels. The most common vowel inventories are the smallest (3-7 vowels) so that in this regard as well French clearly distinguishes itself (one of only ten languages or 3.2 percent) with an inventory of 16 vowels (12 distinct vowel qualities plus four nasal vowels). Observations of the general distributions based on vowel quality show that, while there are (in sum) more mid-range vowels than high or low vowels, front vowels 1019 significantly outnumber back 964 and of these the high front vowel is the most prevalent 481 to 448 high back. “Unrounded vowels are considerably more frequent than rounded vowels, namely 1569 to 981 or 61.5% to 38.5% “ (Maddieson 1984:124).

These statistics are particularly interesting because they illustrate the universal tendency for at least a three-way distinction across vowel inventories (the three extremes of the triangular vowel space) but beyond that a strong preference for high front vowels followed closely by the high back rounded vowel. The front rounded follows much more distantly only four of 317 languages include the front rounded vowel according to Maddieson.

Taking the 317 languages and these distributional statistics to be representative we propose that certain generalisations be made. Despite their consonant-like distribution (glides are largely considered non-nuclear when resulting from suffixation), presence of glides in phoneme inventories appears inherently linked to the presence of corresponding high vowels (presence of /j/ generally implies the presence of /i/, for example). The

generalisations that fall out of the statistics for glides and their corresponding high vowels can be represented as an implicational (markedness) hierarchy (80):

(80) Distributional Generalisations (markedness hierarchy)

a. High Vowels: i » u » y and then for the
b. Glides: j » w » ɥ.

A distributional hierarchy can also be interpreted as a scale of markedness so that the high front unround /i/ would be the least marked followed by the high back rounded /u/ as more marked and finally the high front rounded /y/ being the most marked of the series. Similarly goes the hierarchy for the corresponding glides so that /j/ is the least marked followed by the more marked /w/ and last and most marked is /ɥ/.

In the following section we will look at the lexical frequencies for high vowels and glides in French. We show that with respect to lexical frequency in French, high vowels and their corresponding glides demonstrate the same hierarchical pattern as has been observed in cross-linguistic distribution.

4.5.3 Lexical frequency and distribution of French glides

We have seen from the UPSID data that French, with sixteen distinct vowels and three distinct glides, is hardly a typical language with respect to its phoneme inventory. In this section we examine the implications for glide frequency in the French lexicon. Given the hierarchical relationship that we observed in the cross-linguistic data for distribution of glides, it should be no surprise that glides in French pattern similarly so that the same hierarchical pattern is observed in frequencies in the lexicon. Carton lists the phoneme inventory for French by frequency where first is the most frequent and 35th is the least frequent. According to this ranking the high vowels are ranked with /i/ in seventh position at a rate of 5.3 percent, next is /u/ at 17th at a rate of 2.5 percent and /y/ at 18th or 2.1 percent. As for the glides front non-round /j/ is first among glides at a rate of 1.8 percent which places in 21st position across the phoneme inventory followed by /w/ at a rate of 1.0 percent in 28th position and finally (seecond last in the inventory) is /ɥ/ occurring at a rate of 0.3 percent in 34th position (1974: 72). Glide frequencies in the

French lexicon show the same hierarchical relationship observed in the cross-linguistic data lending further support to the unmarked status of /j/ the more marked status of /w/ and the most marked status of /ɥ/.

Léon 1978 provides an analysis of phonemes that examines each by their most common orthographic transcriptions, and lexical frequencies by position within the word (initial, medial, and final). We examine the lexical frequencies of the three high vowels /i y u/ as compared to their corresponding glides /j w ɥ/ and then we will look at glide distribution within the word.

All three of the high vowels can be found in word initial, word medial and word-final position. Léon notes different frequencies for written versus oral registers, we note both values here where the first value is for written and the second value is for oral register (in case of a single figure noted the frequency is the same for each register). As we have seen previously the broad scheme supports the pattern observed in all the previous data: [i] is the most frequent of the three vowels at 5.6 percent; [y] at 2.71, 2 percent; and [u] at 2.8; 2.7 percent. Glides follow a similar albeit reduced pattern of frequency as compared to the corresponding high vowels: [j] occurs at a rate of 1 percent; [w] at 0.77; 0.9 percent; and [ɥ] at 0.71; 0.7 percent.

Glides of course do not share equal distribution across all positions in a word: "[g]lides typically occur in slots where consonants occur, although not necessarily in all such slots" (Tranel 1987: 115). The unrounded front glide [j] shows the widest distribution occurring in word-initial (*hier*, *iode*, for example) word-medial (*bien*, *riez*, and *billet*, for example) and word-final (*oeil*, *paye*, and *brille*, for example). The rounded glides [w] and [ɥ] also occur in word-initial and word-medial position but they do not occur in word-final position. Yod is the only glide to occur in word-final position.

Heap et al. (1992) study of the variable realisation of elision and liaison for an array of glide-initial words: They provide a list of all GV combinations possible in word initial position (81).

(81) Word initial Glide + Vowel combinations in French

/j/: [ja], [jã], [je], [jø], [jɔ], [jɔ̃], [ju], [jy], [ji]

/w/: [wa], [wɛ], [wɛ̃], [we], [wi], [wɔ̃]

/ɥ/: [ɥe], [ɥi]

(Heap et. al. 1992:175 cf. (13))

Among these combinations are the historical diphthongs [wa], [wɛ̃], and [ɥi]. The data for the word-initial glides in (81) show the same typological pattern that we have observed for all other aspects of glide distribution in French. A pattern noted by these authors, "[n]otons que le yod initial est le plus nombreux des trois non seulement en termes d'entrées lexicales, mais aussi en termes du nombre de combinaisons possibles avec des voyelles différentes." (Heap et. al. 1992: 175).

Whether in cross-linguistic typologies, lexical frequency for French, or across positions within the word, or syllable, yod enjoys the least restricted distribution, the highest frequency in the lexicon and the widest combinatorial possibilities in word-initial position in French. In addition, in an empirical study of the variable behaviour of these word-initial GV sequences Heap et al. find that in matters concerning variable elision these GV sequences do not conform to the hypothesis put forth by Kaye and Lowenstamm (1984): "Il faut donc nous demander s'il existe des preuves indépendentes qui suggèrent que toutes les paires semi-voyelles + voyelles en (14) constituent en fait de 'vraies' diphthongues d'après les critères de K&L." (1992: 182).

We wonder if the variability observed in these word-initial GV sequences is better accounted for by their inherent markedness. Our own empirical study shows that cross-vocoid distributions may be a correlate of the degree of variability observed (cf. Chapter 6).

Word medial glides can be divided into two categories: lexicalised and derived. The lexicalised contexts (82) occur in monomorphemes and as such they tend to show much less variability whereas the derived contexts (83) are found in contexts where the triggering environment for glide formation is morphological.

(82) Lexicalised monomorphemes with glides

/j/: [je], [jɛ], [jø], [jɔ], [jɛ̃], [ju], [jy]
/w/: [wi], [wa], [wɛ], [wɛ̃],
/ɥ/: [ɥe], [ɥi], [ɥɛ̃], [ɥɛ], [ɥa],

Derived forms show fewer combinations and with more balanced distribution since these are restricted by the vowel initial morphemes that exist in the language.

(83) Derived forms resulting in glides

/j/: [je], [jɛ], [jø], [jɔ̃]
/w/:[wi], [we], [wɛ], [wɛ̃],[wa], [wɔ̃]
/ɥ/:[ɥe], [ɥi], [ɥɛ], [ɥɔ̃]

In this section we have discussed the notion of linguistic universals and their implications for markedness. We have demonstrated some of these implications through the examination of the cross-linguistic typology of glides as well as some distributional implications for glides and their corresponding high vowels in French that leads us to conclude that phonemic status aside, a study of glide typology shows that glides are inextricably linked to the presence of the corresponding high vowels in a phoneme inventory.

We turn now to the subject of glide phenomena in child-language acquisition. We have seen in the preceding sections that there is much evidence to point to a link between glides and the corresponding high vowels. In this next section we discuss evidence from gliding phenomena in early acquisition that indicates a link between glides and certain consonants with which they share some gestural similarity. We examine very briefly the phenomenon of cluster simplification and liquid substitution in child-language acquisition.

4.5.4 Markedness and child language acquisition

Child language acquisition may offer some insights into markedness through observations about order of acquisition amongst segments. Children acquire different segments and different combinations of segments (simple versus complex syllable onsets, for example) at different stages and in different ways according to the particular language in question. Acquisition of various segments, combinations of segments or syllable

structures may be said to fall out of the markedness of that particular element within the grammar of a particular language. With respect to glides, however, it cannot be said that child language data follows as neatly as the implicational hierarchies observed in the adult grammar.

In matters concerning glides in child language, there are some glide phenomena observed cross-linguistically that may inform our understanding of glide phenomena in adult grammars, as Goad observes:

> There is a large body of evidence which reveals that children's early productions are segmentally and prosodically unmarked (e.g., Jakobson, 1941/68; Stampe, 1969; Gnanadesikan, 2004). Following from this observation, a commonly-held view has been that these unmarked patterns must reflect early grammatical organization.
>
> (Goad 2006: 104)

Goad explores this question of child grammars demonstrating "rogue" behaviour as compared to the adult grammars they target. Of particular interest to us is the phenomenon of glide substitution in child language production of liquid consonants. This is a well-attested phenomenon for children wherein the lateral approximant /l/ is to be realised as [j] in onset position but as [w] or a similar vowel in coda position well as in other Germanic languages (Inkelas and Rose 2008: 721). According to Goad, "after the period when consonant+liquid (CL) clusters are reduced to singletons, many children learning West Germanic languages, at least, go through a stage during which the liquid in all such clusters is reportedly realized as a glide (G), e.g., target [trɪp] → [twɪp] 'trip' and [pleɪ] → [pweɪ] 'play'" (2006: 104).

In a personal communication Yvan Rose writes:

> Generally, variation is observed between w and j, though children tend to be better with labial w than alveo-palatal j. As for ɥ, the late acquisition may be reflecting the relative complexity of the articulatory combination involved (combination of labial and lingual gestures with tight timing required). More generally, front rounded vocoids tend to be disfavoured across phonological systems. ɥ simply falls into the same picture. More concretely, I do not think that j (yod) » w (oué) holds at all as an encompassing generalization in child language development.

Glide formation can be viewed as a strategy invoked to repair or preserve the unmarked (or less/least marked surface structure according to the language-specific phonological grammar).

These observations of bring us to a discussion of glide data in Optimality Theory (OT). In the following section, we discuss the Optimality Theoretical approaches for glides in French and other languages.

4.6 Optimality Theoretical treatments

We begin our survey of glide treatments modeled in Optimality Theory with Rosenthall (1997) whose cross-linguistic survey demonstrates the pervasive and universal nature of glide formation as a strategy to resolve language-specific markedness and faithfulness constraints thereby, maintaining optimal surface structures according to the phonological grammar. For French in particular the constraints involve aspects of syllable well-formedness interacting with faithfulness to input.

Modelling grammar as a language-specific hierarchy of constraints based on universal relationships of markedness and faithfulness, OT eliminates any need to model the operation of processes and determine their exact representation (rules, levels and domains). Durand and Lyche observe:

> La phonologie d'une langue est réduite à une hiérarchie spécifique de ces contraintes et les alternances observées sont le résultat de leur interaction conflictuelle. Une fois posée une hiérarchie dans une langue donnée, la syllabation observée sera le fruit d'un choix optimal, de la satisfaction maximale des contraintes, et les glissantes apparaîtront à la surface lorsque les voyelles hautes seront analysées comme des segments non moriques afin de satisfaire au mieux les contraintes dominantes.
>
> (Durand and Lyche 1999: 52)

As is observed in preceding sections, Glide Formation in French is an optional process, a strategy motivated by a universal dispreference for the surface realisation of hiatus (VV). For French, when the first segment of a VV sequence is a high vowel (HV), a glide (epenthetic or alternating with a corresponding high vowel) may occur rendering either a GV or VGV sequence in the output (though VV is not ruled out). The form that occurs

depends on the phonological grammar (rules, syllable structure templates or governing relationships). In the OT model each of these three outputs are found amongst a stock of all possible candidates to be evaluated against the most dominant constraints of the French grammar. When a VV sequence is selected as the winning candidate, Faithfulness to the input form dominates the constraint ranking. On the other hand, when either VGV or GV is selected, markedness that privileges a less marked structure (a CV syllable, for example) in a surface realisation is selected at the expense of a violation of a lower ranked faithfulness constraint.

In the following section we discuss various glide treatments in an OT framework.

4.6.1 Rosenthall (1997)

Glide formation is not exclusive to French. Rather, it is a cross-linguistic strategy employed in varying ways (both similar to French and distinctly different from French) by a great many languages, as discussed in Rosenthall (1997). Glide formation should be considered a widely used strategy in a variety of languages to align output forms with the least violable markedness constraints for a given grammar. This phenomenon and its many instantiations are demonstrated in Rosenthall (1997). In view of our discussion of glide typologies, their distributions, and the implications for markedness, we begin our survey of glide phenomena with one of the earliest treatments of glide formation modelled in an Optimality Theoretical (OT) approach. Rosenthall (1997) presents a cross-linguistic OT analysis of glide formation in data from a broad sample of languages in which he examines various instantiations of high vowel glide alternation. Rosenthall examines the various languages' grammars to demonstrate the implications for syllable structure, stress assignment, and underlying representations in the process of glide formation and the OT model. Rosenthall illustrates that each of these individual grammars and their unique outputs for high vowel~glide alternation are succesfully modelled in OT as the result of the interaction of markedness and faithfulness constraints.

Our purpose here is to draw attention the degree of variability observed in these cross-linguistic realisations of glide formation but also to highlight the relationship that pertains

between glide formation and the particular vowels in a hiatus situation as Rosenthall observes:

> There are a number of phenomena that affect underlying vowel sequences, such as vowel elision, nonmoraic parsing of a high vowel (traditionally called Glide Formation), or epenthesis. Furthermore, languages often exhibit more than one of these phenomena, in particular, high vowels surface nonmoraically and nonhigh vowels are elided. The above phenomena are all related insofar as they occur in order to maintain monophthongal vowels on the surface.
>
> (Rosenthall 1997:39)

Rosenthall's survey of glide formation in a sample of divergent languages is premised on the position that glide formation (in these languages, at least) is motivated by language-specific constraints which privilege monophthongal vowels in surface forms. As rosenthall observes, the varied surface forms that arise result from the strategies available in these divergent phonological grammars:

> A common pattern found in these languages is that a high vowel, when followed by another vowel, surfaces as its non moraic counterpart [glide]. Nonhigh vowels in the same environment show different phenomena cross-linguistically. The syllabification of underlying vowel sequences is examined in a number of languages.
>
> (Rosenthall 1997:3)

In a discussion to introduce the formalism of Optimality Theory, Rosenthall devotes a considerable amount of time to the demonstration of the interaction of markedness (syllable structure constraints) and faithfulness (PARSE constraints). Given the breadth of languages examined, each with their own grammar, Rosenthall details a varied list of compensatory measures resulting from GF. By comparison glide formation in French appears less costly than is observed in Rosenthall's sample languages.

In this account, Rosenthall resorts to evoking the hierarchy of prosodic domains: "Optimality Theory makes use of prosodic licensing through constraints which require the parsing of all phonological constituents i.e. segments, morae, syllables, feet. [...] Phonological units are said to be "parsed" insofar as they are properly incorporated into higher prosodic constituents" (Rosenthall 1997:18). For our analysis of French, we will be concerned almost exclusively with the syllabification level(s) and thus we require only a fraction of these constraints.

Before we discuss an OT analysis of French data we examine some of the implications for glide formation as observed in the data of other languages in Rosenthall (1997). This survey shows that there is a strong dispreference for VV (hiatus) sequences. However, the strategies observed cross-linguistically, which may include one or all of vowel elision, glide or other consonant epenthesis and glide formation, are realised with extensive variability depending on the vowel quality and the grammar of a particular language. Different from GF in French, VV sequences and high vowel~glide alternations may be realised by way of an extensive cast of variants (glide formation, secondary articulation, deletion with or without compensatory lengthening).

OT accounts of glide phenomena generally resort to mora theory to represent the change in status (nuclear to non nuclear) of a high vowel. According to mora theory, all vowels are assigned at least a single mora that can act as a measure of syllable weight (length). All segments in a Rime contain at least a single mora and as such when high vowels are realised as their non nuclear counterparts the implication is that the output will reflect the loss of that mora. In Rosenthall (1997) the principles of Mora Theory are represented by the OT constraints PARSE μ (parse all mora from the input) and V-mora (for every vocalic root node V_i there is a mora_i) and GEM-MORA (allow geminate segments to close syllables). Of most interest to our analysis of high vowel~glide alternations is the realisation of compensatory lengthening resulting from the loss of the mora-bearing segment. In some instances, when the high vowel with a co-indexed mora is realised as the non nuclear counterpart (glide) the mora is then transferred to the vowel immediately following which in turn triggers the surface compensatory lengthening.

As an example of the variety of candidates involved we present only a few of Rosenthall's tableaux: Etsako (84) and Luganda (85); these are followed by a summary in (Table 2) of all the phenomena surveyed by Rosenthall. These data demonstrate that in a grammar with a strong preference for monophthongal surface sequences, a secondary articulation as in 1b., can be optimal when markedness constraints dominate faithfulness constraints. In this grammar the constraints NO DIPH, FILL, ONSET are ranked above those faithfulness constraints that would penalise a secondary articulation SECARTIC or the loss of a mora PARSE μ.

(84) Etsako {NO DIPH, FILL, ONSET} » {SECARTIC, PARSE μ}

/ofia/	NODIPH	FILL	ONSET	PARSE PL	SECARTIC	PARSE μ
a. o.fia.	*!					
b. ☞ o.f^{y}a.					*	*
c. o.fi.a.			*!			
d. o.fi.Δa.		*!				

(Rosenthall 1997: 50)

The winning candidate is realised with a secondary articulation on the preceding consonant which necessarily implies the loss of a mora (84b.). We note that the three markedness constraints NODIPH, FILL, and ONSET, are not crucially ordered (shown by the absence of lines between constraints). Since all three suffer fatal violations, their ranking does not factor into this analysis. The winning candidate is realised with a secondary articulation on the preceding consonant which necessarily implies the loss of a mora (84b.).

Luganda data demonstrate an instance of glide formation where the highly ranked markedness constraints ONSET penalises hiatus (85ab.) and SYLL-SEG (if a root node rt_i is linked to the syllable node, then $*\mu_i$) penalises compensatory lengthening (85de.).

(85) Luganda

/ga+uala/	ONSET	SYLL-SEG	V-MORA	PARSE μ	NLV
a. ga.u.a.la.	*!*				
b. ga.ua.la.	*!				
c.☞ ga.wa.la.			*		
d. ga.wa<:>.la.		*!		*	
e. ga.wa:.la.		*!			*

(Rosenthall 1997: 76)

According to Rosenthall "[p]arsing the root node, as opposed to just the place node, is significant because it directly affects the mora assignment constraints. [...] where the glide is the sole member of the onset, because the root node must be parsed in order to parse the place node [thus] SYLL-SEG becomes relevant (1997:77). The winning candidate for the Yoruba grammar is realised with GF (85c.). We will see similar constraint rankings for glide formation in French in the next section.

In the following table we provide a summary of glide phenomena observed in the various grammars examined by Rosenthall. Table 4 below shows the variety of strategies employed in the grammar of the languages surveyed with examples of outputs to demonstrate their various consequences.

Table 4: Cross-linguistic survey of strategies employed in the resolution of /HV+V/

language	Strategies employed	data
Etsako	✓ Glide formation - secondary articulation before onset consonant	CV.V→$C^{G}V$ /ofia/ [of^{y}a] 'king' /alokui/ [alok^{w}i] 'cameleon'
Yoruba	X No glide formation ✓ Elision of first vowel in the series (even high vowels	CV.V→CV /ni+oko/ [loko][12] 'on the farm' /ti+emi/ [temi] 'mine'
Luganda	✓ Glide formation -Secondary articulation before onset consonant -compensatory lengthening of remaining vowel ✓ Elision of first vowel when non high ✓ Glide in Onset	CV.V→$C^{G}V$: /mu+ oyo/ [m^{w}o:yo] 'soul' /li+ato/ [l^{y}a:to] 'boat' CV.V→CV: /ka+oto/ [ko:to] 'foyer' dim. CV.HV.V→CVGV /ga+uala/ [gawala] 'girl'
Kimatuumbi	✓ hiatus -when the vowels in sequence do not involve a high vowel ✓ Glide in the Onset ✓ Glide formation -secondary articulation before an Onset consonant -compensatory lengthening	CV[-haute] + V→CV.V /ma+oto/ [ma.o.to] 'great fires' /ka+ula/ [ka.u.la] 'small frog' VH.V→ GV: /u+a+teleke/ [wa:teleke] 'you must cook' /i+ula/[yu:la] 'frogs' CV.V→$C^{G}V$: /mi+ oto/ [m^{y}o:to] 'fires' /ki+ ula/ [k^{y}u:la] 'frog'
Ilokano	✓ Glide formation - includes mid and mid-high vowels - secondary articulation when preceded by a onset consonant which may trigger - gemination of that preceding consonant ✓ Consonant epenthesis for low vowels in hiatus	CV[±high] + V→ CG[α anterior]V /masahe+en/ [masahyen] 'massage' /saŋo+en/ [saŋwen] 'face forwards' /babawi+en/ [babawyen] 'regret' $CV[±haute] + V→ C$$C^{G}V$ /sakdo/ [pag-sakd^{w}an]'place where water is fetched' source ? /subli/ [pag-subl^{y}an] 'place that one returns to' /luto+en/ [lutt^{w}en] 'cook (goal focus)' /baggi+en/ [bagg^{y}en] 'to have as one's own' /basa+en/ [basaʔen] 'read (goal focus)' /cyenda+an/ [cyendaʔan] 'marketplace'

[12] Rosenthall does not explain the change from /n/ to /l/ here except to say that it is a separate process and for more information readers are directed to consult Bamgbose 1966, Awobuluyi 1978, Pulleyblank 1988a,b.

The five languages studied by Rosenthall illustrate a broad array of data for cross-linguistic glide formation that provides some interesting insights about the phenomenon. The languages here show: No glide formation (Yoruba); Glide formation (Onset only, Kimatuumbi, Luganda); Glide formation as secondary articulation to a preceding onset consonant (Etsako, Luganda, Kimatuumbi, Ilokano). Glide formation is but one possible strategy employed by these languages to avoid hiatus. Other strategies here include: elision of the first segment as in Yoruba (even of high vowels since glide formation is not available), elision of the first segment (when non high) as in Luganda; and consonant (glottal stop) epenthesis when the first vowel in the series is low even when preceded by another glide thus creating a GG sequence (Ilokano). While French glide formation involves exclusively high vowel~glide alternations, cross-linguistically glide formation is not limited to HV+V sequences. In Ilokano for example the mid-high vowels both front and back undergo glide formation. Finally the various instantiations with strong implications for syllable position: glide (Kimatuumbi, Luganda); secondary articulation (Etsako, Luganda, Kimatuumbi, Ilokano); secondary articulation which may trigger gemination (Ilokano), show to what extent the phenomenon is tied to the syllable structure (markedness constraints) of a given language.

Rosenthall's survey has shown that glide formation is a widely used strategy to avoid hiatus for a divergent group of languages. As variable as French may be with its three-glide inventory and two different strategies: GF and epenthesis, relative to the various outcomes surveyed by Rosenthall implications for glide formation, in French syllable structure is comparably straightforward. French does not have secondary articulation (though some may argue for a higher level of coarticulation phenomena than is usually described), compensatory lengthening or gemination. Compare the outputs discussed in Rosenthall with the categories that we have established for glide data in French in (86):

(86) Three environments for derived glides in French

a. variable high-vowel (HV)/glide (G) alternations (HV→G) when followed by another vowel: (C)HVV→ (C) GV as in *tuer* [tɥe]
b. variable epenthetic glide (propagation of HV to a following empty onset): (C)HV_V→(C)HVG̲V as in *trouer* [tʀuwe]
c. invariable glides: G paye [pej]; travail [tʀavaj], GV (true diphthong proie [pʀwa], truite [tʀɥit)

We have shown that glide formation in French and cross-linguistically is motivated by a strong preference within the grammar to maintain desired syllable structure in the phonological grammar. We have briefly introduced the formalism of OT that provides a model of the interaction between syllable structure (markedness) constraints and faithfulness constraints in a particular grammar. It should be clear that glide formation is inherently linked to a language-specific strategy to maintain syllable-wellformedness. Glide formation corresponds to the dominance of markedness constraints (syllable well-formedness) over faithfulness constraints (strict correspondence to input).

We turn now to OT analyses treating glide formation in French.

4.7 Durand and Lyche (1999)

In this section, we discuss an OT account for cross-dialectal variation in glide formation in French. Durand and Lyche (1999) compare glide formation in two varieties of French: Midi French (MF), from the region of Pézenas, Hérault which is described as having a *Langue d'Oc* accent (southern Occitan regions of France) and Standard French (SF) the so-called norm of the *Langue d'oïl* regions of Northern France also referred to in this work as *la norme parisienne*.

Through their extensive examination of the glide formation data for HVV and HV+V from these two varieties of continental French, Durand and Lyche demonstrate that the two varieties can be distinguished one from the other in matters of syllable structure, domain of glide formation and by the segments that may show glide formation. We begin with a brief comparison of glide data as given by these authors. At first glance glide data in MF seem comparable to SF.

(87)	Standard French	Midi French
croix	[krwa]	[krwa]
pluie	[plɥi]	[plɥi]
groin	[grwɛ̃]	[grweŋ]

(Durand and Lyche 1999: 43)

All three glides /j w ɥ/ as well as the historical diphthongs are attested in Midi French. Glides in word initial position in MF behave ambiguously just as in SF: " il faut diviser ces mots en deux classes au regard de leur comportement dans les contextes de liaison, d'élision et de supplétion (par ex. *les[z]oiseaux vs. le(s) westerns ; l'oiseau* vs. *le western; bel oiseau vs. beau western*)" (Durand and Lyche 1999: 42).

On closer examination, however, the distribution of glides in MF proves more restricted as we can see below. Different from SF, OLj sequences are not prohibited, to the contrary, it appears that in MF these sequences are, at least in some cases, preferred.

(88)	Standard French	Midi French
grief	[grijɛf]	[grjɛf]
brièveté	[brijɛvte]	[brjɛvəte]
pliocène	[plijɔsɛn]	[pljɔsɛnə]

(Durand and Lyche 1999: 43)

The forms in (88) involve glide formation in a monomorphemic environment. Durand and Lyche find that the greatest variability (for HVV) occurs in monomorphemic forms.

> Même si nos locuteurs de FM prononcent les mots ci-dessus [(à savoir [bjoloʒi], [djabɛtə], [pjɔŋ], etc.)] avec synérèse, on ne peut exclure que d'autres locuteurs préfèrent la diérèse dans certains de ces mots. Et, pour nos locuteurs, la diérèse est habituelle là où l'usage standard favorise la synérèse : *Lia* [lija] (prénom), *liane* [lijanə], *liasse* [lijasə], *lion* = *Lyon* [lijɔŋ], *miette* [mijɛtə], *via* [vija], *mouette* [muɛtə], *luette* [lyɛtə], *nuage* [nyaʒə]. Un marquage lexical semble s'imposer.

(Durand and Lyche 1999: 43)

In the polymorphemes, they observe a conservative/innovative split amongst speakers of Midi French such that innovative speakers show a tendency towards generalised dieresis after an OL sequence in derived forms (MF conservative: [dizje] [kablje] [etydije]

[plije] versus MF innovative [dizje] [kablije] [etydije] [plije], cf. Durand and Lyche (1999: 46) in some forms whereas the conservative speakers of Midi French show a preference for dieresis. This conservative~innovative distinction is particularly evident when an HVV sequence occurs root internally or in monomorphemic forms.

(89)	Standard French	Midi French (conservative variety)
	lier [lje]	[lije] (with transitional [j])
	tuer [tɥe]	[tye] (no transitional [ɥ])
	nouer [nwe]	[nue] (no transitional [w])
	oublier [ublije]	[ublije] (with transitonal [j])

(Durand and Lyche 1999: 43)

In (89), we observe that, where contexts involve non high vowels, the MF grammar shows a preference for hiatus (dieresis without epenthesis). Epenthesis is exclusive to iV sequences.

> Le fait que certaines séquences après OL interdites en FS soient possibles en FM (ex. *sablions*) trouve à première vue une explication simple : entre autres, l'ensemble des noyaux complexes est plus étendu en FM qu'en FS. Les suffixes flexionnels verbaux *-ions, -iez*, les suffixes dérivationnels *-ier/-ière* (*sablier, peuplier, meurtrière*) et *-ien* (*sartrien*), sont à considérer comme des noyaux complexes. Ce qui reste à expliquer est l'absence de synérèse entre une base se terminant par une voyelle haute et un suffixe à initiale vocalique.
>
> (Durand and Lyche 1999: 49)

Durand and Lyche demonstrate that, for Midi French, glide formation is not conditioned by the phonological environment (preceding environment and syllable structure) as we have discussed for Standard French. Instead they observe that glide formation in Midi French is restricted by a constraint that bans glide formation across a morphological boundary since this would require syllabifying material from two different morphemes (constituents) into a single syllable. Consequently dieresis is strongly preferred in MF. Glide formation is therefore limited to epenthesis when a root morpheme ending in a high vowel is affixed by a vowel initial suffix : *liez* /li+ e/ [lije] and *oubliez* /ubli+e/ [ublije] but *câbliez* /kabl+ije/ [kablje].

Table 5 below summarises these divergent facts for glide formation and the preference for syneresis versus dieresis in comparable data from SF and MF. Since gliding (epenthesis) concerns only the high vowel /i/ these data and the tableaux to follow treat

data involving the front non-round high vowel exclusively.

Table 5: Intragrammatical variation

Input	Standard French	Midi French (conservative)
câbliez /kabl+ie/	[ka.**bli.je**]	[ka.**blje**]
liez /li+e/	[**lje**]	[**li.je**]
cendrier /sãdʀ+ie/	[sã.**dʀi.je**]	[sã.**dʀje**]
oubliez /ubli+e/	[u.**bli.je**]	[u.**bli.je**]

(Durand and Lyche 1999: 44)

Based solely on surface forms such as *câbliez* [ka.**bli.je**] compared to [ka.**blje**] and *cendrier* [sã.**dʀi.je**] compared with [sã.**dʀje**], it appears as though glide formation is permitted following OL sequences in MF. However, if we compare input forms for *câbliez* /kabl+ie/ and *oubliez* /ubli+e/, what on the surface appears as identical gliding contexts, each with the same high vowel /i/ preceded by the same OL cluster /bl/, prove, in fact, to be quite distinct resulting in variable outputs.

Durand and Lyche point to the distinct morpheme roots (and the suffixes added to them), “[e]nfin, dans les mots complexes on observe des différences plus systématiques entre le FS et le FM. Dans les suffixés, la diérèse est obligatoire en FM à la frontière morphologique lorsque la base se termine par une voyelle haute” (1999: 43). Glide formation in MF is therefore morphologically conditioned. In conservative Midi French, root forms with a high vowel in final position do not undergo glide formation when a vowel initial suffix is added. Instead the high vowel in morpheme final position is preserved but epenthesis may occur to avoid hiatus. Epenthesis is, however, limited to yod, “[r]appelons qu’une terminaison en [i] déclenche un [j] de transition (*lions* [lijɔ̃)], *trions* [trijɔ̃)], *liage* [lijaʒ], *triage* [trijaʒ]) alors que [u] et [y] restent tels quels, ce qui produit des hiatus (*louer* [lue], *tuer* [tye])” (Durand and Lyche 1999: 49).

In their analysis, Durand and Lyche demonstrate that these two grammars (as well as the change in progress observed in the innovative speakers of MF) can be modelled in OT simply by ranking and then re-ranking of the various markedness and faithfulness constraints that interact in glide formation. The constraints for this analysis fall out of the

markedness of *TRI-C surface structures in a grammar as well as the faithfulness of output forms. The variability between the two grammars falls out of the different rankings of the crucial constraints according to violability in order to attain the winning output candidates.

To review, we summarise here the conditions for glide formation and their implications for syllable structure constraints in the phonology of French. Glide formation (high vowel alternations with corresponding glide) may optionally occur when a high vowel is the first segment of a VV sequence regardless of an intervening morpheme boundary but only when syllable structure permits (when not preceded by OL sequence). If glide formation is blocked by an OL sequence then glide epenthesis may occur (also optionally). Otherwise when none of the preceding outputs are realised a surface hiatus (VV sequence) may occur. Glide formation in French is captured in OT through the simultaneous evaluations of markedness and faithfulness constraints that fall out of these relationships. Since glide formation is motivated by a language-specific preference for unmarked CV syllables in the output, the various phonological environments described as triggering glide formation are modelled through the markedness constraints that would penalise output candidates containing highly marked VV or OLG sequences. Any failure in an input-to-output correspondence results from a lower-ranked faithfulness constraint violation (penalising deletion or addition of segmental material). The highest ranked are the least violable therefore the most crucial to the evaluation of outputs or winning candidates. The OT analysis by Lyche and Durand here follows from the analysis set out in Rosenthall (1994,1997) including constraints motivated by Mora Theory to account for some blocking effects. They propose the following constraints for their analysis of the data from French and Midi French.

(90) Markedness Constraints

V-MORA	for every vocalic root node V_i there is a $mora_i$
ONSET	all syllables must have an onset
*TRI-C	syllable constituents cannot be ternary
*CG	no glides in the onset (no glide formation)
*BRANCH μ	no branching mora
* BRANCH-AMBI	μ branching mora must not cross syllable constituencies

(adapted from Durand and Lyche 1999: 54-55)

The markedness constraints that factor into an OT analysis of glide formation correspond to the constraints observed in the phonological grammar, most particularly syllable well-formedness constrants. Starting from the canonical unmarked CV syllable, constraints are ranked based on their violability within a particular grammar. We have seen already many of these OT constraints as part of our discussion of Rosenthall (1997). We begin with V-MORA, which according to Mora Theory of this time must operate after the lexicon (after input) and stipulates that every vowel be assigned at least one mora depending on the particular grammar. Within the hierarchy of constraints the un-dominated V-MORA constraint would prevent the realisation of any glide in the output while for glide formation to operate the corollary (dominated V-MORA) would be required. ONSET is a markedness constraint that requires that all syllables have onsets. *TRI-C, which bans ternary branching onset, is a constraint from the *COMPLEX family of constraints which penalises all branching constituents and ternary branching is doubly penalised. If ONSET dominates hiatus will not occur. We will instead observe either epenthesis or glide formation. If glide formation with association to preceding onset does occur then *CG (which bans Consonant+Glide from onsets) is violated. However, when an OLG sequence is realised in SF then *BRANCH μ, which prohibits a single mora being shared in the nucleus, is violated (for clarity of the presentation these authors stipulate the exception of the three true diphthongs of French [wa], [wɛ̃], and [ɥi]). In order to constrain the context where these exceptions might be permitted the authors resort to *BRANCH-AMBI μ, "[s]elon [*BRANCH-AMBI μ], une more branchante ne peut pas traverser une frontière morphologique et n'est autorisée qu'à l'intérieur d'un morphème" (Durand and Lyche 1999: 55).

Given that the data attest to frequent glide formation these final constraints must be dominated in the ranking. The faithfulness constraints penalise variability in input to output correspondence.

(91) Faithfulness Constraints

DEP-IO	All output segments must correspond to input segments (penalises epenthesis)
MAX-RF	all features of a stem final segment in the input are represented in the output (penalises glide formation)

In the following sections, we discuss the interactions between these various constraints in an OT analysis of glide formation. Lyche and Durand (1999) show us that these two distinct grammars are modelled with efficiency and elegance in the OT framework.

4.7.1 OT analysis: Standard French

We begin with a demonstration of constraint ranking for Standard French monomorphemes (lexicalised forms) most commonly realised with a GV (syneresis).

(92) Standard French: Monomorphemic forms

/pie/	*BRANCH μ	ONSET	*CG	V-MORA
a. pi.e		*!		
b. ☞ pj-e			*	*
c. p-je	*!			*

(Durand and Lyche 1999: 57)

Note that while all the constraints in (92) are violated, the ranking of the constraints *BRANCH μ ≫ ONSET ≫ *CG, is crucial in determining the optimal candidate (92b.). The winning candidate *pj-e* with glide formation that creates a CG sequence syllabified to the onset (noted here with a hyphen) is selected because the constraint *CG is ranked lower than *BRANCH μ and ONSET which incur fatal violations by the candidates in (92ac.).

As described earlier the polymorphemic forms in Standard French show a preference for glide formation (syneresis) over glide epenthesis (dieresis) except when the VV sequence is preceded by an OL sequence. To model this aspect of the SF grammar Durand and Lyche resort to the constraint * BRANCH-AMBI μ to penalise branching (bi-constituent) mora and *TRI-C which penalises ternary branching onsets in derived contexts.

These same constraints and the same ranking account for the true diphthong of *loi* in (93) since *CG will eliminate the less optimal candidate (93b.) thus rendering the winning candidate (93c.).

(93) Standard French: Monomorphemic forms *loi*

/lu.a/	*BRANCH μ	ONSET	*CG	V-MORA
a. lu.a.		*!		
b. lw-a.			*!	*
c. ☞ l-wa.				*

(Durand and Lyche 1999: 57)

In the case of *loi* a monomorpheme in (93) the sequence -wa- (93c.) is a nuclear sequence, one of the three branching mora permitted in the nucleus by exception as stipulated in Durand and Lyche (1999: cf. (25)), therefore there is no violation of the constraint *BRANCH μ rather this candidate is selected as the winning candidate.

(94) Standard French: Polymorphemic *loua*

/lu+a/	*BRA-AM-μ	*TRI-C	* BRANCH μ	ONSET	MAXRF
a. lu.a				*!	
b. l-wa	*!				*
c. ☞ lw-a					*

(Durand and Lyche 1999: 57)

The words *loi* and *loua* are homophonous in Standard French and the distinction being made in (93-94) through the re-ranking of constraints reflects their differing morpheme and syllable structures. In (94) *loua*, in spite of an identical surface structure, is distinguished from monomorphemic *loi* as a polymorphemic in which the sequence [w.a] is derived from two different morphemes thus the candidate in (94b.) violates the *BRA-AM-μ constraint (by virtue of linking two morphemes to a single mora in the shared nucleus). The optimal candidate (94c.), by virtue of not being nuclear thus not being moraic, only violates MAXRF for a violation of input-to-output correspondance.

Some questions remain about the role of mora in French phonology. What is the role of weight in French? Certainly the data does not show consequences for the loss of morae such as is observed in the cross-linguistic data in Rosenthall (compensatory lengthening, for example). Given the lack of evidence to support the role of morae in the grammar of French we would suggest that glide data are more efficiently accounted for without mora.

We turn now to the account for the apparent blocking effect of a preceding OL clusters and the variable surface forms in *trois* [trwa] and *troua* [trua], for example (95).

(95) Standard French: monomorphemic *trois*

/tʀua/	*BRA-AM-μ	*TRI-C	* BRANCH μ	ONSET	MAXRF
a. tʀu.a				*!	
b. ☞ tʀ-wa					
c. tʀw-a		*!			

(Durand and Lyche 1999: 58)

These two words (though very similar to *loi* and *loua*) are not homophonous; they thus involve quite different outputs: *trois*, a monomorphemic word like *loi* is produced with syneresis [trwa], compared with *troua*, a polymorphemic word produced with dieresis like *loua*, but in which, the high vowel in hiatus does not undergo glide formation. Durand and Lyche follow Kaye and Lowenstamm (1984) such that in derived glides are syllabified to an empty onset position therefore a preceding OL sequence blocks glide formation. The winning candidate then is (96a.) the form that exhibits hiatus.

(96) Standard French: Polymorphemic *troua*

/tʀu+a/	*BRA-AM-μ	*TRI-C	* BRANCH μ	ONSET	MAXRF
a. ☞ tʀu.a				*	
b. tʀ-wa	*!				*
c. tʀw-a		*!			*

(Durand and Lyche 1999: 58)

ONSET being dominated renders any violation as immaterial. The candidate with hiatus wins out over the other candidates since they each suffer fatal violations. In (96b.) this candidate fatally violates *BRA-AM-μ whereas (96c.) fatally violates *TRI-C. As shown earlier for *loua* and *troua* the constraint *BRA-AM-μ dominates when these polymorphemic forms are evaluated thus preventing the undesirable nuclear glides *[lwa] and *[trwa]. Durand and Lyche observe:

> Si la voyelle haute est le premier élément d'un affixe *BRANCH-AMBI-μ n'est plus opérationnelle et *BRANCH-μ permet de départager les candidats. Lorsque le radical se termine par une seule consonne, la synérèse a lieu normalement mais elle est interdite s'il se termine par un groupe OL.
>
> (Durand and Lyche 1999: 158)

Note that epenthesis does not occur here.

Finally, we discuss some implications for suffixes of Standard French (-ier [je], for example). As discussed in Lyche (1979) and Morin (1976) various suffixes both

derivational (-ier [je]) and flexional (conditional 2nd pers. pl. –iez [je], for example) are accounted for through morpheme internal glide formation. Durand and Lyche offer the example of *charcutier* to account for such forms.

(97) Standard French: Polymorphemic *charcutier*

/ʃarkyt+ie/	*TRI-C	* BRANCH μ	ONSET	MAXRF
a. /ʃar.ky.ti.e/			*!	
b. /ʃar.ky.t-je/		*!		*
c. ☞ /ʃar.ky.tj-e/				*

(Durand and Lyche 1999: 58)

The winning candidate is (97c.) by virtue of a re-syllabified glide. Despite the seemingly identical surface forms, the fatal violations of * BRANCH μ ([je] is not one of the three permissible branching nuclei (cf. Durand and Lyche 1999: (25)) in (98b.) and the violation of ONSET in (97a.) rule out the other candidates. The highest ranked constraint *TRI-C does not factor into the account in (97) but it proves crucial to the forms where an OL sequence occurs in the preceding environment (*tablier*, below for example).

(98) Standard French: Polymorphemic *tablier* (hiatus)

/tabl+ie/	*TRI-C	* BRANCH μ	ONSET	MAXRF
a. ☞ /ta.bli.e/			*	
b. /ta.bl-je/		*!		*
c. /ta.blj-e/	*!			*

(Durand and Lyche 1999: 59)

In (98) we see that to account for a word such as *tablier* the banned OLG sequence in (98c.) causes a violation of *TRI-C. The winning candidate (98a.) maintains a hiatus since ONSET (penalising hiatus) is dominated by *TRI-C (banning OLG) and * BRANCH μ (banning GV in the nucleus) hiatus is optimal. The other two candidates are ruled out because of fatal violations as a result of glide formation. As mentioned above, however, hiatus and epenthesis are equally possible outputs in Standard French with variability depending on the vowel involved, as well as social factors such as idiolect, register and regional variety. It is important to note that Durand and Lyche analyse these different possible outputs separately (an OT tableau for each candidate). The next tableau, however, has been modified. We adopt the convention of unordered constraints to account for the possibility of variable outputs. Unlike (98), the candidates in (99) include

two equally possible outputs.

(99) Standard French: Polymorphemic *tablier* (both epenthesis and hiatus)

/tabl+ie/	*TRI-C	* BRANCH μ	ONSET	DEP-IO	MAXRF
a. ☞ /ta.bli.e/			*		
b. ☞ /ta.bli.je/				*	
c. /ta.bl-je/		*!			
d. /ta.blj-e/	*!				

Based on the analysis by Durand and Lyche (1999), and a modified Tableau in (99) using the convention of a broken line between constraints to indicate that constraints are unordered, thus allowing for the possibility of two equally optimal output candidates. Note also that we have added the faithfulness constraint DEP-IO since epenthesis ought to be constrained, but we point out that this account for the two winning candidates corresponds to the same analyses used in Tableaux (98) and (99). The two equally possible winning candidates are accounted for by using the unranked constraints ONSET and DEP-IO. The candidate in (99a.) with hiatus violates ONSET while the candidate in (99b.) with epenthesis actually satisfies ONSET but violates the lower ranked faithfulness constraint DEP-IO. According to Prince and Smolensky (1993) all candidates must be crucially ordered which would not allow an account of the variability observed in French. We follow the model of un-ordered constraints (cf. Anttilla 1997, Reynolds and Nagy 1997).

For ease of presentation, Durand and Lyche limit this part of their analysis to epenthesis and glide formation involving /i~j/. In the next section, we will discuss an alternative OT account proposed in Bullock (2002) that offers a model for the variability of epenthesis and glide formation across the full three-glide inventory based on the principles of the implicational hierarchy described earlier and evident in the data discussed here. Let us turn now to the analysis of the Midi French grammar in OT.

4.7.2 OT analysis: Conservative Midi French

The greatest single factor to distinguish the grammar of MF from that of SF in an OT account is the constraint ranking for each. In the SF analysis for example the faithfulness

constraint MAXRF (faithfulness to the stem final segment features) is low ranked and is consequently more violable without serious consequence. Therefore glides may occur in a winning candidate. The markedness constraints *TRI-C and *BRANCH μ (or *BRA-AM-μ) dominate in Standard French, however, to account for the MF data this situation (with respect to MAXRF) is reversed so that MAXRF dominates all other constraints while *BRANCH μ and ONSET remain unordered: MAXRF ≫*BRANCH μ, ONSET ≫ *TRI-C. We note also that the constraint *BRA-AM-μ becomes superfluous in the analysis (indicated with shading). Following Durand and Lyche and for the purpose of comparison we begin with the polymorphemic forms *loua* (100) and *troua* (101).

(100) Midi French: Polymorphemic *loua*

/lu+a/	MAXRF	ONSET	*BRA-AM-μ	*TRI-C
a. ☞ lu.a		*		
b. l-wa	*!			*
c. lw-a	*!			

(Durand and Lyche 1999: 60)

(101) Midi French: Polymorphemic *troua*

/tru+a/	MAXRF	ONSET	*TRI-C	*BRA-AM-μ
a. ☞ tru.a		*		
b. tr-wa	*!			*
c. trw-a	*!		*	

(Durand and Lyche 1999: 60)

For each of these polymorphemic candidates (regardless of the preceding environment) dieresis prevails; the winning candidate contains hiatus. The OT account for these two forms (*loua* and *troua*) is therefore identical. In contrast to SF where the markedness constraints dominate, the tableaux in (102-104) demonstrate that the dominance of faithfulness as defined by the constraint MAXRF is integral to derived forms in the MF grammar. The undominated MAXRF prohibits any change in the features of a stem-final segment (i.e. glide formation). Glide formation at a morpheme boundary is, therefore sub optimal. The dominant constraint MAXRF will not factor into the analysis of any monomorpheme or polymorpheme that does contain a stem final high vowel (102-103).

(102) Midi French: monomorphemic *trois*

/trua/	MAXRF	* BRANCH μ	ONSET	*TRI-C
a. tru.a			*!	
b. ☞ tr-wa				
c. trw-a				*!

(Durand and Lyche 1999: 60)

The crucial constraint for this next tableau is * BRANCH μ since only the licit branching mora -wa- can be determined optimal (102b.). In (103) we show an analysis for the polymorphemic form *tablier*. Here again there is no need of re-ranking the constraints required since * BRANCH μ and ONSET dominate *TRI-C.

(103) Midi French: Polymorphemic *tablier*

/tabl+ie/	MAXRF	* BRANCH μ	ONSET	*TRI-C
a. /ta.bli.e/			*!	
b. /ta.bl.je/		*!		
c. ☞ /ta.blj-e/				*

(Durand and Lyche 1999: 59)

Tableau (103) models the derived form *tablier* of MF where a surface OLG is observed but these authors remind us that this is not a case of a high vowel in the root but a high vowel + vowel in the suffix. The winning candidate is (103c.) where an OLG sequence, which violates the low-ranked *TRI-C, is optimal. Although optimal when the root morpheme ends in a high vowel as in (101), hiatus (103a.) is eliminated as because it fatally violates ONSET. The root is preserved while glide formation occurs in the suffix creating an illicit branching nucleus -je- which fatally violates * BRANCH μ.

We turn now to the MF innovative speakers who, like for Standard French, opt for the form in dieresis.

4.7.3 OT analysis: Innovative Midi French

Durand and Lyche observe that in an OT account of MF the innovative speakers' grammar does not follow from a complete shift towards the SF grammar: "[c]omme le FM innovateur n'autorise pas la synérèse dans *nié*, le changement en cours n'est pas attribuable à un relâchement du rôle des contraintes de fidélité, il s'agit plutôt de la

montée d'une contrainte de marque" (1999:61). The grammar of the innovative MF speaker is accounted for by simply raising one markedness constraint without displacing the dominance of faithfulness in the MF grammar. While the grammar of the MF innovative speaker may still more closely resemble that of the Conservative Midi speaker, the minor re-ranking however moves in the direction of the norm (towards the greater prominence of markedness) or more precisely a change from above. OT models this change in progress quite elegantly through a single shift in dominance so that *TRI-C is more highly ranked and dominates ONSET: MAXRF ≫ BRANCH μ, *TRI-C≫ ONSET while the faithfulness constraint MAXRF continues to dominate the hierarchy (104).

(104) Innovative Midi French: Polymorphemic *tablier*

/tabl+ie/	MAXRF	*BRANCH μ	*TRI-C	ONSET	DEP-IO
a. /ta.bli.e/				*!	
b. ☞ /ta.bli.je/					*
c. /ta.bl.je/		*!			
d. /ta.blj-e/			*!		

(Durand and Lyche 1999: 59)

We have seen here the crucial role of the stipulation allowing the three historical diphthongs as exceptions to the constraint *BRANCH μ. In (104bc.), they are identical but for their nuclear non-nuclear glide associations.

Analysis of variability in glide formation data in the grammar of Standard French versus Midi French demonstrates the significance of the distinction between historical (also nuclear or lexicalised) diphthongs and derived glide contexts.

In the following section we see that Bullock (2002) takes exception with this position which she attributes to a longstanding bias in the literature.

4.8 Bullock (2002)

According to Bullock (2002) the standard descriptions and formal treatments that are based on *le français de référence* (also Standard French) tend to perpetuate a general bias in the description of glide formation in French. This bias has lead to a considerable focus on the distinction between historical diphthongs (true or nuclear GV sequences) and the

derived diphthongs (apparent or non nuclear GV sequences). These accounts according to Bullock fail to consider the linguistic external factors which more adequately explain the variability observed for glide formation:

> In fact, the external evidence points in exactly the opposite direction; that is, speakers consistently parse *all* non-initial GV sequences as nuclear so that the distinctivity of the 'true diphthongs' amounts to nothing more than a slightly more liberal distribution in lexical items. Further, the interplay between syneresis and dieresis in actual pronunciation is quite variable and dependent on such things as regional variation, affective pronunciation and speech rate.
>
> (Bullock 2002:13)

In a survey of data from various sources, Bullock builds a case for the position that all GV sequences are treated as nuclear concluding that "[i]f external evidence of this sort is taken to be a good proving ground for testing the observational adequacy of a theoretical analysis, then the standard approach clearly comes up short" (Bullock 2002:17)[13].

In addition to her position on nuclear GV sequences Bullock proposes that the variability of glide formation attested in various regional varieties is better accounted for using an implicational hierarchy that reflects the role that markedness plays in glide distribution in French as we mentioned earlier in our discussion of glide typology (cf. section 4.4).

(105) Implicational Hierarchy for glide typology
j ≫ w ≫ ɥ or X ≫ Y ≫ Z

The implications of this hierarchy are that markedness is represented as a sliding scale where X is the least marked (or unmarked), Y is more marked than X but less marked than Z. In a Greenberg perspective presence of Z implies Y and X, and presence of Y implies X, and so it goes for the facts in French suggests Bullock. She extends the reading of the generalisation to apply to distribution and participation of glides in the various phenomena in French. Bullock points to further implications that fall out of the glide distribution in French such as margin fitness: "The choice of making the labial velar

---

[13] Sources such as Child Language acquisition (François et al.: 1977), language games (*Verlan* cf. Lefkowitz 1991, Scullen 1993, and Seguin and Teillard 1996); *Largonji des louchébèms* cf. Mandelbaum-Reiner 1991, Plénat 1985; *Javanais* Plénat 1991), as well as regional and dialectal variation (Morin 1976, Brasseur 1980)

more harmonic as a margin than the front rounded glide is weakly motivated by the slightly more liberal distribution [w] enjoys in French, but, more important, it may simply reflect a more universal markedness of the front round glide" (2002:18).

(106) Inverse Implicational Hierarchy in French Grammar

*ɥ ≫ *w ≫ *j

(Bullock 2002:18)

This implicational hierarchy can be applied to glides occuring in epenthesis or to glides occuring in the nucleus by GF. In either environment the rounded glides are always more marked than the non round. The implication here is that if a glide is observed it will be in the order of yod first then w followed by ɥ. We have of course discussed evidence to support this observation in our discussion of glide phenomena in French. Based on the evidence of glide phenomena in French, Bullock proposes a markedness constraint hierarchy to reflect the relationships observed for glides in syllable margins (107). Each of the constraints is independently ordered so that they may interact with other markedness and faithfulness constraints.

(107) Margin Fitness for glides in French

*M/ɥ ≫ *M/w ≫ *M/j

(Bullock 2002:18)

The hierarchies in (107-109) are motivated by universal typologies for glides but more specifically they reflect the facts observed for glides in the French data: "[t]he distributional facts alone suggest that there is a fairly high-ranked constraint in French that penalizes parsing glides, particularly the rounded glides, into syllable margins […]" (Bullock 2002:18).

Turning to the regional variability of glide formation, Bullock refers to Tranel's representation of the pan dialect variation in France. Adapted from Bullock (2002) following various sources (cf. Martinet and Walter 1973, Tranel 1987, Lyche and Girard 1995, and Durand and Lyche 1999), Heap and Kelly (2005) offer a modified version of the summary first proposed by Tranel (cf. 1987: 122). The table, to which Heap and Kelly add the affective pronunciation of media speech (cf. Lyche and Girard 1995), shows a relationship between the glide hierarchy and variable realisations with syneresis

versus dieresis across some of the regional varieties (cf. Table 6 below reproduced from Table 1 below).

Table 6: Pan dialectal variation

	nier	*nouer*	*nuer*	
a. Parisian French	[nje]	[nwe]	[nɥe]	synérèse
b. Northern French	[nje]	[nwe]	[ny.e]	diérèse
c. Midi French	[nje]	[nu.e]	[ny.e]	
d. Media French	[ni.e]	[nu.e]	[ny.e]	

Heap and Kelly (2005), based on Bullock (2002: 20) after Tranel (1987: 122), and Lyche and Girard (1995).

To this summary we might also add the Belgian variety of French with a two-glide inventory (*puis* *[pɥi]~[pwi]). According to Bullock: "[c]rucially some varieties of French even neutralize the labial-palatal approximant to the labial-velar" (Bullock 2002: 18) which further supports the more marked status of /ɥ/ in French.

To account for the GF triggering morphological environment /HV+V/, following Bakovic (2000: 23) (not unlike MAXRF in Durand and Lyche (1999)), Bullock proposes an IDENT constraint (108) that can be ordered amongst all other constraints to account for the various degrees of glide realisations in Table 3.

(108) S(tem)A(ffix)-IDENT[F]

> A segment in an affixed form [*stem + affix*] must have the same value of the feature as its correspondent in the stem of affixation [*Stem*].

(Bullock 2002: 21)

The independently ordered conjoined markedness and faithfulness constraint SA-IDENT ranked against the markedness constraint ONSET allows for a greater degree of variability within an account. Bullock accounts for each of the different registers represented in Table 6 as follows (note that we have added here the register d. Media French).

(109) S(tem)A(ffix)-IDENT[F]
Register a.
ONSET ≫ [SAIDENT[V] and *ɥ]: [SAIDENT[V] and *w]: [SAIDENT[V] and *j]
Register b.
[SAIDENT[V] and *ɥ] ≫ ONSET ≫ [SAIDENT[V] and *w]: [SAIDENT[V] and *j]
Register c.
[SAIDENT[V] and *ɥ]: [SAIDENT[V] and *w] ≫ ONSET ≫ [SAIDENT[V] and *j]
Register d.
[SAIDENT[V] and *ɥ]: [SAIDENT[V] and *w] :[SAIDENT[V] and *j] ≫ ONSET
(Bullock 2002:21)

Reducing the variable French glide distribution to positional fitness markedness hierarchies, there is no longer any need to make reference to morae in the analysis. In addition to the constraints above, Bullock proposes a series of peak fitness constraints that are specific to a particular grammar and for which rankings are determined separately for each licit GV based on the distributional facts; "The 'true diphthongs', then, are simply the least marked. Because this ranking is language specific, we must stipulate its order, placing [ɥi], [wa] and [wɛ̃] lower than all other combinations (2002:21). Bullock includes [jɛ] as a peak albeit a very marked peak "[i]n fact, there is good evidence from the history of French that the series [jɛ] at one time enjoyed the same liberal distribution as the so-called true diphthongs (Bullock 2002:21 original emphasis).

(110) Peak Fitness for glides in French

…*P/jɛ ≫ *P/wa, wɛ̃ ≫ *P/ɥi ...

"What the hierarchy in [110] recognises is that all complex nuclei are not equally harmonic." (Bullock 2002:22); as a syllable peak [ɥi] is least marked followed by [wa], [wɛ̃] and lastly [jɛ].

> The implication of the markedness constraints on branching nuclei is that there is no global constraint in French that prohibits CLGV series. Instead there is a distributional constraint on *what kind of branching nuclei* may follow a complex onset. In essence all surface GVs following a syllable initial onset are rising diphthongs, some are simply less marked in that position than are others.
> (Bullock 2002:22)

The approach that Bullock proposes eliminates need of ad hoc stipulations to account for branching nuclei. All GV sequences are admitted to the syllable nucleus as rising diphthongs that follow the markedness hierarchy in (110) above. With respect to nuclear glides the so-called true diphthongs are ranked with respect to one another. However this hierarchy would also be ranked with respect to other nuclear (peak) material. Anything ranked above the hierarchy of GV sequences is more marked and anything ranked below is less marked. Finally Bullock examines the implications for the complex onset preceding GV sequences. Acknowledging that throughout the history of French the implications for a preceding CL sequence have been in evolution, Bullock proposes a markedness hierarchy that reflects the interaction of complex onsets and branching nuclei over time. Based on observations made by the 17th century grammarian Ménage who "attributes the penchant for dieresis in words like *meurtrier* "murderer" to Corneille" (Bullock 2002:21), she proposes a constraint hierarchy that captures the pre-Corneille 17th century norm (men's speech at the time of Ménage) and the post-Corneille 17th century innovative speech (women's speech at the time of Ménage; cf. also section 1.5.2, 1.5).

(111) Ménage era production constraint

a. Pre-Corneille 17th (Men's speech)
ONSET ≫ *CL[jɛ] ≫ *CL[ɥi] ≫ *CL[wa]

b. Post-Corneille 17th (Women's speech)
*CL[jɛ] ≫ ONSET ≫ *CL[ɥi] ≫ *CL[wa]

(Bullock 2002:22)

Bullock cites indeterminacy as a reason for leaving out the diphthong [wɛ̃] adding that "[...] breaking down the global constraint of *CLGV into a markedness constraint on diphthong distribution allows for a good deal of variation and change." (Bullock 2002:22). It is still necessary however to account for the absence of CLGV in derived contexts. Bullock proposes that the constraints in (111) above be conjoined with the faithfulness constraint SA-IDENT (111), "[t]he affect of such conjunction is to elevate low-ranked markedness conditions within the hierarchy, prohibiting features which surface in lexical forms from appearing in derived environments" (Bullock 2002:22). Let

us look now at Bullock's account of derived *troua* (112) and the monomorphemic *trois* (113) using the conjoined constraints.

(112) Conjunction in derived forms: *troua*

/trou+a/		[SAIDENT[V] and *CC[wa]]	ONSET
a.	☞ [trua]		*
b.	[tr[wa]]	*!	

(Bullock 2002:23)

The conjoined constraint allows that when [SAIDENT[V] is violated and a CCGV does occur the specific outcome (CC[wa] here) is banned.

(113) Non-effect of markedness in the lexicon: *trois*

/trwa/		ONSET	*CL[wa]
a.	[trua]	*!	
b.	☞ [tr[wa]]		*

(Bullock 2002:23)

Bullock's proposal using conjoined constraints is however controversial. Not only is it extremely powerful, but it also seems just as much an ad hoc solution—targeting some very specific forms—as Durand and Lyche's exceptions to their branching mora constraint. Recall too that Durand and Lyche's account also proposed the use of a faithfulness constraint MAXRF similar to the SA-IDENT as well as the constraint *BRA-AM-µ in order to constrain derived outputs.

Durand and Lyche's study of the crucial distinction between Standard French and Midi French grammar as modelled in OT reveals the dominant role of Markedness over Faithfulness in the SF grammar versus the reversal of that dominance as demonstrated by the Faithfulness over Markedness ranking for the Midi grammar. Bullock shows us that in fact much of the glide formation data can be reformulated as markedness relationships in the grammar. In each case however the most difficult aspect of glide distribution is in accounting for the presence of CLGV sequences in the lexicon but their absence as an optimal surface form in derivations.

So it seems that the greatest challenge to any account of French glide formation is accounting for the presence of nuclear GV sequences (true diphthongs) when they are

prohibited in derived environments. In the next section we examine an alternate approach to the particular issue of derived environment blocking for OT, from Hall (2006).

4.9 Hall (2006)

Hall (2006) examines blocking effects in phonological and morphological processes across several languages. Of particular interest to us, he examines the phenomena of Derived Environment Blocking (DEB) in French glide formation to demonstrate the role of a general interaction of markedness (M) and faithfulness (F) constraints within the OT framework.

> Derived Environment Blocking (DEB) occurs if a phonological process is prevented from deriving a sequence of sounds [XY], but underlying (i.e. nonderived) /XY/ sequences are permitted to surface as [XY]. It will be argued below that Derived Environment Blocking effects can be captured in Optimality Theory in terms of a general ranking involving FAITHFULNESS and MARKEDNESS constraints and that individual languages invoke a specific instantiation of this ranking.
>
> (Hall 2006: 803)

The situation that Hall describes above is the exact situation in French for glide formation blocking following OL while OLG exists underlyingly in forms such as *trois*, for example. Surveying blocking effects cross-linguistically Hall observes a generalised scheme for grammars with these blocking effects. We include these various blocking phenomena and their effects as summarised by Hall below (note that the capitals on the left represent markedness (M) and faithfulness (F) and A and B are the different constraints):

	Constraint ranking:	Blocking effect:
(114)	a. MB » FB:	A process P of the form /Z/→[X]/ [Q] applies. MB penalizes [ZQ]. FB penalizes the change from /Z/ to [X].
	b. MA » MB:	A PB effect: Process P does not apply if the output would consist of a sequence of sounds [XY] (or [YX]). MA penalizes surface sequence [XY] (or [YX]).
	c. FA » MA:	A phonologically nonderived input /XY/ (or /YX/) surfaces as [XY] (or [YX]). FA penalizes any change which prevents /XY/ from surfacing as such.

(Hall 2006: 811)

According to Hall, DEB can be seen to fall out of a general ranking in the grammar: Faithfulness F_A » Markedness M_A » Markedness M_B » Faithfulness F_B. Other language specific data (German assibilation and glide formation as well as French glide formation) may involve other intervening constraints but will still follow this general pattern. Hall considers a large number of French glide data that spans initial and medial positions as well as representing all the possible preceding environments in order to build his case that French glide formation falls out of a language-specific instantiation of the general ranking for DEB effects.

(115) General ranking for French
DEP-μ (F_A) » *CLG (M_A) » ONSET (M_B) » MAX-μ (F_B).

(Hall 2006: 841)

As we have established in preceding sections, the greatest challenge to any account for glide formation in French is in capturing the distinction between true diphthongs and derived GV sequences (apparent diphthongs). Hall takes a different approach to the problem: "[m]y analysis of glide formation is based on the assumption that all vowels are underlyingly moraic […]. I assume that underlying glides [historical diphthongs] are represented as the corresponding high vowels which are not linked to a mora" (Hall 2006: 841-842). Morae are therefore assumed underlyingly and as such there is no need of V-MORA to assign a mora to each vowel. This seems to us an ad hoc stipulation for the purpose of accounting for true diphthongs in French that we find somewhat motivated, since after all, there is no compensatory cost to GF in French (unlike the various results observed in Rosenthall 1997 cf. 4.5.1). Now that the historical non-moraic high vowels are stipulated in the lexicon there is no need to account for the presence of [wa], [wɛ̃], and [ɥi] following OL sequences while derived glides are blocked in this very context.

Hall's stipulation simplifies the analysis of glide phenomena in French though morae remain a necessary part of his account, "I capture glide formation in French with the ranking ONSET » MAX-μ [...] my analysis also requires reference to a constraint penalizing output forms in which a root node is linked to a mora and to a following syllable node, i.e. NOBRANCH, which is ranked ahead of MAX-μ" (Hall 2006:834-835).

The crucial markedness constraints are *CLG penalising these tri-consonant sequences in the output, ONSET which penalises hiatus, and finally NOBRANCH which penalises "output forms in which a root node is linked to a mora and to a following syllable node" (2006: 834-835), in other words Hall proposes the addition of the NOBRANCH constraint for French in order to penalise glide epenthesis (comparable to the BRANCH-AMBI-μ in Durand and Lyche). Following this discussion, we will show an alternative analysis to illustrate that with the particular goal of considering variable outputs (as in the case of glide formation or glide epenthesis), we need not penalise these attested data for French and as a consequence we will not need mora theory.

We begin with the derived form *louer* (116). Just as in all preceding accounts glide formation is accounted for with a markedness dominant ranking. Hall gives us unordered ONSET and NOBRANCH that dominate MAX-μ. This ranking penalises marked structures like hiatus or epenthesis at the cost of glide formation. In this aspect, Hall's analysis offers little different from the preceding analyses. The constraint ranking in (116) corresponds to the lower end of the general ranking scheme proposed by Hall (115). We see more of Hall's general ranking as he progresses. While occasionally a new constraint is required, at no time is re-ranking of the general schema given in (115) required.

(116) *louer*

/lu-e/	ONSET	NOBRANCH	MAX-μ
a. lu.e	*!		
b. ☞ lwe			*
c. lu.we		*!	

(Hall 2006 (36))

According to this account which shows faithfulness constraints dominated by markedness constraints the winning candidate is (116b.), the form with glide formation (syneresis). The candidates in (116a. hiatus) and (116c. epenthesis) are excluded since each suffers a fatal violation of the markedness constraints that dominate. While Hall acknowledges that glide formation may be optional (at least for i/j), in (116) above GF is represented as categorical. He also represents epenthesis for all three glides stating:

> In French phonology it is usually assumed that there is a palatal glide between the [i] and following vowel, e.g. [plije] and not [plie] (see, for example, Tranel 1987). For other authors (e.g. Kaye and Lowenstamm 1984) there are homorganic glides between [u]/[y] plus vowel sequences as well. I follow the latter approach in my analysis.
>
> (Hall 2006: 835, note 31)

Crucially in any OT account of derived forms, Markedness must dominate Faithfulness to allow for glides as optimal output candidates. However, instances of OL sequences in the preceding environment are seen to block GF. A new markedness constraint banning an OLG sequence is required to constrain the process of GF in French grammar. The following tableau (117) demonstrates the blocking effects of a preceding OL cluster on glide formation with the addition of the constraint *CLG for the verb *clouer*.

(117) *clouer*

/klu-e/	*CLG	ONSET	NOBRANCH	MAX-μ
a. [klu.e]		*!		
b. [klwe]	*!			*
c. ☞ [klu.we]			*	

(Hall 2006 (42))

Hall's tableau shows unordered *CLG and ONSET (since the result is unchanged regardless of which constraint dominates). We note however that if the tableau were to show unordered ONSET and NOBRANCH (see 118 below), two output candidates are possible (118a. hiatus and 118c. epenthesis) allowing us to model more variability in the realisations.

(118) *clouer* (revised version)

/klu-e/	*CLG	ONSET	NOBRANCH	MAX-μ
a. ☞ [klu.e]		*		
b. [klwe]	*!			*
c. ☞ [klu.we]			*	

With unordered constraint ranking as shown in (118) above, we observe that ONSET is dominated therefore hiatus (dieresis) is potentially an optimal surface structure when the alternative, glide formation, is blocked by a *CLG constraint. We note that in with the addition of the *CLG constraint, the significance of the faithfulness constraint MAX-μ is greatly diminished in this analysis of the word *clouer* (indicated by the darker shading). It

remains however as part of our demonstration of a progression towards Hall's general ranking.

Looking at the monomorphemic *trois* (119), the candidates showing glide epenthesis and hiatus (119ab.) incur fatal constraint violations although these are both optimal candidates in (118). The faithfulness constraint DEP-μ must dominate the ranking in (119), since the output is identical to the input. Any potential variability in the glide realisation must be constrained. Derived forms with simple onsets (C) as in (116-117) show a constraint ranking that corresponds to the bottom end of Hall's general ranking (115). Derived environments that show blocking effects require the addition of yet another markedness constraint (*CLG) which serves to ban glide formation after OL. To further illustrate the significance of markedness contraints to derived environments Hall gives this tableau for the representation of a lexicalised diphthong *trois* [tʀwa] in (119).

(119) *trois*

/trwa/	DEP-μ	*CLG	ONSET	NOBRANCH	MAX-μ
a. [tru.a]	*!		*		
b. [tru.wa]	*!			*	
c. ☞ [trwa]		*			

(Hall 2006:840 (45))

In (119) we find the entire general ranking proposed by Hall in (115). The addition of the dominant faithfulness constraint DEP-μ is crucial to this analysis. Hiatus (119a.) and epenthesis (119b.) are prohibited since they incur fatal violations of the constraint DEP-μ. Once again we retain MAX-μ in the tableau to show its relative ranking, though it has no role in determining the winning candidate.

The general ranking proposed by Hall to account for the blocking effects of derived environments in (115) is indeed evident, however, this position depends entirely on the stipulation that some glides exist underlyingly and may appear therefore as the input form in these isolated cases. In all other instances, the low end of the general ranking proposed by Hall factors most prominently such that glide formation (118) can be seen as a result of the markedness dominant constraint ranking which contrasts with that of (118) which shows the winning candidate is the most faithful.

In his account of the lexicalised diphthong occuring following an OL sequence, Hall avoids the difficult entanglements of accounting for the nuclear diphthongs that show a banned OLG sequence by stipulating the diphthongs in the input. Thus, these are underlyingly non-moraic. While it does make accounting for these non alternating glides much easier, it is not to our mind substantially different from stipulating an historical diphthong (cf. Durand and Lyche 1999:25).

In the next chapter we undertake to study the state of affairs for glides in Modern French. We are motivated by many authors (Durand 2006, Morin 1974) who question the constant reuse of a common set of data in descriptions and analyses as well as by Encrevé who states: “[l]a réconciliation entre recherche empirique et recherche théorique ne se contente pas d’instaurer entre elles quelque coexistence pacifique, mais veut les enrichir l’une l’autre pour pousser plus loin l’exigence explicative” (1988: 14-15).

Our survey of OT treatments shows that glide formation data present complex issues for the grammar not only in French but also cross linguistically. Glide phenomena in French necessarily involve interactions of different Markedness constraints (in particular ONSET, but also .e.g. *CLG, NOBRANCH,) ranked above Faithfulness constraints such as Max μ, DEP-IO, MAXRF.

Chapter 5

5 Judgement Task as a methodological tool

Data for the present study come from the corpus *La phonologie du français contemporain* (PFC, Durand, Jacques, Bernard Laks & Chantal Lyche 2002, 2009). An international collaboration, the PFC project under the co-direction of Jacques Durand (Toulouse II), Bernard Laks (Paris X), Chantal Lyche (Oslo), and Marie-Hélène Côté (Laval) unites researchers from a variety of countries, who on the basis of a common protocol share the goal of recording over 500 speakers from approximately 50 different locations in the francophone world to attain the broadest coverage of varieties of contemporary French. The project coordinators give the following description:

> Le projet PFC part de la constatation qu'il est nécessaire de poursuivre le travail de description entrepris depuis au moins un siècle par tous les spécialistes de la communication parlée pour (1) fournir une meilleure image du français parlé dans son unité et sa diversité ; (2) mettre à l'épreuve les modèles phonologiques et phonétiques sur le plan synchronique et diachronique ; (3) favoriser les échanges entre les connaissances phonologiques et les outils du traitement automatique de la parole ; (4) permettre la conservation d'une partie importante du patrimoine linguistique des espaces francophones du monde, et ce en contrepoint aux corpus déjà constitués ; (5) encourager un renouvellement des données et des analyses pour l'enseignement du français.
>
> (Durand, Lyche and Laks, Bulletin PFC 1 2002)

At the core of the PFC project is the description of phonological systems with emphasis placed on establishing a phonemic inventory, the study of *liaison* and *le 'e' muet* while providing a general survey of French as it is spoken in francophone regions around the world at the beginning of the twenty-first century. Collaborators conduct field surveys according to the PFC protocol and contribute to the PFC database (http://www.projet-pfc.net/) orthographically transcribed interview excerpts, a reading passage, and a word-list which have been coded according to prescribed methodology for *liaison* and *e muet*. The PFC database provides collaborators and researchers access to audio recordings with orthographic transcriptions of French exhibiting several speech styles and representing a geographic diversity that until now has not been available. Teams that conduct the surveys and contribute the material as outlined in the

PFC protocol (http://www.projet-pfc.net/bulletins-et-colloques/cat_view/918-bulletins-pfc/919-bulletin-pfc-nd1.html) are free to pursue their own research interests beyond that outlined by protocol. The materials contained in the PFC database are therefore intended to contribute to a broader and more accurate description of French for phonological models and descriptive literatures.

5.1 The PFC: Methodology and Protocol

With a goal of contributing breadth to the data informing phonological description and theoretical discussions of modern French, PFC project coordinators have created a protocol that, with the four tasks, yields speech samples that represent a varying degree of formality. According to Durand, "[i]n the ideal case, the four types of recording as presented here yield an ascending scale of formality but the interpretation of the data needs a great ddeal of care. In particular, terms like 'formal' and 'informal' should not be taken at face value. As far as the conversations are concerned, much depends on the fieldworkers involved, their experience and their ability to set up proper conditions for the recordings. The two spoken styles are not always as sharply differentiated as one might have wished." (2006: 72).

We have therefore chosen to work with data from the PFC word-list, which elicits relatively formal speech from participants to create a judgement task (to serve as a methodological tool). Findings from this preliminary study are used to inform our analysis of excerpts of guided conversation (primary study) to examine the reality of glide realisations in Hexagonal French. Analysis of judgement task results centres primarily on determining the most prevalent acoustic qualities that distinguish an HVV realisation with syneresis from one realised with dieresis. Based on the unanimous findings from a panel of Hexagonal French consultants, we determine those cues that will contribute to our spontaneous speech study. In the sections that follow, we discuss the different tasks of the PFC protocol and illustrate how each task might be exploited in the study of glide contexts.

5.1.1 The PFC Word-list

The PFC word-list includes 94 words (listed below). The words included are familiar, relatively short, chosen as “those most often cited in the descriptive literatures and studies to ensure comparability with previous findings” (Durand 2006: 67).

The entire PFC word-list (94 words in total) is provided below:

Table 7: The PFC word-list

1.	roc	33.	liège	66.	peuple
2.	rat	34.	baignoire	67.	rauque
3.	jeune	35.	pécheur	68.	cinquième
4.	mal	36.	socialisme	69.	nier
5.	ras	37.	relier	70.	extraordinaire
6.	fou à lier	38.	aspect	71.	meurtre
7.	des jeunets	39.	niais	72.	vous prendriez
8.	intact	40.	épais	73.	botté
9.	nous prendrions	41.	des genêts	74.	patte
10.	fêtard	42.	blond	75.	étriller
11.	nièce	43.	creux	76.	faites
12.	pâte	44.	reliure	77.	feutre
13.	piquet	45.	piqué	78.	quatrième
14.	épée	46	malle	79.	muette
15.	compagnie	47.	gnôle	80.	piquais
16.	fête	48.	bouleverser	81.	trouer
17.	islamique	49.	million	82.	piquer
18.	agneau	50.	explosion	83.	creuse
19.	pêcheur	51.	influence	84.	beauté
20.	médecin	52.	mâle	85.	patte
21.	paume	53.	ex-mari	86.	pâte
22.	infect	54.	pomme	87.	épais
23.	dégeler	55.	étrier	88.	épée
24.	bêtement	56.	chemise	89.	jeune
25.	épier	57.	brin	90.	jeûne
26.	millionnaire	58.	lierre	91.	beauté
27.	brun	59.	blanc	92.	botté
28.	scier	60.	petit	93.	brun
29.	fêter	61.	jeûne	94.	brin
30.	mouette	62.	rhinocéros		
31.	déjeuner	63.	miette		
32.	ex-femme	64.	slip		
		65.	compagne		

The task of reading a word-list aloud elicits a formal style. The word-list includes five minimal pairs that contrast vowels (numbers 85-94) and a number of consonants in initial, medial and word-final environments to aid collaborators in establishing a phonemic inventory. Speakers are asked to read the list aloud including the number that precedes each word for future indexation purposes. Each word is then contained in a quasi carrier phrase, as in *vingt-cinq: épier* or *soixante-dix-neuf: muette*, mitigating some of the formality inherent in reading a list of words in isolation.

For the study of high vocoids involved in gliding contexts, the list can be reduced to the words shown in Table 8:

Table 8: All instances of lexicalised /GV/ and derived /HV+V/ contexts

6.	fou à lier	37.	relier	68.	cinquième
9.	nous prendrions	39.	niais	69.	nier
11.	nièce	44.	reliure	72.	vous prendriez
25.	épier	51.	influence	78.	quatrième
28.	scier	55.	étrier	79.	muette
30.	mouette	58.	lierre	81.	trouer
33.	liège	63.	miette		

As in the lexicon, the contexts involving high vocoids in the word-list are dominated by the high front unrounded vocoid /I/. Nonetheless, the PFC co-directors were careful to include contexts for all three high vocoids in a variety of environments. The reduced list above can be divided further into lexicalised (Table 9) and derived (Table 10) contexts:

Table 9: Lexicalised GV contexts

11.	nièce	58.	lierre	78.	quatrième
25.	épier	63.	miette	79.	muette
30.	mouette	68.	cinquième		

Note that the minimal set: 63 *miette*~30 *mouette*~79 *muette*, provides an excellent opportunity to compare each of three high vocoids in identical environments while the stimuli 68. *cinquième* and 78. *quatrième* allow us to compare a single lexicalised morpheme in different environments: preceded by a simple onset (C) and preceded by an complex onset with an OL sequence. Among the derived stimuli shown in the table

below, we find stimuli for all three high vocoids /I, U, Y/ in comparable environments including one stimulus for each high vocoid preceded by an OL sequence: /I/ 72. *vous prendriez,* /U/ 81. *trouer*, /Y/ 51. *influence*. In Table 10 we show all derived stimuli in the PFC word-list.

Table 10: Derived /HV+V/ contexts

6.	fou à lier	39.	niais	55.	étrier
9.	nous prendrions	44.	reliure	69.	nier
28.	scier	51.	influence	72.	vous prendriez
				81.	trouer

Derived stimuli shown above are similarly dominated by the high front unrounded vocoid /I/ as is observed for stimuli in lexicalised contexts. We find 8 stimuli involving /I/, and one each involving the rounded high vocoids /U/ *trouer,* /Y/ *influence*.

All HVV contexts identified here are word medial preceded by at least a single consonant with three instances of HVV preceded by OL sequences. These contexts are expected to show realisation with dieresis (with or without epenthesis): *pren<u>drions</u>* [dʀi(j)ɔ̃], *trouer* [tʀu(w)e], for example. Ten of the twenty words listed are considered monosyllabic since, whether they present a derived or a lexicalised HVV context, it is expected they are realised with syneresis as in *nièce* [njɛs]*, lierre* [ljɛʀ], for example; five words show HVV contexts in the final syllable of words considered to be disyllabic (*cinquième*, *relier*, for example). The remaining five are considered polysyllabic since the gliding context preceded by an OL cluster is expected to be realised with dieresis /V(G)V/: *pren<u>drions</u>*, *in<u>fluence</u>*, *é<u>trier</u>, pren<u>driez</u>* and *qua<u>trième</u>*.

Additionally, we note a gliding context observed in token 6. *fou à lier*. In the first part of the phrase *<u>fou à</u> lier* the high vowel of the word *fou* [u] is followed by a vowel initial context in the preposition *à* [a]. We observe that this HV#V context may result in glide formation so that *fou à* is realised as [fwa]; one of several possible realisations [fu.a;fu.wa;fwa]. This context, which we will refer to as a derived cross-word-boundary glide formation, is not generally discussed as a gliding context in descriptions of gliding phenomena. Recall from our previous discussions that it is described by Passy (cf. chapter 2) also Nyrop (1899: 216-217); Lyche discusses contexts for glide formations or

glide insertions (1979: 15-22, 327-328, cf. section 3.4.2). In discussions of the phonological word (PW) as a gliding domain, Hannahs (1995:1131), Tranel (1987: 119) and Johnson (1987: 893) discuss some cross-word contexts as evidence of a strong boundary expected to prohibit glide formation (cf. section 4.4).

Descriptions of gliding phenomena are generally concerned with word-medial contexts where a target sequence straddles a morpheme boundary: /HV+V/ resulting in glide formation: HV→ GV (realised with syneresis), or glide epenthesis: HV→ HV.GV (dieresis). Previous studies of the French of Windsor, Ontario (Kelly 2005, Poiré et al. 2007) have shown this to be a context quite favourable to glide formation (word-list, reading passage and directed conversation). We examine cross-word-boundary contexts more closely in our second study. If a form is attested in the most formal style then we can rightfully assume that it is being produced quite frequently in the everyday speech of that participant (Durand 2006). In the following section, we will discuss the PFC reading task.

5.1.2 The PFC Text

The text, intended to read without difficulty, is comparable to stories found in a regional newspaper article. According to the project directors, content is "axé sur la simplicité lexicale et grammaticale [...] on obtient alors un style soutenu certes, mais dans certains cas un style moins surveillé, voire naturel" (Durand et al. 2002: 42).

The text itself includes segmental material enough to confirm a phonemic inventory while also exploiting a range of phonological phenomena typical of Continental French; Schwa and liaison are systematically tested but other features are also included such as palatalisation, gliding and various types of assimilation (Durand 2006, Durand, Lyche and Laks 2002, 2009). A copy of the complete PFC text is found in the Appendix A.

In Table 11 below we list glide contexts in the PFC reading passage. Note that these include some lexicalised, all derived and cross-word-boundary contexts. Earlier work with reading data from the Windsor Ontario PFC survey (Kelly 2005) provide insights

that inform the present studies, in particular, regarding glide formation across word boundary.

Table 11: /HV+V/ and GV contexts in the PFC Text

pied	premier	valu à	officielles	soutien
quatrième	quatrième	habituels	vu à	désespoir
qui a	inquiet	qui ont	se vouer	Saint-Pierre
italiennes	entier	se multiplier	vérifier	Jonquière

Table 11 shows 16 tokens for word medial lexicalised and derived HVV contexts as well as four cross-word-boundary contexts: *qui a, valu à, qui ont, vu à*. Our findings show that these contexts show potential for glide formation or glide epenthesis. Based on these findings we include cross-word-boundary contexts in our primary study (spontaneous speech) discussed in chapter 6.

We turn now to the PFC protocol for conducting interviews with survey participants.

5.1.3 PFC guided discussion and informal conversation

Finally, the PFC protocol includes a guided interview (formal) between the primary investigator and the participant, and a free conversation (informal) between members of a group of participants. Ideally, the investigator will not be present during the informal conversations, but this can add the complication of having to reconstitute part of what is being said "[o]n travaille d'autant mieux sur des enregistrements qu'on a été présent et les deux enquêteurs se compléteront utilement pour les transcriptions et la phase d'analyse," (Durand, Laks & Lyche 2002: 10). The informal conversation exploits pre-existing relationships among participants in an effort to mitigate the observer's paradox (Labov 1972: 209). According to the project directors, the difference between the two conversations is largely the role of the investigator: "[...] dans le premier cas (discussion libre), il n'y a aucune asymétrie de rôle. L'enquêteur, s'il est présent, est un simple membre du groupe. Dans le deuxième cas (discussion guidée), c'est un enquêteur (Durand, Laks & Lyche 2002:10)".

The guided interview elicits a more formal register which may also present challenges for the investigator: "[l]'entretien guidée est un exercice périlleux car il doit à la fois laisser

le témoin s'exprimer mais aussi permettre de dresser un portrait sociolinguistique détaillé." (Durand, Laks & Lyche 2002: 42). During the interview, investigators seek to obtain essential information about the participant such as education, travel, work and languages spoken in order to complete the ***Fiche signalétique*** for each participant (cf. appendix C) while avoiding the perils of closed questions that elicit one-word responses. To this end, investigators are encouraged to invite participants to share personal stories; all the while guiding the discussion.

Our examination of glides in French includes a judgement task using PFC word-list data, the results of which inform our primary study of glide realisations in a spontaneous speech sample of PFC guided discussions. Our judgement task asks three linguistically informed native speakers of Hexagonal French to listen to audio recordings of words extracted from the PFC word-list recordings and then judge their realisations. From the results obtained, we identify tokens that are judged categorically to be realised with syneresis or dieresis. We examine the acoustic signal for all tokens resulting in categorical judgements with the goal of identifying those accoustic cues that correlate with realisation with syneresis or dieresis. The cues are then used to inform our own analysis of tokens in our second and primary study.

In the next section we discuss the methodology used in the judgment task.

5.2 Methodology

The judgement task is populated with audio stimuli from the PFC word-list. Our goal is to establish that visual acoustic cues may reliably indicate GV versus V(G)V sequences in a token to distinguish a realisation with syneresis versus one with dieresis. In order to achieve this goal we want to examine a diverse set of HVV data. To ensure diversity of the sample, our data is drawn from word-list recordings from eleven different PFC surveys of continental French including from Belgium: Tournai (bta), Liège (bla) and Gembloux (bga), from Switzerland: Nyon (sva). Hexagonal French includes surveys from *Occitan* regions: Douzens (11a), Toulouse (31a) and from *Île de France*: Paris (75c) and Brunoy (91a). Finally from the *Rhône Alpes*: Lyon (69a) and Grenoble (38a). To avoid

priming our judges for regional identifications in advance of hearing the stimulus to be judged, stimuli are isolated from the preceding numbers that accompany them in the reading task. To create our sample, we identified all lexicalised and derived glide contexts in the word-list and then reduced the number of words based on the high vocoid involved and the phonological environments in which they occur. The stimuli for our judgement task were spliced from the original PFC list recordings using the slicer.Praat script (Copyright (C) 2008 Pablo Arantes) written for the Praat acoustic analysis program (Boersma and Weenink 2013).

The goal of the Judgement Task is two-fold: first, to identify HVV tokens that obtain categorical results for syneresis (GV) and dieresis V(G)V (with or without epenthesis) from three judges who are native speakers of Hexagonal French. To these tokens we compare judgements provided by this author. Second, using only tokens judged by consensus from our panel, we examine visual acoustic cues observed on a spectrogram in PRAAT (amplitude, formant strength and intensity contour) to identify those cues on the acoustic signal most indicative of syneresis compared to dieresis. These cues are intended to inform our analysis of /HV(+)V/ contexts in our study of spontaneous speech.

From the PFC word-list (cf. Table 2 §5.1.1), we retain only a subset of words that offer lexicalised and derived contexts involving each of the three vocoids /IYU/. The words retained are shown below.

Table 12: PFC word-list, stimuli retained for judgement task

6.	(fou à) lier	51.	influence	72. (vous) prendriez
28.	scier	63.	miette	81. trouer
30.	mouette	79.	muette	

Given the breadth of our sample many audio samples were eliminated due to issues of acoustic quality (noise or distortion). Our final judgement task includes 782 tokens from 90 speakers, though the tokens are not balanced across speakers.

Modelled on the Multiple Forced Choice (MFC5) listening experiment provided in the Praat user's manual, our judgement task uses the stimulus-dependent run text with

randomization and replay button[14]. Our judges include three linguistically informed native Hexagonal French speakers and PFC collaborators who were solicited as consultants for their linguistic training, their familiarity with Praat and with SAMPA transcriptions. Familiarity with the Praat program and SAMPA transcription is essential. Each judge is provided Praat scripts (three different sessions) as well as instructions to launch the task script in the Praat program and save results upon completion of each session of the judgement task. In Praat, each judge runs the scripts provided (three in total) as time permits.

Once the judgement task is launched in Praat, further instructions appear on screen as shown in the figure below:

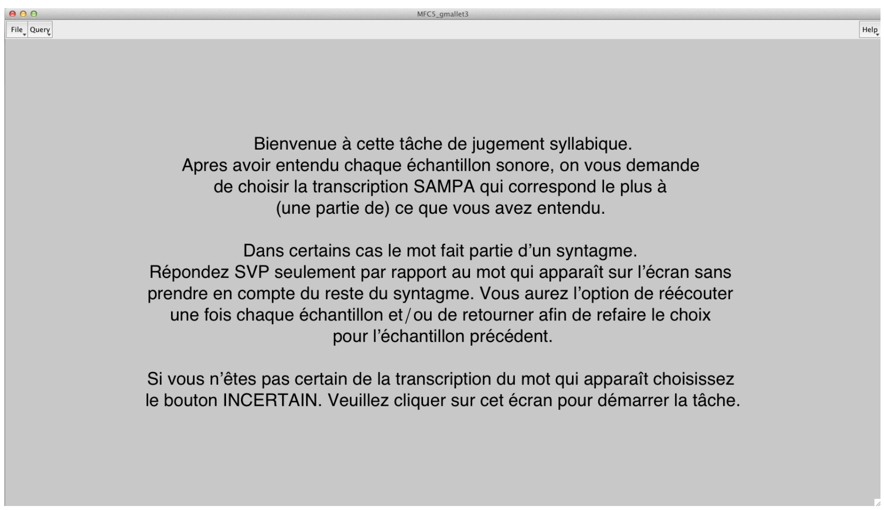


Figure 2: Judgement task, initial screen

As the task proceeds judges hear an audio file and they see on screen four options (screen 'buttons') with SAMPA transcriptions from which they choose the transcription that corresponds best to the stimulus just heard. When the judges click on the screen the option that, to their judgement, corresponds best to the audio stimulus, the result is

---

14 I would like to thank Paul Boersma for support provided while adapting the MFC5 model available in the Praat user's manual.

recorded and the task advances to the next word. A sample of the screen of options provided is shown below:

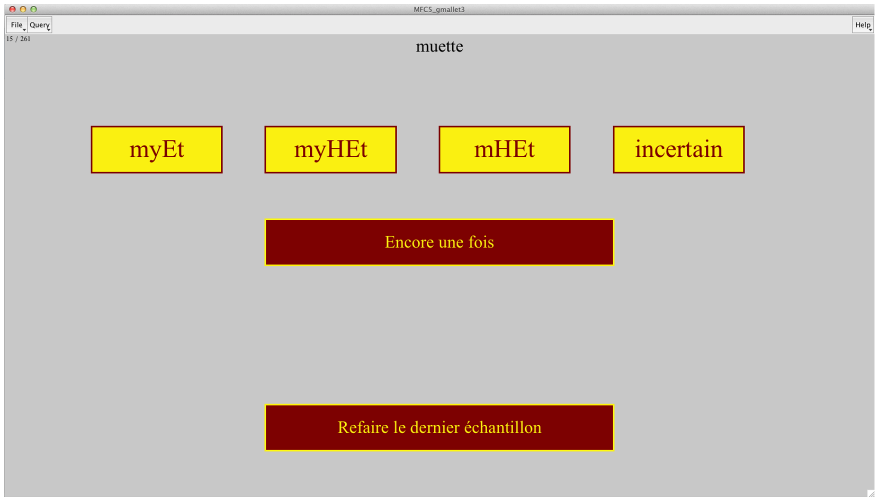


Figure 3: Judgement task, SAMPA transcription options for the stimulus *muette*

When shown the screen in Figure 3 above, judges hear the audio stimulus for *muette* (noted on the top of the screen) and are given four options to choose from. These options appear transcribed in boxes across the top of their screen (Figure 3), they are asked to choose the option that best fits the stimulus just heard. From left to right and provided in a SAMPA transcription, the options include realisation with: dieresis (hiatus), dieresis (epenthesis), syneresis, or uncertain. Stimuli may be replayed only once or judges can return to the previous stimulus. The judgement task includes 782 stimuli divided into three separate sessions of 260, 261 and 261 stimuli with breaks suggested after each set of 40 (see screen below).

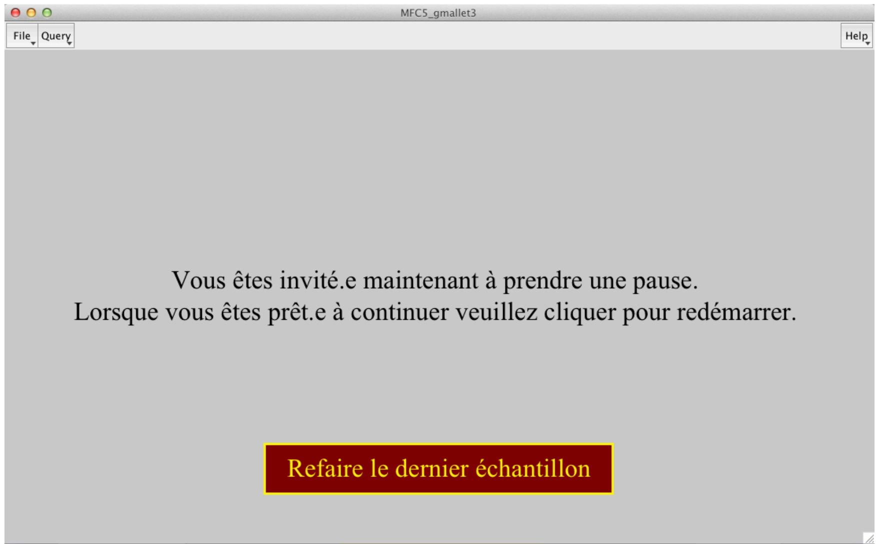


Figure 4: Judgement task, Invitation to take a break

Throughout the sessions, judges can see their progress by consulting the counter provided in the upper left portion of the screen.

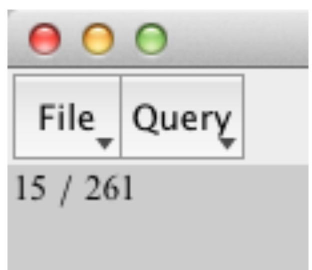


Figure 5: Judgement task, stimuli counter

Upon completion of a session, judges see their final instructions asking them to extract and save their results file.

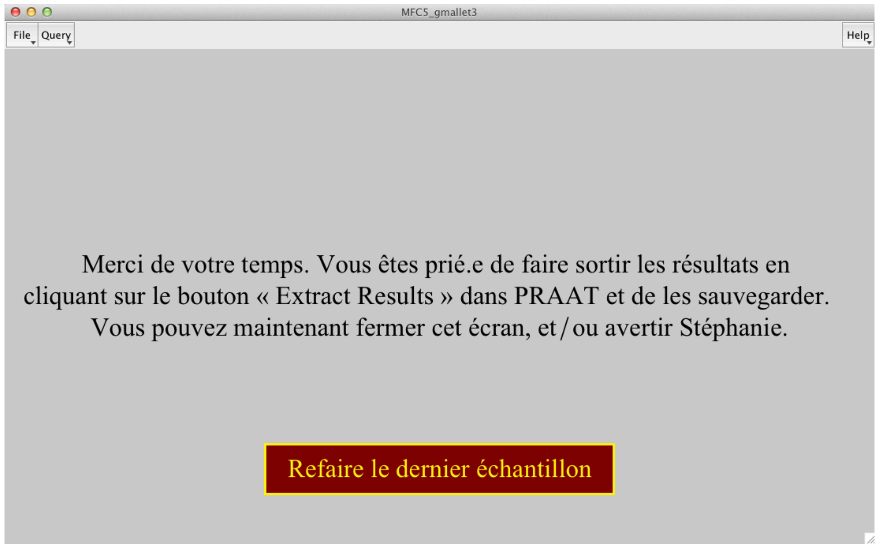


Figure 6: Judgement task, Final instructions to judges upon completion

Each script is named uniquely for each judge so that results can be stored and later indexed by judge.

In the following section, we present results for the judgement task. We begin with overall results and discuss the categorical results of the judgement task.

5.3 Judgement Task: Results

Each of our three judges evaluated identical sets of stimuli. Care was taken to assure an equal distribution amongst stimuli across eight words retained for the task. After removing stimuli for audibility issues a slight imbalance is observed. While measures of statistical significance might potentially offer some insights, we choose to report descriptive results in the following sections without such calculations, given the rather small Ns which are at times imbalanced across regions and speakers.

Table 13: Judgement task distribution of stimuli by word; 8 stimuli: N = number of iterations, % of total tokens calculated.

Stimuli	Judges 1-2-3		Total	
	N	%	N	%
6 (*fou à*) *lier*	104	13.30	312	13.30
28 *scier*	98	12.53	294	12.53
30 *mouette*	103	13.17	309	13.17
51 *influence*	103	13.17	309	13.17
63 *miette*	**67**	**8.57**	**201**	**8.57**
72 (*vous*) *prendriez*	104	13.30	312	13.30
79 muette	101	12.92	303	12.92
81 *trouer*	102	13.04	306	13.04
Total	**782**	**100**	**2346**	**100**

Table 13 shows distribution across eight stimuli for 782 tokens. Stimuli are represented in the task with relative parity; each represents approximately 13% of tokens, with the exception of 63 *miette* showing only 8.5%. The unequal number of instances for 63 *miette* is a product of token exclusion due to audible noise or distortion.

Given our goal to identify acoustic cues that distinguish realisations with syneresis from realisations with dieresis across three high vocoids, selection of stimuli is intended to represent, in as balanced a manner as possible, each high vocoid across lexicalised and derived glide contexts. Our primary study involves the examination of realisations of HVV contexts as they occur in spontaneous speech. Analysis of realisations of HVV contexts in spontaneous speech is based on acoustic cues identified here. In an effort to verify reliability of this author's judgements, we have also completed the judgement task. The tables that follow show findings for four judges (1, 2, 3 and K, this author) and for each of eight stimuli beginning with derived followed by lexicalised contexts.

Table 14 below shows task results by judge for each of the derived stimuli: *lier*, *scier*, *vous prendriez* /I/, *trouer* /U/ and *influence* /Y/. These contexts are not balanced across high vocoids because tokens involving /I/ dominate. The first two derived words show an /I+V/ context preceded by a simple onset (C); the last two words involve a preceding complex onset (OL) for the high vocoids /I/ and /U/. We will examine findings for lexicalised contexts in Table 14 below.

Table 14: Judgement task, results for four derived stimuli across three judges (1, 2, 3 and this author, K): N= number of tokens, percentages (%) calculated by response for each stimulus

Derived /I/	**1**		**2**		**3**		**Total**		**K**	
6 *(fou à) lier*	N	%	N	%	N	%	N	%	N	%
syneresis	**71**	**68.3**	**68**	**65.4**	**78**	**75.0**	**217**	**69.6**	**71**	**68.3**
dieresis	28	26.9	24	23.1	25	24.0	77	24.7	33	31.7
uncertain	5	4.8	12	11.5	1	1.0	18	5.8	--	--
Total	104	100	104	100	104	100	312	100	104	100
28 *scier*	N	%	N	%	N	%	N	%	N	%
syneresis	17	17.3	23	23.5	21	21.4	61	20.7	19	19.4
dieresis	**75**	**76.5**	**69**	**70.4**	**75**	**76.5**	**219**	**74.5**	**78**	**79.6**
uncertain	6	6.1	6	6.1	2	2.0	14	4.8	1	1.0
Total	98	100	98	100	98	100	294	100	98	100
72 *(vous) prendriez*	N	%	N	%	N	%	N	%	N	%
syneresis	34	32.7	--	--	24	23.1	58	18.6	3	2.9
dieresis	**56**	**53.8**	**101**	**97.1**	**79**	**76.0**	**236**	**75.6**	**98**	**94.2**
uncertain	14	13.5	3	2.9	1	1.0	18	5.8	3	2.9
Total	104	100	104	100	104	100	312	100	104	100
Derived /U/	**1**		**2**		**3**		**Total**		**K**	
81 *trouer*	N	%	N	%	N	%	N	%	N	%
syneresis	4	3.9	14	13.7	11	10.8	29	9.5	13	12.7
dieresis	**97**	**95.1**	**73**	**71.6**	**89**	**87.3**	**259**	**84.6**	**86**	**84.3**
uncertain	1	1.0	15	14.7	2	2.0	18	5.9	3	2.9
Total	102	100	102	100	102	100	306	100	102	100
Derived /Y/	**1**		**2**		**3**		**Total**		**K**	
51 *influence*	N	%	N	%	N	%	N	%	N	%
syneresis	1	1.0	38	36.9	25	24.3	64	20.7	9	8.7
dieresis	**102**	**99.0**	**61**	**59.2**	**75**	**72.8**	**238**	**77.0**	**88**	**85.4**
uncertain	-	--	4	3.9	3	2.9	7	2.3	6	5.8
Total	103	100	103	100	103	100	309	100	103	100

Results for derived stimuli show interesting consistencies across judges. We observe considerable consensus among judges for the same stimulus; at its lowest judges show agreement for the stimulus *(fou à)lier* at a rate of 69.6%, while at its highest judges agree in their judgements at a rate of 84.6%. We have bolded majority judgements for each stimulus and compare these findings to our own judgements. We find our own

judgements to be mostly in line with judgements observed among our Hexagonal French speaking judges. Dieresis is preferred for 72 (*vous*) *prendriez* (80.3%) and 81 *trouer* (84.6%); however, realisations for each stimulus across three judges show realisations are not as categorical as is usually described. *Trouer*, for example, shows findings for dieresis that range from 95.1% to 71.6% across all judges, while results for *vous prendriez* show greater variability ranging between 53.8% to 97.1%. Results for the stimulus *influence* show considerable inter-judge variability: dieresis is privileged to differing degrees (59.2%- 99.0%) across three judges. Important, judgements from this author (K) for each derived stimulus fall consistently within the range observed for three native speakers with the exception of *scier* where identification as dieresis averages 74.5% across three judges (70.4%-76.5%) while this author shows 79.6%.

In another matter, when we compare results for the stimulus *fou à lier* to the stimulus *scier*, the different findings for these two stimuli are telling. Here are two environmentally similar CHVV tokens where the high vocoid /I/ is preceded by a simple onset: fricative in one instance and liquid in the other; tokens for *lier* are judged to privilege syneresis while tokens for *scier* privilege dieresis. Because these are derived contexts involving the same high vocoid, we must ask if there is an environmental factor to be considered in the different realisations. An analysis of the possible conditioning effect of the preceding environments is merited but it is beyond the scope of this preliminary study. We discuss some effects of the preceding environment in our primary study (cf. section 6.3.3).

We turn now to results for lexicalised contexts among the word-list data. Table 15 below shows results for four lexicalised stimuli across a panel of three judges and this author (K).

Table 15: Judgement task, results for four lexicalised stimuli by judge (1, 2, 3 and this author, K): N= number of tokens, percentages (%) calculated by response for each stimulus

Lexicalised /I/	1		2		3		Total		K	
63 *miette*	N	%	N	%	N	%	N	%	N	%
syneresis	**61**	**91.0**	**61**	**91.0**	**63**	**94.0**	**185**	**92.0**	**48**	**71.6**
dieresis	6	9.0	3	4.5	4	6.0	13	6.5	19	28.4
uncertain	-	--	3	4.5	-	--	3	1.5	-	--
Total	67	100	67	100	67	100	201	100	67	100
Lexicalised /Y/	**1**		**2**		**3**		**Total**		**K**	
79 *muette*	N	%	N	%	N	%	N	%	N	%
syneresis	8	7.9	30	29.7	17	16.8	55	18.2	23	22.8
dieresis	**87**	**86.1**	**62**	**61.4**	**79**	**78.2**	**228**	**75.2**	**77**	**76.2**
uncertain	6	5.9	9	8.9	5	5.0	20	6.6	1	1.0
Total	101	100	101	100	101	100	303	100	101	100
Lexicalised /U/	**1**		**2**		**3**		**Total**		**K**	
30 *mouette*	N	%	N	%	N	%	N	%	N	%
syneresis	**63**	**61.2**	**77**	**74.8**	**72**	**69.9**	**212**	**68.6**	**72**	**69.9**
dieresis	38	36.9	23	22.3	30	29.1	91	29.4	31	30.1
uncertain	2	1.9	3	2.9	1	1.0	6	1.9	-	--
Total	103	100	103	100	103	100	309	100	103	100

According to these findings, all lexicalised stimuli show variability of realisation as judged by our panel and by this author (K). The stimulus *miette* shows the greatest consensus in results across our panel and among all other tokens examined in this judgement task; only three tokens elicit uncertainty and 13 tokens (6.5%) show dieresis with remaining tokens privileging syneresis at a rate of 92%. Results for this author also privilege syneresis; the rate (71.6%) is considerably lower than that of our panel. For all other stimuli (*muette,* and *mouette*), findings for this author (K) fall within the range observed across three judges. Results for the lexicalised minimal set *miette~muette~mouette* are intriguing. Not unexpectedly, *miette* and *mouette* show a preference for syneresis (86.9% and 68.9% respectively), however, *muette* is judged more often to be realised with dieresis (75.2%). We look at variability across high vocoids more closely in the next section when we examine findings for consensus across judges. According to these findings, the stimulus *muette* privileges realisation with dieresis.

Given the identical environment observed in each word of the set, we are left to conclude that the high vocoid itself is a conditioning factor in realisations observed. This observation can be seen to support Bullock's proposed markedness hierarchy: *M/j » *M/w » *M/ɥ, that stipulates /ɥ/ is least fit for margins (2002: 18). According to Bullock 2002, "the choice of making the labial velar glide more harmonic as a margin than the front rounded glide is weakly motivated by the slightly more liberal distribution [w] enjoys in French but, more important, it may simply reflect a more universal markedness of the front rounded glide: (10) *ɥ » *w » *j (cf. also section 4.8). We discuss aspects of cross-glide divergence further in our primary study findings (cf. section 6.7).

Across all lexicalised and derived data this author's judgements are largely in line with findings from our panel. For the stimuli *scier* and *miette*, we find divergence between this author's results and those of our panel which further supports the need of/search for acoustic cues that signal acoustic differences in realisations with syneresis versus those with dieresis to eliminate the effect of the individual judges on findings.

In the following section, we examine categorical results by stimulus.

5.4 Judgement Task: categorical responses

In this section, we examine categorical findings for each stimulus across our panel of three judges and compare these to results including this author's judgements to determine the degree of consensus across our panel and then to this author. Table 16 below shows the rate at which all three judges responded alike for each of eight stimuli thus a categorical finding.

Table 16: Judgment Task, categorical results across 3 judges (1, 2, 3), by stimulus, N= number of categorical results from a panel of three judges, % calculated for realisations with syneresis and realisations with dieresis.

Stimulus	Categorical syneresis		Categorical dieresis		Total Categorical		Total Tokens
	N	%	N	%	N	%	N
6 (*fou à*) *lier*	**57**	**54.8**	17	16.3	74	71.2	104
28 *scier*	16	16.3	**65**	**66.3**	81	82.7	98
30 *mouette*	**60**	**58.3**	19	18.4	79	76.7	103
51 *influence*	1	1.0	**51**	**49.5**	52	50.5	103
63 *miette*	**58**	**86.6**	1	1.5	59	88.1	**67**
72 (*vous*) *prendriez*	-	--	**53**	**60.0**	53	60.0	104
79 *muette*	6	5.9	**55**	**54.5**	61	60.4	101
81 *trouer*	1	1.0	**65**	**63.7**	66	64.7	102
Total	199	25.5	**326**	**41.7**	**525**	**67.1**	**782**

We note that categoricity is high for all eight stimuli; 525 (67.1%) of overall tokens are judged by consensus by our panel. At lowest, categorical responses show 49.5% tokens for the stimulus *influence*, privileging dieresis over syneresis 51:1. At highest, the stimulus *miette* shows 86.6% of tokens are judged categorically privileging syneresis over dieresis 58:1. Not unexpectedly, HVV contexts preceded by OL sequences show strong preference for dieresis, however, results show that categoricity at its highest shows a mere 63.7% (81 *trouer*). The stimuli *trouer*, and *influence* each show a single token with categorical syneresis while *vous prendriez* shows categoricity for dieresis only. The near identical derived contexts *lier* and *scier* show near equal categoricity but in opposite directions; *lier* privileges syneresis (54.8%) but *scier* privileges dieresis (66.3%). Finally, findings across the minimal set *miette-mouette-muette* provide a comparison of identical contexts for each high vocoid. The stimulus *muette* involving the front rounded high vocoid /Y/ shows a strong preference for realisation with dieresis (54.5%), while *mouette* (58.3%) and miette (86.6%) privilege syneresis.

Overall, stimuli involving /I/ show the greatest categoricity: *miette* 88.1% and *scier* 82.7%, however, their individual results are distinctly different. 58 of 59 tokens for *miette*

are judged with categorical syneresis while *scier* shows dieresis privileged occurring 65/81 times (66.3%) with the remaining 16 tokens (16.3%) judged with syneresis. Results for *lier* and *mouette* pattern similarly: syneresis is privileged for each (54.8% and 58.3%, respectively); as observed for *scier*, variability is much more prevalent: dieresis is observed in just under 20% of tokens. Across eight stimuli, tokens privilege realisation with dieresis (41.7%). Syneresis is observed in (25.5%) of tokens, which we attribute to contexts preceded by an OL sequence coupled with the (perhaps unexpected) findings for *scier* and *muette*. Where we observe consensus for syneresis, our findings show a clear majority amongst tokens for that stimulus: at highest *miette* shows 86.6%, at lowest *lier* shows 54.8%. Consensus for dieresis is lower (ranging between 41.7% and 63.7%). Consensus among judges is lowest for the stimulus *influence* at 50.5% of tokens judged by consensus realised with dieresis. This finding is followed closely by the stimulus *(vous) prendriez* by consensus for dieresis at 51%.

As mentioned earlier, rates of consensus across our panel of judges for these stimuli also reflect a distributional hierarchy; the stimulus *miette* contains /I/ and shows the highest rate of consensus for syneresis (88.1%) followed by *mouette* containing /U/ (76.7%) for syneresis whereas *muette* shows consensus for dieresis (54.5%). Finally, a most surprising outcome for this task; we note consensus amongst our panel of judges for realisation with syneresis for a single token of the stimuli 81 *trouer* and 51 *influence* both of which are preceded by an OL sequence. A single token with syneresis is not by itself noteworthy however consensus across a panel of three judges constitutes genuine outliers.

Let us turn now to categorical findings across four judges (our panel of judges and this author). In Table 17 below we show categorical finding for all eight stimuli.

Table 17: Judgement task, categorical results across 4 judges (1, 2, 3, K) by stimulus, N= number of categorical results from a panel of three judges, % calculated for realisations with syneresis and realisations with dieresis.

Stimulus	Categorical syneresis X 4		Categorical dieresis X 4		Total categorical X 4		Total Categorical X 3		Total tokens
	N	%	N	N	N	%	N	%	N
6 (*fou à*) *lier*	**49**	**47.1**	14	13.5	63	(60.6)	74	71.2	104 (10.6)
28 *scier*	12	12.2	**64**	**65.3**	76	(77.6)	81	82.7	98 (5.1)
30 *mouette*	**58**	**56.3**	18	17.5	76	(73.8)	79	76.7	103 (2.9)
51 *influence*	1	1.0	**51**	**49.5**	52	50.5	52	50.5	103
63 *miette*	**48**	**71.6**	1	1.5	49	(73.1)	59	88.1	**67 (14)**
72 (*vous*) *prendriez*	-	--	**53**	**51.0**	53	(51.0)	53	60.0	104 (9.0)
79 *muette*	4	4.0	**50**	**49.5**	54	(53.5)	61	60.4	101 (6.9)
81 *trouer*	1	1.0	**61**	**59.8**	62	(60.8)	66	64.7	102 (3.9)
Total	173	22.1	**312**	**39.9**	**485**	(62.0)	**525**	**67.1**	**782 (5.1)**

First, we point out that the outliers for syneresis following an OL sequence from our previous table remain amongst categorical findings here; one token each for the stimuli 81 *trouer* and 51 *influence* are judged categorically to be realised with syneresis. We will discuss these tokens further in the following acoustic analysis. Across four judges we find less consensus with overall rates dropping (5.1%). The drop in consensus (± 5%) is consistent for a majority of stimuli (lower consensus marked by parentheses). The stimulus 51 *influence* is judged at parity while others: 6 *fou à lier* (10.6%) and 63 *miette* (14%), show more dramatic decrease in consensus. Stimuli such as 30 *mouette* (2.9%) and 81 *trouer* (3.9%) show similar results compared to findings for our panel of three judges while 28 *scier* (5.1%) and 79 *muette* (6.9%) are consistent with the overall average. Change in categorical findings for tokens realised with syneresis range from no change for 51 *influence*, 72 (*vous*) *prendriez*, 81 *trouer* to a difference of 15% for the stimulus 63 *miette* and categorical findings with dieresis range from no change for 51 *influence*, 63 *miette*, 72 *(vous) prendriez*) to a high of 5% (79 *mouette*). Overall, the greatest difference is observed for tokens judged realised with syneresis: 25.3% is reduced to 22.1% however the loss in consensus is less dramatic for tokens judged to be

realised with dieresis (41.7% to 39.9%). This author's judgements vary from our panel yet overall tendencies in judgements (syneresis versus dieresis) by stimuli are in evidence nonetheless despite being a native speaker of English. Some variability we attribute to the effect of surrounding environments particularly HVV contexts preceded by [m] present difficulty (for this author, also mentioned by some judges).

This change in consensus observed across judges speaks to the reliability of this author's judgements. Our own judgements pattern in the direction of the panel on the whole. However, we examine closely visual cues to identify syllabicity in realisations of HVV contexts. We examine findings from our judgement task for tokens showing consensus across three Hexagonal French speaking judges for realisations with syneresis and dieresis. In the following sections, we will compare tokens for each stimulus that obtains consensus for syneresis and dieresis from our panel of judges with a goal to identify visual acoustic cues that distinguish syneresis from dieresis. These cues will serve as the basis for our analysis of HVV contexts in our primary spontaneous speech study.

5.5 Acoustic cues: categorical syneresis and dieresis

In this section, we examine tokens judged categorically by our panel of three judges for realisation with syneresis and with dieresis to identify the most prominent visual acoustic cues for each on the PRAAT spectrogram. Specifically, we examine peaks of intensity on the intensity contour as an indication of syllable count: one syllable = one peak = syneresis, two syllables = two peaks = dieresis. In an effort to support observations from the intensity contour, we examine also wave amplitudes and formant energy for the same period of vocalic production. Our goal here is to evaluate the reliability of intensity supported by formant energy and amplitude as a cue to the identification of syneresis versus dieresis in HVV sequences.

Before turning to our analysis of visual acoustic cues, we discuss briefly those phonetic qualities most commonly attributed to glide production and their role in our analysis.

At the sub-phonemic level, glides are taken to be identical to their corresponding high vowel. According to Hume (1995), for example, "[glides] are identical in terms of feature

composition differing only in syllable position; a vowel is a syllable nucleus whereas a glide never is (Sievers 1881, Pike 1943, Clements and Keyser 1983, Kaye and Lowenstamn 1984, Levin 1985)." Hume (1995:1).

As discussed in detail in Chapter 1, phonetic descriptions of glides generally centre on their distribution within the word as well as their consonant-like qualities. In the next sections we discuss duration and constriction.

5.5.1 Duration

Analysis of the realisation of /HV+V/ sequences may rely on duration as a factor in distinguishing syneresis from dieresis. Duration should be a reliable cue since a V(G)V sequence is inherently longer than a GV sequence; however, we find that inherent difficulties related to variable speech rate in spontaneous speech data make duration an unreliable measure to suit our goals. In an earlier study of HVV sequences in the Canadian French of Windsor Ontario, Poiré et al. (2007) find that duration of HVV sequences varies across speech styles (PFC reading passage versus the semi-directed conversation), and based on the position in the accentual phrase. The study is based on the model developed by Cedergren and Perreault (1995, 1994):

> Ce modèle probabiliste [...] met en lumière trois principaux facteurs conditionnant la variation de la durée segmentale: i. la complexité de la syllabe, ii. la complexité de la rime et iii. la position relative de la syllabe dans le domaine de réalisation du contour intonatif.
>
> (Poiré et al. 2007)

Given the number of factors that may condition segmental duration, we have chosen not to use measurements of duration as a cue to distinguishing syneresis from dieresis. A laboratory-controlled study, wherein cross speaker normalisation allows for more reliable comparison of duration, is not ruled out for future studies.

5.5.2 Constriction: consonant-like quality of glides

> A consonant is the result of audible friction, squeezing, or stopping of the breath in some part of the mouth (or occasionally of the throat). The main distinction between vowels and consonants is that while in the former the mouth configuration merely modifies the vocalized breath, which is therefore an essential element of the vowels, in consonants the narrowing or stopping of the oral passage is the foundation of the sound and the state of the glottis is something secondary. (Sweet 1877: 31)

As discussed in Chapter 2, Passy (1913) describes high vowels as the least sonorous and therefore the more consonant-like of vowels. Glides, he says, are fricatives distinguished from vowels by their friction (*frottements*). According to Grammont "Les semi-voyelles sont encore éminemment des spirantes et aussi bien des fricatives et des constrictives" 1963:77). According to Jones (1932), glides are inherently short combined with absence of stable state renders them least sonorous among vowels therefore more consonant-like: a frictionless continuant. Saussure observes that the closed property of high vowels, common also to consonants, justifies the term semi-vowel:

> Par rapport aux autres voyelles, ces sons [i, y u] supposent une fermeture encore considérable, assez voisine de celle des consonnes. Il en résulte certaines conséquences [...] qui justifient le nom de *semi-voyelles* donné généralement à ces phonèmes. (1982:75 original emphasis)

These characteristics provide a composite image of consonant-like qualities; glides are closed, less sonorous, and constrictive. These are features intended to distinguish glide production from that of a corresponding high vowel. Though commonly used to describe the articulatory quality of glides (Sweet 1907, Straka 1964, Maddieson and Emory 1985), constriction is not always a reliable measure of glide production (Straka 1964). Pike states "[t]here seems to be no articulatory measuring rod for degree of constriction or obstruction which marks the consonant-vowel border", (1943:67, cf. also Chapter 1):

Our study of the acoustic signal in Praat aims to identify visually those acoustic cues that set apart high vowel production from glide production but more precisely to contribute to distinguishing syneresis from dieresis. We examine in particular the intensity contour of tokens that are judged categorically by our panel as realised with syneresis versus with dieresis.

We turn now to an analysis of tokens judged with categoricity for realisations with syneresis and with dieresis.

5.6 Analysis of tokens: categorical judgements for syneresis and dieresis

For this analysis of tokens judged by our panel, we identify and segment HVV contexts following methodology set out in Poiré et al. (2007): "[n]ous cherchons donc à savoir comment les différentes sous-parties de la séquence (stabilité formantique initiale, transition, stabilité formantique finale) réagissent aux variations de durée causées par la position dans le syntagme intonatif", (2007: 344). In this earlier study, HVV tokens are divided into three sub parts: periods of stable formant activity followed by transitional formant activity and potentially another period of stable formant activity. The goal of our earlier study (Poiré et al. 2007), is to examine the relationship between duration (of these subparts) and position in the accentual phrase. In the present study, the same subparts are identified by formant behaviour (energy, stability and transition) in Praat with the goal of identifying syneresis versus dieresis. Subparts are labelled for the high vowel or corresponding glide (tokens for categorical syneresis, for example). Tokens judged for dieresis are marked for the high vowel, the transition (TR), and for the following non-high vowel.

In the figures to follow, segmentation and transcriptions will reflect this tri-partite system such that tokens judged to be realised with syneresis (GV) are transcribed with the glide involved since the presence of glide is unambiguous. Tokens for dieresis are transcribed with a glide when judgements show consensus for epenthesis. If for a given token, our panel's findings show discrepancy (categorical dieresis but findings for both hiatus and epenthesis, for example), we use TR in our transcription (HV-TR-V). In all cases, periods of transition are marked on the spectrogram with a double-headed arrow to indicate approximately where formant transitions occur.

We note also some co-articulation effects in the production of HVV sequences. These relate especially to OL cluster simplification: liquid deletion, liquid assimilation to voice

of a preceding voiceless obstruent, and schwa epenthesis between O and L. Other phonetic details such as distinction of variants for /ʀ/ or variation in vowel qualities are not included here. Parentheses ( ) are used in our transcriptions to indicate the absence (deletion) of an expected segment (syncope of the liquid in an OL cluster, for example, is transcribed as [t(ʀ)]). Where /ʀ/ is present but voiceless it is transcribed as such [ʀ̥].

In many cases, we will examine tokens that obtain categorical results from our panel of judges (all three judges find syneresis (GV) or dieresis (HV(G)V); in some cases, however, we use tokens for which all judges agree that a realisation is made with dieresis but discrepancies occur across judges as to the realisation of epenthesis (HVGV versus hiatus HVV). Since our goal here is to assess the effectiveness of the intensity contour (corroborated by the spectrogram and oscillogram) as a visual cue to indicate number of syllables (syneresis versus dieresis) with which a token is produced, we are satisfied that these judgements are comparable. Some discrepancies between judges' findings for hiatus versus epenthesis, in these situations the judges findings are noted in footnotes; the presence or absence of an epenthetic glide is beyond the scope of this study. We have followed the methodologie of Poiré et al. 2007 in segmenting the GV or VGV sequences:

> Si la segmentation d'un énoncé oblige à fixer une frontière segmentale entre cette transition et une période de stabilité formantique, le choix du moment de ce passage d'un segment à un autre n'est jamais totalement satisfaisant du point de vue de l'analyse acoustique.
>
> (Poiré et al. 2007: 341)

In each token presented here, we show selection lines that approximate vocalic boundaries mark all sub parts, once identified.

We begin our analysis with tokens for the same stimulus that exhibit syneresis and compare these to tokens with dieresis starting with a stimulus having a simple onset preceding (28 *scier* /I/) then we turn to a stimulus with a complex onset (81 *trouer* /U/). These tokens provide a sample of two of three high vocoids preceded by a simple and a complex onset so that we can examine the role of the various visible acoustic cues in differentiating tokens with dieresis from those with syneresis. After establishing the cues we will focus on in this analysis, we will examine tokens for the remaining task stimuli:

three tokens for complex onsets involving each of three high vocoids, two stimuli exhibiting simple onsets for the high vocoid /I/ and then our minimal set with three stimuli with identical environments for each of the three high vocoids.

Figure 7 below shows the oscillogram and spectrogram for a token of the stimulus 28 *scier* realised with dieresis (epenthesis) judged by consensus by our panel. The images used are taken from Praat spectrograms; each token is adjusted for speakers' sex: male speakers measure on a scale of 0-5000 Hz and female on a scale of 0-5500 Hz, while intensity contour is overlaid in Praat on a scale from 50-100 dB.

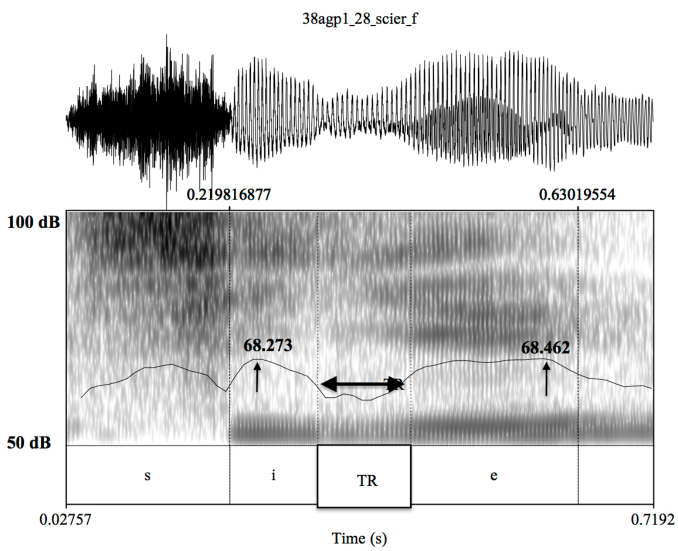


Figure 7: Praat spectrogram with intensity contour and oscillogram for the token 38agp1 28 *scier* (dieresis with epenthesis)

This token, judged categorically for dieresis, is realised with epenthesis. Vertical lines mark the approximate boundaries for vowel segments (time in seconds for the vocalic period noted above). If constriction is indeed characteristic of glide production, we take it to be a matter of perception since signs of constriction are not identified on the spectrogram here; rather we observe energy loss (shown in [TR] on the spectrogram). We note that transition from the preceding fricative [s] exhibits a considerable dip before climbing dramatically in correspondence to production of the high vowel [i]. This token, judged to be realised with dieresis with epenthesis is marked by weakened formant

energy on the spectrogram which corresponds with a dip (to 59.50 dB) in intensity observed on the intensity contour. Further study of the perception of the presence of a glide, while warranted, is beyond the scope of the present study; we do note however, the potential link between energy loss and the perception of constriction (cf. Howe 2014 on intensity and lenition or assibilation in Spanish stops).

The intensity contour shows that each vowel corresponds to a peak in intensity (shown with vertical arrows): [i] (68.36 dB) and [e] (68.43 dB); vowels are clearly separated by a dip in intensity that measures (8.86 dB) between peaks. Characteristic signs of formant transition are not obvious, however, peaks of intensity, formant energy patterns on the spectrogram, corroborated by wave amplitude on the oscillogram are unmistakeable. For comparison, we turn now to a token for the same stimulus 28 *scier* judged by consensus to be realised with syneresis (Figure 8 below).

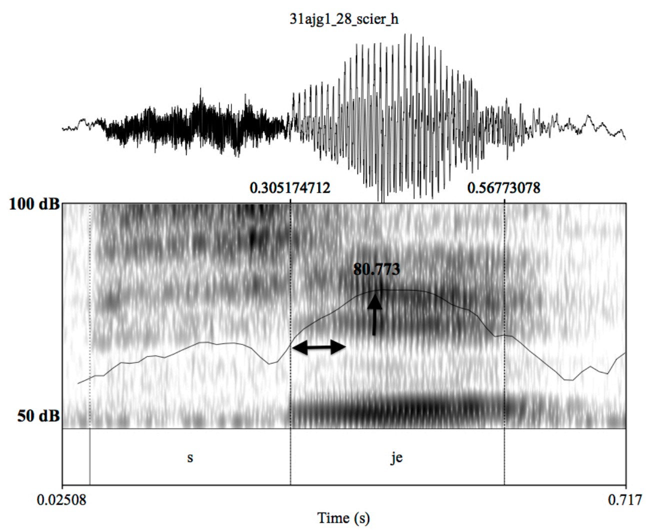


Figure 8: Praat spectrogram with intensity contour and oscillogram for the token 31ajg1 28 *scier* (syneresis)

As in our preceding token, we observe a transition from the preceding [s] with a considerable dip in intensity from 69.11 dB to 64.19 dB (a difference of 4.92 dB) during [s] production, before ascending to a single peak at 80.77 dB during the vocalic production that follows; vertical lines mark the (approximate) boundaries for vocalic production (time in seconds for the vocalic period noted above). We note that the peak

intensity is achieved late in the vocalic period compared to what was observed for the token with epenthesis above; we attribute this difference to high vowel production in the preceding token (cf. Fig. 7) versus glide production here. As was observed in the preceding token, we note the absence of formant transitions which we attribute to the preceding fricative [s] since we can observe formants during the production of [s] which suggests that there is some assimilation or co-articulation present. In contrast to the preceding token, we observe signs of energy building in formants at voicing onset (double-headed arrow). Although the characteristic signs of formant transition are absent, the shape of the intensity contour is clear. For this vocalic period, intensity ranges from 70.20 dB to 80.77 dB (vertical arrow). Differences in peak intensity between tokens can be attributed to individual variations in speech. The single peak of intensity is consistent with our judges' finding for realisation with syneresis.

Our analysis of the two tokens for 28 *scier* illustrates that peaks of intensity correspond to vowel production and as such we can expect that our judges' findings for syneresis (one syllable) or dieresis (two syllables) are reflected in the shape of the intensity contour: one syllable = one peak.

We turn now to an examination of variable realisations for the stimulus 81 *trouer.* Figure 9 below shows a token for the stimulus 81 *trouer* realised with dieresis (hiatus). In contrast to the preceding tokens for the stimulus 28 *scier*, this stimulus presents the high vocoid /U/ preceded by a complex OL onset.

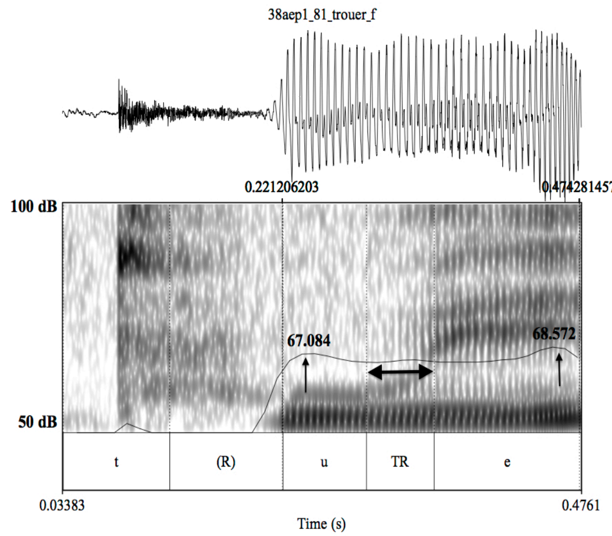


Figure 9: Praat spectrogram with intensity contour and oscillogram for 38aep1 81 *trouer* (dieresis with hiatus)

By the intensity contour in the OL onset, we observe signs of simplification through liquid deletion; while we should expect intensity to drop to zero for the production of [t], we should also expect intensity to increase during the production of [R] then transition to high vowel production. In this token, however, intensity begins from zero immediately preceding vocalic production which implies absence of [R] that can be attributed to OL simplification through liquid deletion (cf. Ladefoged 2005, Chow and Poiré 2007ab.). The transition from zero to the first peak corresponding to [u] is immediate in contrast to the delayed peak in intensity we observed for a glide plus vowel sequence or syneresis (cf Fig. 8 above and Fig. 10 below). This token provides a clear contrast to the preceding tokens which involved the high vocoid /I/; the first two formants are much closer for /U/ here than is observed for /I/ such that formant transition is much more evident on the spectrogram above than is observed in preceding tokens. Vertical lines mark the approximate boundaries for vowel production and the transition that separates the two vowels produced in *trouer*. Time in seconds of each vocalic period is marked above the spectrogram. Intensity peaks are less pronounced, which we attribute to the high back vocoid /U/. We identify, nonetheless, two moderate peaks beginning and ending the intensity contour. Two peaks are separated by a transition clearly marked by characteristic signs of formant transition as well as signs of energy loss. Peaks of

intensity (marked by vertical arrows) correspond to vowel production: [u] (67.08 dB) and [e] (68.57 dB), while the transition (TR) between vowels shows only a moderate dip (to 65.27 dB). This is a difference of only 1.28 dB as compared to the more considerable dip of 8.641 dB observed for 28 *scier* realised with epenthesis [ije] (Fig. 7 above). Further study would be required to determine any relationship to presence or absence of glide. In the following example we examine the same stimulus realised with dieresis.

Judges' findings are infrequently categorical for epenthesis versus hiatus which attests to the difficulty in judging for the presence of glide versus absence of glide. In Figure 10 below, we examine another token for the stimulus 81 *trouer* which, based on our judges' findings, is realised categorically with dieresis (one judgement for dieresis with epenthesis and two for dieresis with hiatus).

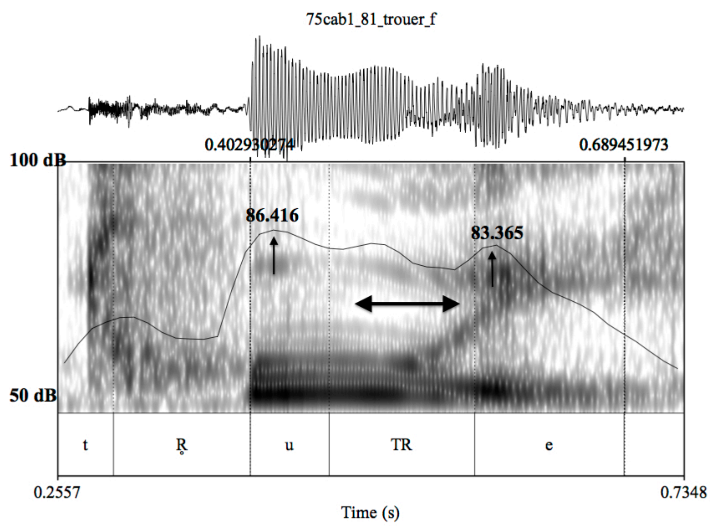


Figure 10: Acoustic cues, Praat spectrogram with intensity contour and oscillogram, for the token 75cab1 81 *trouer* (dieresis)

Figure 10 shows *trouer* realised with dieresis corroborated by two sharp peaks, marked by arrows with measurements in dB, on the intensity contour at the beginning and end of vocalic production. The intensity contour immediately preceding the vocalic production is more dynamic than is observed in the preceding token; we take the peak observed in the onset before transition to vocalic production to indicate the presence of a voiceless liquid [ʁ̥]. In contrast to the preceding token, this OL onset does not undergo

simplification and intensity peaks are much more prevalent. Intensity transition from the preceding liquid begins at 64.44 dB rather than zero and attains a first peak of 86.42 dB during the high vowel [u]. From that first peak, intensity is in a steady fall and continues to fall to its most pronounced dip during vocalic production before resuming its climb to the peak for [e] production. As noted, formant positions reflect /U/ compared to /I/. The tell-tale signs of transition are easily identified by the sharp rise in F2 which only stabilises for production of the front mid vowel [e] marked also by the notable increase in energy on the spectrogram. The spectrogram shows a considerable loss of energy in F2 and F3 after high vowel production and before the mid front vowel which, in turn, corresponds to the most pronounced dip of intensity and a section of low amplitude on the oscillogram preceding [e] production.

Note that it is difficult to ascertain the exact point at which a glide ends and production of the vowel [e] that follows begins. Ladefoged discussing the difficulty related to the segmentation of the sequences involving the bilabial glide, says: "[t]here is no separate w." (2003:140). This token is judged with some discrepancy to be realised with dieresis which is corroborated by the visible acoustic cues.

We turn now to the examination of a token for the same stimulus 81 *trouer* (Figure 11 below) judged categorically by our panel for realisation with syneresis. We note that this finding is unexpected for a HVV context preceded by an OL sequence.

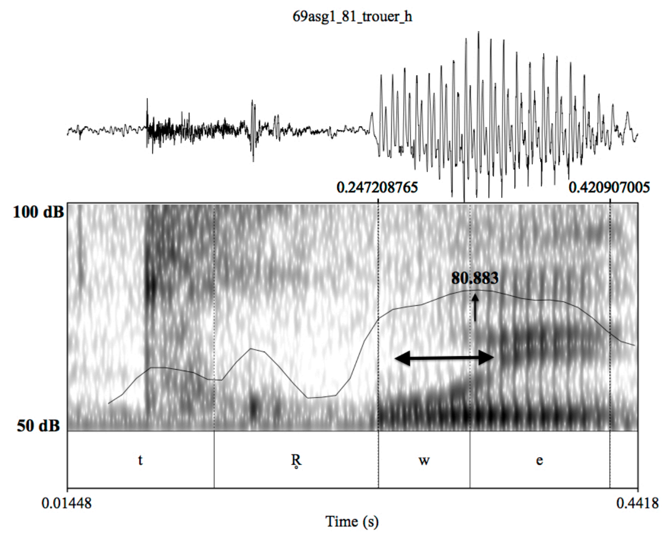


Figure 11: Praat spectrogram with intensity contour and oscillogram for the token 69asg1 81 *trouer* exhibiting syneresis [we].

Given the unexpected finding for this token, we might expect a realisation with syneresis after OL only after simplification of the preceding OL sequence (liquid deletion, for example); however, we observe on the spectrogram an indication of the presence of a liquid (intensity peak in onset). The shape of the intensity contour observed here resembles that which we observed on the previous token wherein the liquid, though present, is voiceless. Intensity peaks only once during the vocalic period of this token corroborated by the immediate transition of formants and the single peak in wave amplitude on the oscillogram. These indications are consistent with our judges' consensus in finding for a realisation with syneresis [we]. Syneresis is further indicated by the formant transition observed in F2 at the onset of vocalic production (double-headed arrow) before attaining a sustained period of stability at the production of [e]; note that F2 and F3 gain strength and stability to the right of the double-headed arrow. The intensity contour for the vocalic period begins relatively high 76.56 dB despite a considerable dip through [R] then peaks at 80.88 dB at the onset of [e] production.

Evidence from the preceding tokens illustrates how we are able to reliably distinguish tokens for syneresis from those realised with dieresis using peaks of intensity on the

intensity contour with corroborating evidence from other acoustic cues visible on the spectrogram and oscillogram.

In the case of syneresis, we observe a single peak of intensity with intensity loss associated with transition in evidence only at the onset of vocalic production. The single peak in intensity occurs later in the vocalic period coinciding with the production of the following non-high vowel. Dieresis, on the other hand, is distinguished by two peaks in intensity separated by a period of transition (with or without glide production). Peaks of intensity may appear less evident depending on the high vowel involved and the preceding environment; transition from the onset to the following vocalic period is directly influenced by the preceding consonant. We will discuss this further in the token analysis to follow. Transition between intensity peaks is marked by a dip in intensity on the intensity contour, formant transition often accompanied by evidence of energy loss on the formant contours, and diminished wave amplitude on the oscillogram.

We examine next variable tokens with dieresis versus syneresis for each of the remaining stimuli. For comparison purposes, we begin with tokens preceded by an OL sequence (51 *influence* /Y/, 72 *(vous) prendriez* /I/). We will examine the effectiveness of the intensity contour as a diagnostic tool for distinguishing dieresis from syneresis while also noting differences in the acoustic signal for each high vocoid and differences that result with changes in the surrounding environments. Figure 12 below shows 51 *influence* judged unanimously to be realised with syneresis.

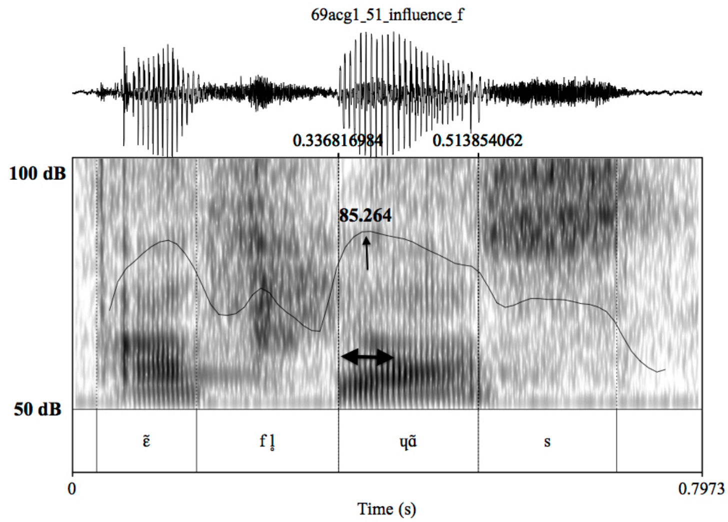


Figure 12: Acoustic cues, Praat spectrogram and intensity contour with oscillogram for the token 69acg1 51 *influence* (syneresis)

Vocalic production following /fl/ shows only a single peak of intensity corresponding a single period of peak wave amplitude on the oscillogram (comparable to that of [ɛ̃] preceding). The vocalic period of the final syllable begins by a period of formant transition and energy loss corresponding to glide production; /Yɑ̃/ is realised as a GV sequence [ɥɑ̃]. The spectrogram shows no sustained period which might indicate production of the high vowel [y], rather the vocalic production begins with low energy in F2 followed immediately by formant transition towards the low back vowel [ɑ̃]. The single peak of intensity measuring 85.264 dB corroborates the judges' findings for syneresis. The transition of front rounded high vocoid /Y/ shows the descent of F2 on the formant contour and the early rise to a peak of the intensity contour which we attribute to the nasal vowel. The preceding onset /fl/, shows that intensity is greatly diminished during the production of the voiceless fricative [f] then climbs to a moderate peak during production of the voiceless liquid [l̥].

Figure 13 below shows a token for the same stimulus 51 *influence* judged by consensus from our panel of judges to be realised with dieresis (hiatus).

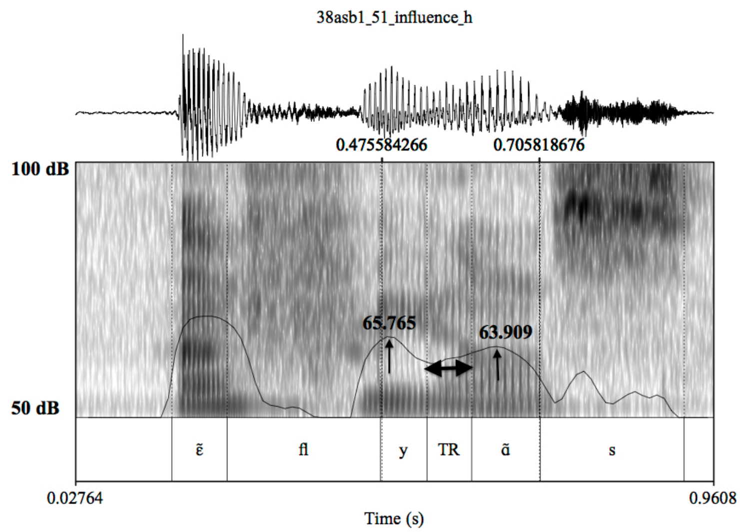


Figure 13: Acoustic cues, Praat spectrogram with intensity contour and oscillogram (above) for the token 38asb1 51 *influence* (hiatus)

Figure 13 shows the final period of vocalic production exhibits two peaks separated by a marked dip on the intensity contour during the HVV production. As compared with a single peak and early transition observed in our previous token for syneresis, the distinction between syneresis and dieresis (here) is clear. The second period of vocalic production bounded by vertical lines corresponds to the final two syllables of this token (judged by our panel to exhibit hiatus). During this period of vocalic production the intensity contour shows two distinct peaks separated by a distinct yet smooth transition. The first sustained peak of intensity for [y] is followed immediately by transition of F2 (double-headed arrow) before the production of the final nasal vowel [ɑ̃] (second intensity peak).

In Figure 14, we examine a token for the stimulus 72 *(vous) prendriez*. This stimulus obtains categorical results for dieresis from our panel with discrepancy between judgements for hiatus versus epenthesis. Two judges find for epenthesis and one for hiatus.

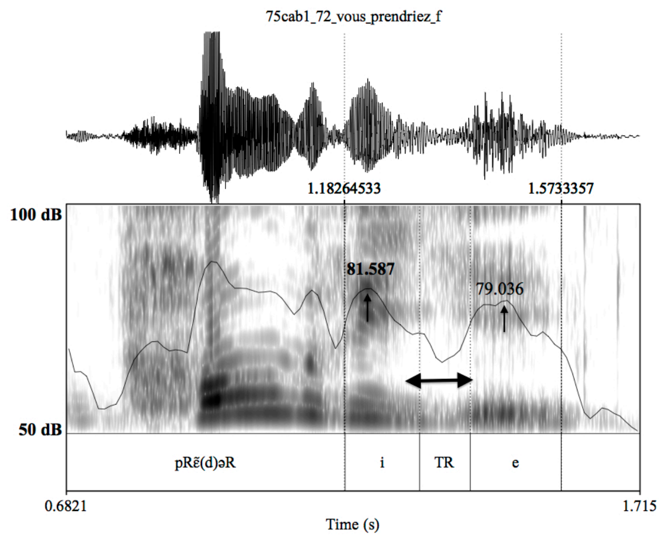


Figure 14: Acoustic cues, Praat spectrogram with intensity contour and oscillogram (above) for the token 75cab1 72 *(vous) prendriez* (dieresis)

Here we observe evidence of OL simplification through schwa insertion in the onset evident in the second peak on the intensity contour, which we take to be [ə] (cf. Poiré and Chow 2007, Colantoni and Steele 2007). We observe also that the shape of the intensity contour, its dramatic peaks and unmistakable dip (65.42 dB) during the final period of vocalic production (marked by vertical lines), shows clear evidence for dieresis. During the same period, we observe corresponding peaked waveforms on the oscillogram. Close examination of the spectrogram for this period shows sustained energy in the first three formants that corresponds to production of the vowel [i] followed by a sustained period of energy and intensity loss (TR) before both the intensity and the energy climb again to a second peak corresponding to the production of [e].

Our analysis, thus far, examines variable realisations of the same stimulus (tokens judged by our panel to be realised with syneresis or with dieresis) involving three high vocoids preceded by a simple onset (in 28 *scier*) as well as complex onsets for 51 *influence*, 72 *(vous) prendriez*, and 81 *trouer*. We have examined HVV realisation in both lexicalised and derived contexts and for all three high vocoids. We have shown that an examination of peaks on the intensity contour, corroborated by evidence from formant contour on the

spectrogram and wave amplitude on the oscillogram, proves to be a strong diagnostic tool in visually distinguishing syneresis from dieresis in an instrumental analysis of the signal.

Each of the following stimuli is preceded by a simple onset: containing consonant phonemes [l], [s], and the nasal [m]. In Figure 15 we examine a token for the stimulus 6 *(fou à) lier* judged categorically to be realised with syneresis.

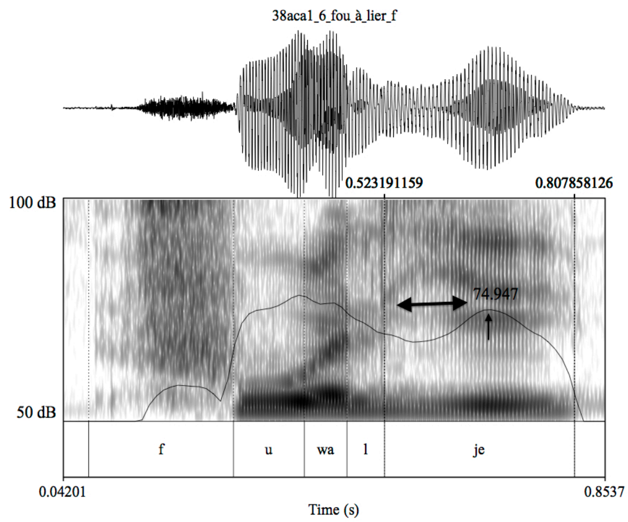


Figure 15: Acoustic cues, Praat spectrogram with intensity contour and oscillogram (above) for the token 38aca1 *6 (fou à) lier* (syneresis)

The token above shows the stimulus 6 *fou à lier* wherein the word *lier* is judged categorically by our panel to be realised with syneresis. Intensity shows a steady fall from the preceding syllable and continues to fall through the preceding [l] then attains its lowest measure before increasing to a single peak indicating syneresis. We observe early transitions at the onset of vocalic production (double headed-arrow) with simultaneous loss of energy and intensity on the spectrogram. Rising intensity is simultaneous with formant transitions and peaks with formant stabilisation and increased energy for the following mid vowel [e]; a single peak in wave amplitude on the oscillogram corroborates the cues.

Figure 16 below shows the token 38acm1 6 *(fou à) lier* judged to be realised with dieresis.

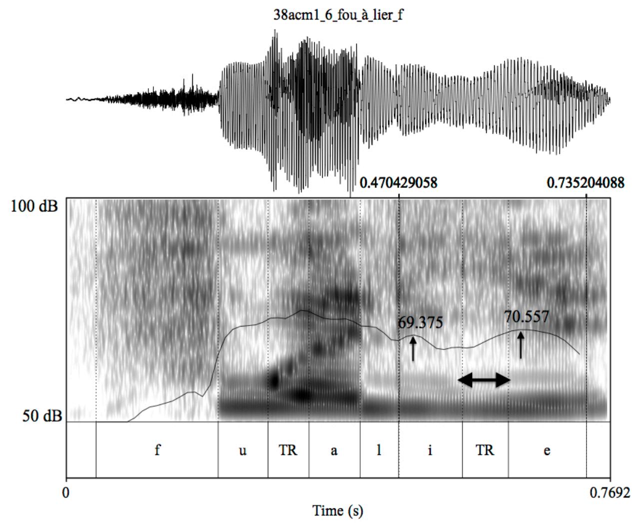


Figure 16: Acoustic cues, Praat spectrogram with intensity contour and oscillogram for the token 38acm1 6 (fou à) lier (dieresis)[15]

On the intensity contour here, we observe transition from [l] to [i] that is less evident in the preceding token. We take this as an indication of presence of the high vowel here versus glide production as is observed in the previous token; here intensity begins its ascent immediately after production of [l], whereas in the preceding figure the intensity contour shows continued descent until attaining a low during the production of [j]. We identify two peaks in intensity during vocalic production divided by a period of transition marked by a dip in intensity, and energy loss on the spectrogram. Two peaks in intensity after [l] indicate dieresis, as corroborated by our judges' findings.

We turn now to the analysis of tokens for the stimulus 28 *scier* realised with syneresis and with dieresis. Figure 17 below shows the oscillogram, spectrogram and intensity contour for the token 31ajg1 28 *scier* judged categorically to be realised with syneresis.

---

15 Judgements vary for this token: two epenthesis and one hiatus.

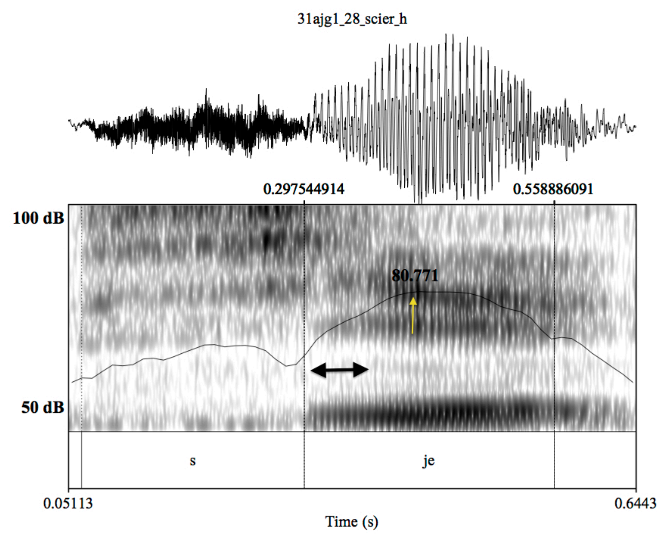


Figure 17: Acoustic cues, Praat spectrogram with intensity contour and oscillogram (above) for the token 31ajg1 28 *scier* (syneresis)

We note that signs of F2 and F3 transitions are not as evident as is observed in other tokens. Energy at the onset of vocalic production (marked by the double headed arrow) is weak which we take to indicate glide production. Increased energy in F2 and F3 correlates with a single peak of intensity which correlates with the production of [e]. Maximum intensity (80.77 dB) is attained at the peak of [e] production. A single peak of wave amplitude on the oscillogram corroborates these observations and a finding of syneresis.

In Figure 18, we examine a token for the same stimulus 28 *scier* that is judged by consensus to be realised with dieresis. Judgements show categorical dieresis: 2 for epenthesis and 1 for hiatus. We contrast our findings to the token above realised with syneresis.

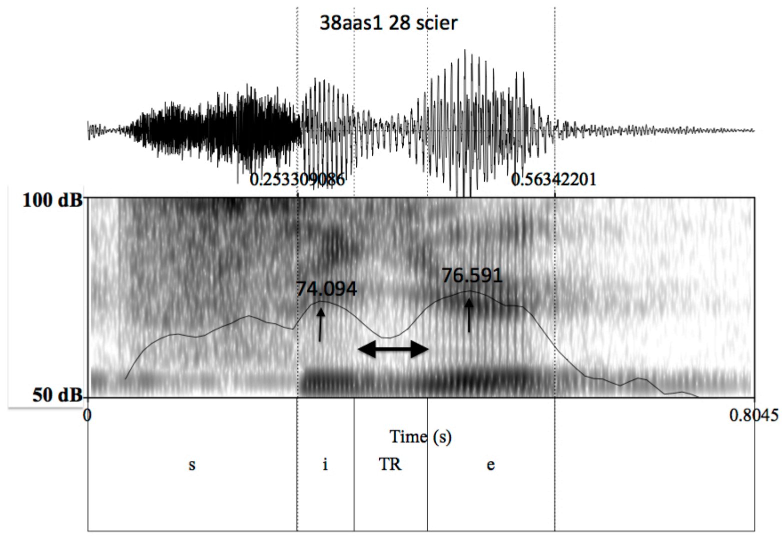


Figure 18: Acoustic cues, Praat spectrogram with intensity contour and oscillogram for the token 38aas1 28 *scier* (dieresis)

The contrast with the preceding token, Figure 17, is striking: we note in particular that the [s] in the preceding token shows signs of formant structure whereas here this is not the case. On the figure above, formants are mostly evident during vocalic production which shows two clear peaks in energy (oscillogram) separated by considerable energy loss during transition. Two peaks on the intensity contour corroborate dieresis; the first measures 74.09 dB at its maximum and the second sustained peak measures 76.59 dB at its maximum. Intensity from the preceding [s] shows a somewhat more dramatic transition to the high vowel [i] as compared to the gradual ascent observed in the preceding example for production of a glide [j]. Cues from intensity, amplitude, formant transition and energy indicate unambiguously that this token is produced with dieresis. We observe F2 and F3 transitions from [i] to [e], marked also by considerable energy loss.

In the following discussion, we examine spectrographic images with intensity contour and wave amplitudes (on oscillogram) for the minimal set *miette~muette~mouette*. We observe tokens for each stimulus that, by consensus, are judged to be either syneresis or dieresis. Examining the minimal set provides an opportunity to compare tokens for syneresis and dieresis for each high vocoid /I, Y, U/ in an identical phonological

environment /m__ɛt/. The preceding tokens exhibit HVV contexts involving the mid-closed vowel [e] (with the exception of tokens for the nasal vowel observed in the stimulus 51 *influence*); the minimal set *miette~muette~mouette* involves the mid-open [ɛ]. We will examine differences observed in transitions from the nasal consonant for each high vocoid, as well as visual acoustic cues in the period of vocalic production.

The figure below shows the token 69asg1 63 *miette* judged by consensus to be realised with syneresis.

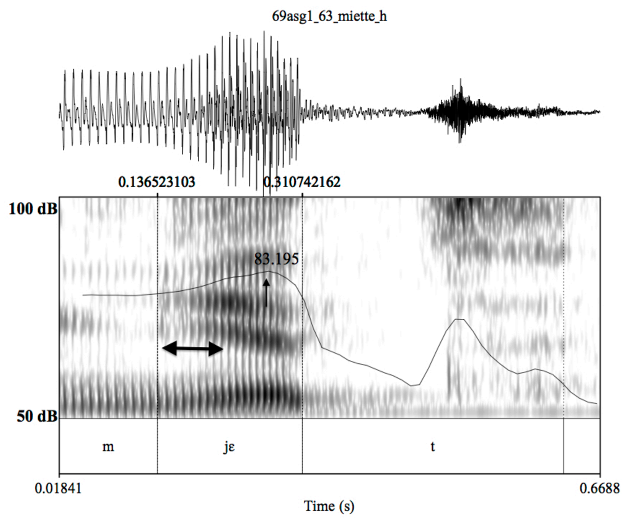


Figure 19: Acoustic cues, Praat spectrogram with intensity contour and oscillogram for the token 69asg1 63 *miette* (syneresis)

Different from other preceding tokens showing non-nasal onset consonants, the nasal consonant [m] shows consistent intensity throughout its production averaging 77.80 dB. Transition to the following high vocoid is, therefore, much less pronounced; marked by a very gradual ascent to a single peak at 83.20 dB. Vocalic production begins with weakened formant energy which we take to indicate glide production before the following mid-open vowel [ɛ]. Vocalic production shows a single peak of intensity corresponding to [ɛ]. This single peak of intensity corroborated by a single period of enhanced formant energy on the spectrogram and a single peak in wave amplitude on the oscillogram points to a realisation with syneresis and is confirmed by our judges' unanimous finding for syneresis.

To this point, transition from an onset consonant to the following high vocoid has been easily identified by noticeable energy loss, considerably lowered intensity and diminished wave amplitude during consonant production. Here, the more sonorous nasal consonant shows a reduced transition to the following high vocoid.

We turn now to a token for the same stimulus judged to be realised with dieresis. Figure 20 below shows a token for the stimulus 63 *miette* realised with dieresis.

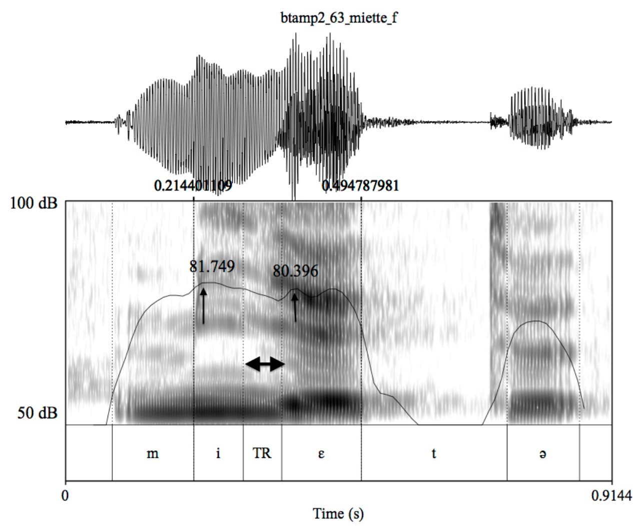


Figure 20: Acoustic cues, Praat spectrogram with intensity contour and oscillogram for the token btamp2 63 *miette* (dieresis)[16]

In contrast to the preceding token, the intensity contour shows a more evident transition. We observe a considerable ascent on the intensity contour from [m] to [i]. Here, where the token for dieresis begins with the high vowel [i] rather than [j], the intensity contour climbs swiftly to a peak during production of the high vowel. On the spectrogram, we observe that vocalic production shows two distinct periods separated by a clearly discernable transition (double headed arrow). This too, is in contrast to the preceding token which shows a single (gradual) peak of intensity during production of [ɛ] (syneresis). From this first peak, intensity descends gradually throughout the period of transition between vowels reaching its lowest measure at 77.66 dB before rising again for

[16] Judgements for this token vary: 2 for epenthesis and 1 for hiatus.

the production of [ɛ]. Note that given more sonorous quality of the preceding nasal, intensity peaks during vocalic production are less dramatic then we observe in our preceding tokens.

We turn now to analysis of syneresis versus dieresis for the stimuli 79 *muette* containing the high vocoid /Y/, Figure 21.

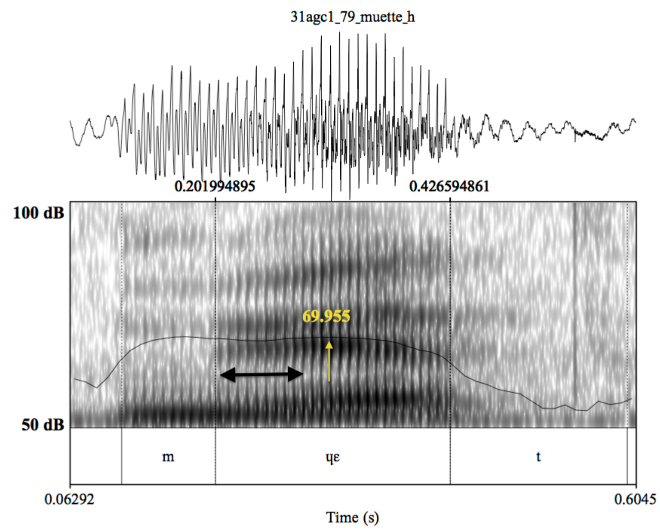


Figure 21: Acoustic cues, Praat spectrogram with intensity contour and oscillogram for the token 31agc1 79 *muette* (syneresis)

Once again the preceding nasal environment has a dramatic effect on the shape of the intensity contour. The transition from the preceding [m] to the following front rounded high vocoid /Y/ shows a very gradual dip of intensity that reaches its lowest measure (69.23 dB) during transition from the glide [ɥ] (note also the slight drop in wave amplitude on the oscillogram). This dip, in contrast to the ascending contour observed in the preceding token for 63 *miette* /I/ realised with syneresis (cf. Fig. 20), is a reflection of the different high vocoid /Y/. This change in high vocoid also produces only minor formant transitions from the glide [ɥ] to [ɛ]. Nonetheless, after the glide we observe a period of sustained formant strength during [ɛ]. Finally, this token shows a single (very subtle) peak on the intensity contour (confirmed by an intensity listing in Praat 69.96 dB) corroborating our judges' findings for syneresis.

We turn now to a token for the same stimulus 79 *muette* judged by our panel to be realised with dieresis.

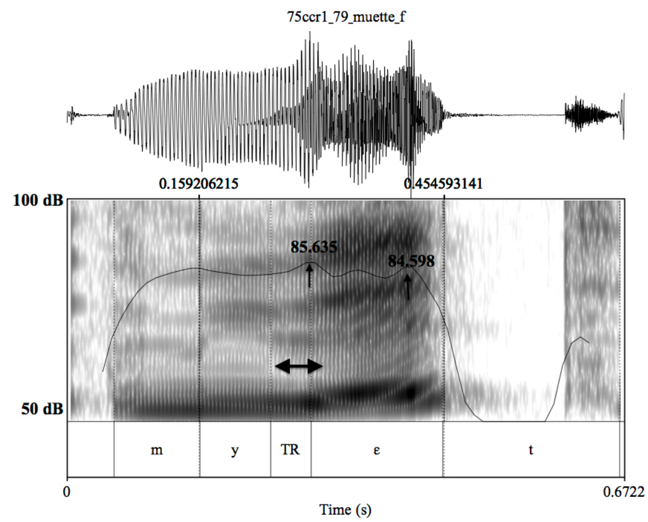


Figure 22: Acoustic cues, Praat spectrogram with intensity contour and oscillogram (above) for the token 75ccr1 79 *muette* (dieresis)[17]

As observed for the preceding token involving the high vocoid /Y/, transition from the onset nasal is marked by a dip of intensity immediately after [m]. Intensity continues to dip throughout the production of [y]. It appears that it is only in transition that intensity climbs to its greatest peak (85.64 dB) then transitions to a secondary peak during the vowel [ɛ] (84.60 dB). The spectrogram above indicates that this token begins with a sustained high vowel [y] but peak intensity is not attained until after formants begin transition toward [ɛ]. We identify, nonetheless, production of two distinct vowels, although the first is much weaker than the second with respect to formant energy. Transition is observed (double headed arrow) with a moderate energy loss before energy peaks during front mid-open vowel production. The preceding nasal influences the high vocoid production differently than is observed for [l] and [R] (also sonorants). The nasal consonant [m] appears to attenuate intensity of the high vocoid. These observations, corroborated by our judges' findings, confirm that this token is realised with dieresis.

[17] Findings show consensus for dieresis with 2 judgements for hiatus and 1 epenthesis.

Finally, we examine tokens for the stimulus 30 *mouette* realised with syneresis and with dieresis. This stimulus allows us to compare the high vocoid /U/ to the high vocoids /I/ and /Y/ in identical environments. Figure 23 below shows a token for 30 *mouette* realised with syneresis.

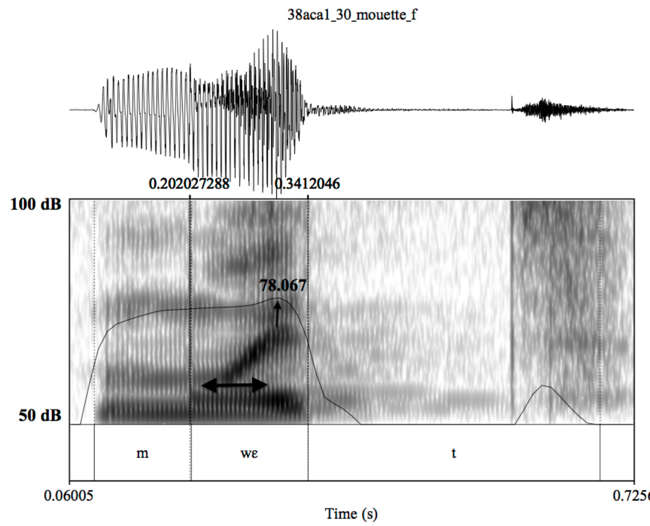


Figure 23: Acoustic cues, Praat spectrogram and oscillogram with intensity contour for the token 38aca1 30 *mouette* (syneresis)

The spectrographic and oscillographic evidence shown is clear; this token exhibits a single period of vocalic production, evident in the intensity contour, corroborated by formant energy on the spectrogram and the single peak of wave amplitude following [m] on the oscillogram indicates syneresis. From the oscillogram, we observe that vocalic production immediately following the burst of [m] shows low energy which is further corroborated by the formant energy observed on the spectrogram. The earliest part of glide production shows a modest increase in intensity transitioning from [m]. Beyond this initial period of low energy, we observe increasing energy in F1 and F2, increasing intensity as well as the unmistakable transition of F2 towards the production of [ɛ] where formants strengthen and stabilise. During this same period, we observe the steady increase in intensity which attains its peak at 78.07 dB corresponding to the peak in the waveform on the oscillogram. Acoustic cues observed here corroborate our judges' findings for syneresis.

Figure 24 below shows a token for the stimulus 30 *mouette* judged categorically realised with hiatus (dieresis). In this particular token, we observe that the effects of the surrounding environment /m__ɛt/ render peaks of intensity less obvious. As we mentioned previously intensity during vocalic production is attenuated by the preceding nasal. We do identify two periods of increased energy separated by a considerable dip in amplitude on the oscillogram; however, the distinction between vowel production (syllables) is less perceptible on the intensity contour.

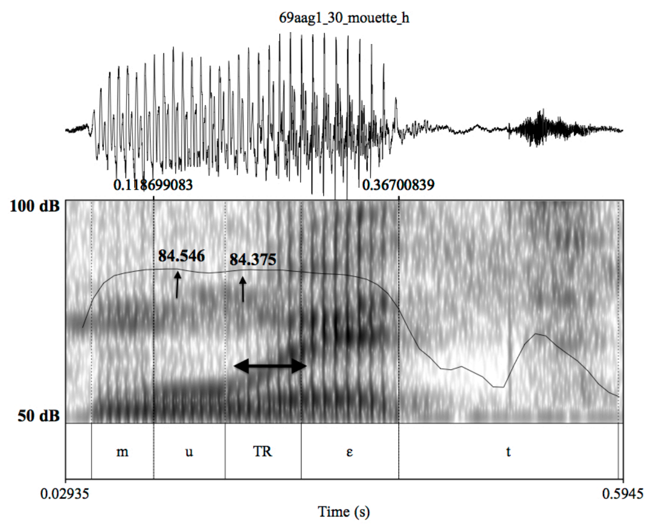


Figure 24: Acoustic cues, Praat spectrogram and oscillogram with intensity contour for the token 69aag1 30 *mouette* (dieresis)

The near imperceptibility of peaks in intensity in this context demands that we look to other visual cues such as wave amplitude and formant strength and transitions for a reliable diagnostic. Based on the strength of implications from these cues, we consult the intensity listing provided in Praat to confirm our diagnostic. Intensity listing in Praat confirm two moderate peaks of intensity (dieresis) identified on the figure above.

Our analysis of tokens from the minimal set *miette~muette~mouette* has shown that the preceding environment and the high vocoid quality will influence the visual cues. The intensity contour shows variable transitions from preceding consonant to the high vocoids /I/ and /U/ intensity climbs after [m] whereas /Y/ drops slightly after [m]. Transitions (TR) between vowels (dieresis) are much less evident on the intensity contour however

other acoustic cues are largely unmistakable and can be relied upon in order to make a reliable diagnostic.

Since we use the observations made here to inform our analysis of spontaneous speech data in Chapter 6, we consider these factors. We have shown here that in most cases, distinctions between syneresis and dieresis are readily identified on the acoustic signal. However, in less obvious cases (as observed for tokens of the minimal set) a reliable diagnostic may be made using corroborating evidence across a spectrogram and oscillogram confirmed by the intensity listing in Praat.

Our examination of the word-list data shows that distinction between syneresis and dieresis is made reliably using the intensity contour supported with evidence of other acoustic cues available in Praat. We have shown that despite inter speaker variation (of frequency, speech rate, and even accentuation), cues from the intensity contour provide a reliable visual cue for the identification of syllable peaks; syneresis and dieresis can be reliably identified using the visual cues of acoustic analysis in Praat.

Note that judgements obtained for the different tokens observed here show that tokens exhibiting syneresis are largely uncontroversial, their findings show consensus among judges, whereas tokens judged realised with dieresis show consensus (epenthesis versus hiatus) much less frequently. More study to examine the role of different preceding environments (voiced and voiceless stops, fricatives and the full gamut of nasals) is merited. Given the frequent reports of the different results for GF particularly concerning preceding [l] versus preceding [ʀ] and the blocking effect of Obstruent + Liquid sequences, any future study of the effect of preceding environments on GF must include these preceding liquids.

In the following section, we discuss our methodology including the role of acoustic cues, and findings for our study of spontaneous speech data. Our analysis of spontaneous speech data will rely on the intensity contour supported with corroborating evidence from other acoustic cues such as energy observed on the spectrogram and wave amplitude observed on the oscillogram to distinguish between dieresis (presence of a high vowel

plus vowel sequence HV(G)V) versus syneresis (corresponding glide plus vowel sequence GV).

We turn now to our primary spontaneous speech study of /HV(+)V/ contexts using the PFC interview data. In Chapter 6, we discuss the methodology and results for this study.

Chapter 6

6 Empirical Study: Glide phenomena in Spontaneous Speech

> Language reality is far more varied and far less homogenous than many descriptivists would be willing to concede.
>
> (Martinet 1962:4)

Martinet's observation about the varied nature of language reality dates to more than 50 years ago, and yet the studies of variability in modern French are relatively few. It is our goal then to examine the variability of glide realisations in an empirical study that investigates the reality of glide phenomena in modern Hexagonal French.

The goals of the present study are to examine glide phenomena, specifically variability, in a fresh set of data while also examining the classic glide contexts that have become central to any discussion of French glides with the goal to accurately describe glide phenomena at the beginning of the 21st century. Informed by the findings from our judgement task as a methodological tool, we examine spontaneous speech data from excerpts of the guided discussion in the PFC database for six surveys from three regions of Hexagonal French: *Île de France*, *Normandie* and the *Rhône Alpes* regions, with the goal to code then describe gliding phenomena and the prevalence of variability (if found) across the various glide contexts. In the following sections, we discuss the PFC protocol while also emphasizing some of the advantages of the various tasks for our studies of glide phenomena. Our study of glide phenomena is conducted in two parts. Our discussion of glide descriptions and subsequent theoretical models illustrates the prominent role of glides in descriptive literatures and glide phenomena in theoretical models of Hexagonal French. The goal of this study is to examine glide contexts in the spontaneous speech of Hexagonal French speakers to observe the degree to which /HVV/ realisations are as predictable as is described (realised with syneresis versus with dieresis) across lexicalised and derived contexts and each high vocoid: /I Y U/. To this end, we examine all glide contexts both lexicalised and derived to observe glide behaviour in a

fresh set of data[18]. As mentioned above in the results of the judgement task IN SECTION XX descriptive results are reported below without measures of statistical significance, given the small Ns that are somewhat imbalanced across regions and speakers.

Recall from our discussion of variability in Chapter 1 (cf. Table 1 reproduced in Table 18 below), we discussed regional tendencies for realisations with syneresis versus dieresis as schematised in Table 18 below:

Table 18: Pan dialectal variation (Heap and Kelly 2005 based on Tranel 1987 and Lyche and Girard 1995)

Word	*nier*	*nouer*	*nuer*	
a. Parisian French	[nje]	[nwe]	[nɥe]	synérèse
b. Northern French	[nje]	[nwe]	[ny.e]	diérèse
c. Midi French	[nje]	[nu.e]	[ny.e]	
d. Media French	[ni.e]	[nu.e]	[ny.e]	

Our study examines glide realisations in lexicalised GV and derived /HV+V/ contexts observed in the spontaneous speech samples from PFC guided interviews. Given the variability as schematised in Table 18 above, we have chosen six PFC surveys from three distinct regions: *Normandie*, *Île de France*, and *Rhône Alpes* (we examine the less described southeast rather than the southwest (Midi French). We begin with PFC surveys from *Île de France* (75c *Paris* and 91a Brunoy). *Île de France* being the epicenter of the standard variety (noted as Parisian French above) upon which so many descriptions are based. These data provide a baseline against which all other results may be compared. The second region to be examined is *Normandie*, a northern coastal region situated to the northeast of *Île de France* which includes data from the PFC surveys 50a *Brécey* and 61a *Domfrontais*. Finally, we examine data from the southeastern region of *Rhône Alpes* including the two surveys 38a *Grenoble* and 69a *Lyon*.

[18] We have chosen to exclude derived contexts where the existence of a glide in the radical is uncertain: nettoie[j]~nettoyer.

Each transcribed interview excerpt provides minimally five minutes (300 s.) of discourse for the examination of glide contexts. Using the acoustic analysis program Praat, we search the PFC orthographic transcriptions using orthographic conventions for high vowels in order to identify all /HV+V/ and GV contexts (120 below). Once identified, glide contexts are examined in Praat using the evidence from acoustic cues studied in our judgement task data (intensity contour with corroboration from wave amplitude and formant energy, cf. section 5.5). Tokens are coded for high vocoid involved, category: lexicalised versus derived, and realisation with syneresis or dieresis. The oft-described contexts that we examine are shown below (reproduced from 21 below).

(120) French glides

Derived Glides: /HV+V/→[GV] or [HVGV]

a. Glide formation (high vowel~glide alternations): /HV+V/→[GV]
  iv. [i]→ [j]: *lier* /li+e/→ [lje] 'to tie'
  v. [u]→ [w]: *jouer* /ʒu + e/ → [ʒwe] 'to play'
  vi. [y]→ [ɥ]: *suer* /sy + e/ → [sɥe] 'to sweat'

b. Epenthetic glide (high vowel spreading): /OLHV+V/ → [OLHV(G)V]
  vii. [i] → [i.j]: *plier* /pli+e/ → [pli.je] 'to fold'
  viii. [u] → [u.w]: *trouer* /tʀu+e/ → [tʀu.(w)e] 'to pierce'
  ix. [y] → [y.ɥ]: *fluer* /fly + e/ → [fly.e], [fly.(ɥ)e] 'to flow'

According to canonical descriptions, derived glide contexts necessarily involve a high vowel in stem final position. When a vowel initial suffix is added to a morpheme ending in a high vowel the potential HVV (hiatus) is avoided through glide formation resulting in a surface [GV] (120a.). We include among derived glide contexts realisations with dieresis and an epenthetic glide (120b., V.GV). Given the prevalence of cross-word-boundary contexts resulting in glide formation, we include also derived cross-word-boundary contexts (e.g. *fourni avec* [$nja], *qui avait* [$kja]) in our study.

Lexicalised contexts (121 below reproduced from 31 below), by their nature, are generally described as invariable, however, studies have shown (cf. Heap et al. 1992) that speakers treatments do not always conform to these descriptions. Exceptional variation may be attributed to an individual speaker's stylistic choice.

(121) French glides

Lexicalised Glides: [GV], et [G] onset [j, w], coda [j], nucleus [GV]

c. Absolute initial [G]:
   i) *hi*V- : *hier* [jɛːʀ] [i.jɛːʀ] [i.ɛːʀ] 'yesterday'; *hiatus* [(ˈ)ja.tys] 'hiatus'
   ii) *i*V- : *iambe* [jɑ̃b] 'iambe'; *iode* [jɔd] 'iode', *ion* [jɔ̃] 'ion'
   iii) *y*V- : *yaourt* [jaurt] 'yoghurt', *yeux* [jø] 'eyes', *yéti* [jɛ.ti] 'yeti'
   iv) *w*V- : *wiskey* [wi.ski] 'whiskey', *wallon* [wa.lɔ̃] 'walloon', *watt* [wat] 'watt'
   v) *ou*V- : *ouest* [wɛst] 'west', *ouïe* [wi] 'hearing'

d. Medial (onset): V$jV
   i) -V*y*V- [j]: *crayon* [cʀɛ.jɔ̃] 'pencil', *moyen* [mwa.jɛ̃] 'average'
   ii) -V*ille*V-: *ailleurs* [a.jœːʀ] 'elsewhere',

e. Nuclear: [GV]
   i) *-oi-* [wa]~[wɛ̃]: *proie* [pʀwa] 'prey', *coin* [kwɛ̃] 'corner'
   ii) *-ui-* [ɥi]: *huitre* [ɥitʀ] 'oyster', *truite* [tʀɥit] 'trout', *huit* [ˈɥi(t)] 'eight'

f. Lexicalised: CjV (historically nuclear)
   i) *-ien* [jɛ̃], *bien* [bjɛ̃] 'well', *chien* [ʃjɛ̃] 'dog', *rien* [ʀjɛ̃] 'nothing',
   ii) *-ier* [je] : *acier* [asje] 'steel', *panier* [panje] 'basket'
   iii) *-e-*/*-ie-*: venir/vient [vjɛ̃] 'to come/comes', *tenir*/*tient* 'to take/takes'

g. Final (coda) $Vj
   i) -V*il*: *œil* [œj] 'eye', *soleil* [sɔ.lɛj] 'sun', *éventail* [evɑ̃.taj] 'fan'
   ii) -V*ille: volaille* [vɔ.laj] 'poultry', *groseille* [gʀo. zɛj] 'currant'

h. Lexicalised suffixes
   i) *–ien* [jɛ̃] *Paris* → *parisien* adj.
   ii) *–ier* : *palmier, orangier, poirier*, *saladier, ouvrier, mûrier*
   iii) *–ième* [jɛm] [ijɛm]: *dixième* [di.zjɛm] ; *quatrième* [ka.tʀi.jɛm]

(Rey 1963, Juilland 1965)

In the following sections (6.1.1-6.1.4), we discuss our methodology for studying glide contexts in the spontaneous speech sample provided by the guided interviews from the PFC corpus. We begin with a description of the PFC surveys and the regions studied (section 6.1.1). In section 6.1.2, we discuss the distributions of tokens for speaker and numbers of speakers across region. In section 6.1.3, we discuss our methodology for token identification and analysis in PRAAT. In 6.1.4 we elaborate our system of coding the tokens observed. In section 6.2, we discuss the overall results of our study. Section 6.3 examines the variability observed across derived and then lexicalised contexts.

Section 6.4, we discuss the distribution of tokens across six surveys in three regions before turning to an analysis of findings for each region in sections 6.5-6.7. In section 6.8 we provide a summary of our findings.

6.1.1 HVV environments in Spontaneous Speech Data

The goal of this our primary study is to examine distribution and realisations of lexicalised and derived glide contexts across three high vocoids in the spontaneous speech of Hexagonal French speakers. Each token identified is coded in Praat for the context (lexicalised or derived), the environment in which it occurs, the high vocoid involved and the realisation observed (syneresis versus dieresis). The spontaneous speech sample comes from excerpts of guided conversation interviews from the PFC corpus for the surveys we have selected. These excerpts are taken from six PFC surveys: 50a *Brécey* and 61a *Domfrontais* (Normandy); 75c *Paris Centre ville* and 95a *Brunoy* (*Île de France*); 38a *Grenoble* and 69a *Lyon* (*Rhône Alpes*), we have retained recordings for analysis based on their acoustic quality. Our analysis necessarily relies on the visibility of the acoustic signal and in particular the various acoustic cues we identified in section 5.5 below. Some recordings where the cues are not clearly visible have been eliminated.

6.1.2 Distribution of tokens across speakers

Across three regions (*Normandie*, *Île de France*, *Rhône Alpes*), our study of glide contexts in spontaneous speech results in 3415 tokens identified in six PFC surveys. Our sample includes 31 speakers, of which 17 (54.8%) are female and 14 (45.2%) male. Numbers of speakers for each region are roughly representative of the populations for each region i.e. more tokens in *Île de France* than in *Rhône Alpes* than in *Normandie* (Table 19 below).

Table 19: Number of speakers by region (percentage calculated over total speakers in sample) compared to population by region (percentage calculated over total population for regions)[19]

	Speakers by region	Population by region
Île de France	13 speakers (41.9%)	11 728 240 (59.4%)
Normandie	8 speakers (25.8%)	3 303 822 (16.7%)
Rhône Alpes	10 speakers (32.3%)	6 174 040 (31.3%)
Total	31 speakers	19 735 222

The tables below provide a summary of our findings for each region by speaker for each survey beginning with *Île de France* (below). *Île de France*, the most populous of the three regions in our sample, represents 41.0% of overall tokens (3415 identified).

[19] **Source**: Insee http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/populations-legales/france-regions.asp?annee=2009

Table 20: tokens by speaker for each PFC survey (*Île de France*), each speaker (identified by PFC speaker code), age, sex, N= number of tokens (for each speaker)

Île de France	Speaker (13)	Age[20]	Sex	N
Paris *Centre ville* (75c)	75cab1	70	F	169
	75ccm1	34	M	192
	75ccr1	35	F	75
	75csb1	26	F	159
	75cvl1	26	F	152
Total	**5 speakers**		**F(4)/M (1)**	**747**
Brunoy (91a)	91aal1	27	F	68
	91acs2	22	M	80
	91db1	54	M	64
	91aed1	73	F	84
	91ael1	73	F	74
	91ajc1	63	M	103
	91amb1	62	M	74
	91asl1	64	F	91
	91asl1	64	F	91
Total	**8 speakers**		F (5)/M (4)	**638**
Region total	**13 speakers**		F (9)/M (5)	**1385**

Unfortunately, balance for the age/sex is not achieved in our data since our primary concern for the visibility of the signal requires that sone interviews be excluded for reasons of acoustic quality. Unfortunately, imbalance for age and sex across speakers by survey prevents meaningful cross-speaker analysis of data by social factor.

In Table 21, we provide an overview of tokens observed for each survey by speaker (with age, and sex) for *Normandie* (least populous region examined).

---

[20] Ages not provided in the PFC database have been approximated; according to the information provided the survey was conducted over a two-year period between 2004 and 2006. Using 2005 as an average date of survey, we have estimated speaker's age from the date of birth provided in on-line database: http://www.projet-pfc.net/enquetes.html.

Table 21: tokens by speaker for each PFC survey (*Normandie*), each speaker (identified by PFC speaker code), age, sex, N= number of tokens

Normandie	Speaker (8)	Age	Sex	N tokens
Brécey (50a)	50ajp1	19	M	76
	50alb1	55	F	158
	50app1	69	F	148
	50apb1	42	M	203
Total	**4 speakers**		**F (2) /M (2)**	**585**
Domfrontais (61a)	61agr1	74	F	134
	61ahd1	83	M	96
	61ahl1	20	M	111
	61ajh1	48	F	207
Total	**4 speakers**		**F (2) / M (2)**	**547**
Total Region	**8 speakers**		**F (4) / M (4)**	**1133**

Normandy, our least populous region retains the least number of speakers yet we observe considerably more tokens overall (33.2%) averaging 141.6 per speaker. The region combined shows good age/sex equilibrium across speakers.

Table 22 shows a summary of the tokens observed by speaker for *Rhône Alps*.

Table 22: tokens by speaker for each PFC survey (*Rhône Alpes*), each speaker (identified by PFC speaker code), age, sex, N= number of tokens

Rhône Alpes	Speaker (10)	Age	Sex	N tokens
Lyon (69a)	69aag1	51	M	76
	69ajd1	46	F	99
	69all1	74	F	63
	69asg1	25	M	83
Total	**4 speakers**		**F (2) / M (2)**	**321**
Grenoble (38a)	38aas1	68	M	135
	38aca1	49	M	102
	38agp1	66	F	55
	38aym1	28	M	91
	38amb1	22	F	115
	38aep1	33	F	78
Total	**6 speakers**		**F (3) / M (3)**	**576**
Region total	**10 speakers**		**F (5) / M (5)**	**897**

Rhône Alpes, our second most populous region with the second highest number of speakers, shows the lowest rate of tokens per speaker (89.7 tokens). The 897 tokens for the region represent 26.3% of overall tokens identified.

In the following section, we discuss the methods used to identify and code all glide contexts in the PFC interviews using the Praat program. Glides are coded in Praat (version 5.2.17). We retain the PFC speaker code for indexation purposes.

6.1.3 Acoustic Analysis: token identification (methodology)

The figure below shows the Praat spectrogram and oscillogram with the PFC aligned transcription tier. The speaker code 75ccm1 identifies the survey Paris – *Centreville* survey (75c). The individual speaker is identified as cm1. Demographic data for each speaker (age, sex, years educated, years lived in or away from the survey location, languages spoken etc.) is recorded in the *fiche signalétique* and stored in the PFC database. Figure 25 below shows a selection wherein we have identified the lexicalised glide context *suite*.

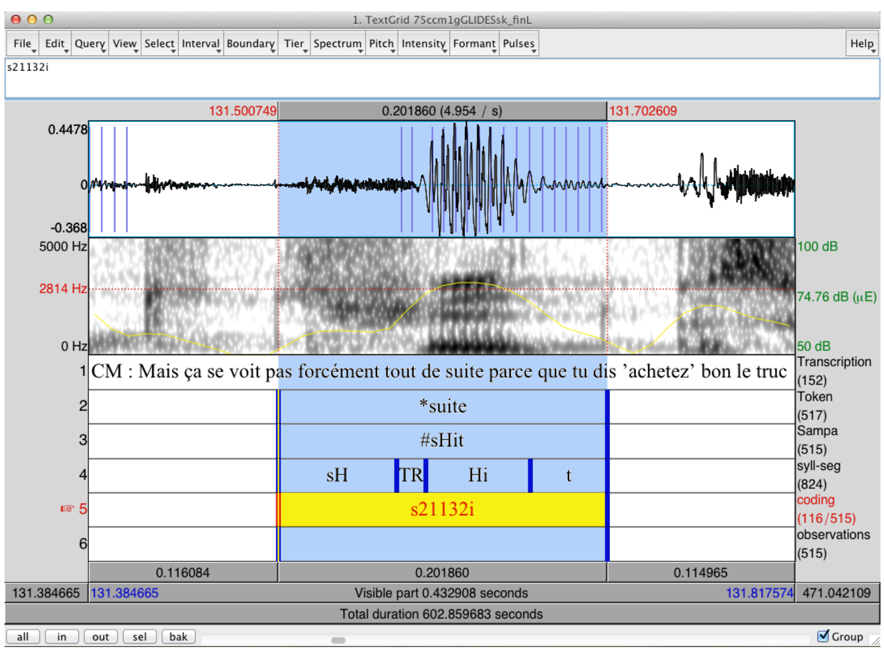


Figure 25: Praat coding tiers

Figure 25 above shows the token *suite* in Praat with oscillogram, spectrogram and Praat.textgrid. The first tier labelled Transcription is an orthographic transcription aligned with the audio recording of the discussion provided in the PFC database. By searching the orthographic transcription provided in the PFC database using orthographic conventions for high vocoids followed by another vowel, we are able to identify all glide contexts (lexicalised and derived). In each Praat.TextGrid file, we add five supplemental tiers (orthographic transcription, SAMPA transcription, segmentation, coding and observations) in order to record all pertinent information. Not all information recorded will factor into this analysis but may be exploited in future studies. On Tier 2 we label each token in isolation using orthographic transcription. Tier 3 shows a Speech Assessment Methods Phonetic Alphabet (SAMPA) transcription for the token *suite*[21]. On Tier 4, we conduct an initial segmentation and on Tier 5 we add our code reflecting the objectives of our analysis: lexicalised versus derived glide contexts, high vocoid involved, all potential variants (syneresis, dieresis, and high vocoid deletion), preceding and following phonological environments, observations. All added intervals are aligned at the left and right edges. Aligned intervals are extracted from each tier separately maintaining their order and time relative to the original recording (.wav) file for indexation potential to reference to original recordings. All intervals, compiled by tier, are entered in an Excel spreadsheet where we have constructed a searchable data set.

In the following section, we explain our coding system used to identify each glide context as shown on Tier 5 (above).

6.1.4 PRAAT: coding system

Each glide context is coded using a (minimally) seven-figure code to capture all pertinent information for our analysis including the high vocoid involved, context (lexicalised or derived) and the surrounding phonological environment. Table 23 below shows a

[21] http://www.phon.ucl.ac.uk/home/sampa/

character-by-character breakdown of our coding system using the example of the lexicalised token *suite* s21132i:

Table 23: coding system broken down for the token *suite* s21132i: each character is given in the left hand column with a summary of the significance for each character given to the right

Code for *suite*	Context
s	Preceding phonological context (up to three possible consonants – s, sk, skr, for example, or word boundary #.
2	High Vocoid (1, 2, 3) = /I, Y, U/, respectively
1	Phonological status: **lexicalised**: 0 = word medial intervocalic; 1 = all positions in the word **derived**: 2 = phonologically derived across morpheme boundary/HV+V/ 3 = phonologically derived across-word-boundary /HV#V/
1	Realisation of glide context with syneresis: 0 = syncope 1 = GV syneresis 2 = dieresis hiatus 3 = dieresis epenthesis
3	Preceding Segment: 0 = vowel or word boundary 1 = nasal consonant 2 = non nasal voiced 3 = non nasal voiceless ,
2	Co-articulatory effects: 0 = not pertinent 1 = devoiced vocoid 2 = palatalisation of preceding segment 3 = syncope 4 = preceding segment is a liaison consonant 5 = preceding segment shows regressive assimilation 6 = schwa epenthesis
i	Following phonological context (vowel immediately following a high vocoid). SAMPA transcription (exception for nasal vowels coded as follows: 1= e~ /ɛ̃/, & = a~ /ɑ̃/, 3 = o~ /ɔ̃/, 4 = 9~ /œ̃/

Orthographic representations for preceding and phonological environments may include up to three characters for preceding consonants ([str] for example) and a single character for a following vowel. Vowel notations are as follows: twelve oral vowels, (i/j), (y/H), (u/w), e, E, a, A, O, o, 2 (mid-closed ø), 9 (mid-open œ) @. In order to reduce nasal codes to a single character and to avoid conflict with characters already used in the oral

vowel system the four nasal vowels are coded in SAMPA as follows: 1= e~ /ɛ̃/, & = a~ /ɑ̃/, 3 = o~ /ɔ̃/, 4 = 9~ /œ̃/. Table 23 above shows a breakdown of our code for the token *suite* (#sHit, in SAMPA transcription). The token contains the lexicalised GV sequence [ɥi] coded: s21132i. Central to the code is a five-digit code which captures all information about the high vocoid and the type of context in which it is found. The first three digits encode three primary criteria for our study: vocoid (I, Y, U), lexicalised/derived context (GV/HVV), realisation with syneresis (GV) or with dieresis (hiatus HVV or epenthesis HVGV). The glide context is preceded by *s*, the digits 211 indicate that a front rounded vocoid (2 encodes the High Vocoid /Y/ (1=/I/; 2 = /Y/; 3= /U/), occurs in a lexicalised context (1 = lexicalised GV sequence), realised with syneresis *1* is preceded by a voiceless obstruent *3* which exhibits some palatalisation (coarticulation with a preceding voiceless obstruent (*tout de suite* (tuts'ɥit) *2*. Finally, the glide context is followed by *i*.

A final note: two digits (shaded grey area on table above) were added to our code in order to codify notable observations made during our analysis of high vocoids in these contexts but do not figure in our analysis. Many observations occur with such regularity that we find it necessary to note them. Given their frequency, we have found it more efficient to include these as part of our coding system. These codified observations include numeric plus parentheses (_), to indicate a word initial/word-final glide and by implication possible liaison (initial position) or resyllabification (final position) contexts, deletion of the high vocoid or a preceding segment (usually a liquid involved in a preceding complex onset (OL), or co-articulation effects such as segment deletion, devoicing, and/or palatalisation.

In the following sections, we examine the distribution of lexicalised and derived tokens found in spontaneous speech data from all 31 speakers in our sample.

6.2 Spontaneous speech analysis results for lexicalised and derived data by region

We continue our discussion of our spontaneous speech study with the overall distribution of data across regions followed by distribution of HVV across specific contexts: lexicalised, derived (including cross-word-boundary), and word medial derived (cross morpheme boundary). Finally in the sections to follow, we discuss the variability of realisations observed. Since lexicalised forms are less variable we will focus on variability observed in the derived glide contexts and the surrounding phonological environments which may be a factor in the variability observed.

Table 24 shows the distribution of all tokens for each region across three high vocoids.

Table 24: Distribution of tokens for all /HVV/ environments across three high vocoids by region, N= number, percentages (%) for each high vocoid /I U Y/, calculated by region

Region	Vocoid							
	I		U		Y		**Total**	
	N	%	N	%	N	%	**N**	**%**
Île de France	770	55.6	371	26.8	244	17.6	**1385**	**40.6**
Normandie	656	57.9	320	28.2	157	13.9	**1133**	**33.2**
Rhône Alpes	466	52.0	269	30.0	162	18.1	**897**	**26.3**
Grand Total	**1892**	**55.4**	**960**	**28.1**	**563**	**16.5**	**3415**	**100**

Our data show consistently that /I/ is most prevalent averaging (55.4%) followed by /U/ averaging 28.1% then /Y/ occurring on average at 16.5% of all contexts across all three regions. The rates at which each high vocoid occurs are consistent across all three regions and reflects descriptions of glide distributions as a distributional hierarchy: I ≫ U ≫ Y (Léon 1978, Tranel 1987, Bullock 2002 cf. section 4.8) which corresponds to universal typologies (Greenberg 1963, Maddieson 1984 cf. section 4.5.2).

In the following tables, we discuss findings for lexicalised /GV/ versus derived /HV+V/ glide contexts. Table 25 shows all tokens for lexicalised and derived glide contexts across three regions.

Table 25: Derived versus lexicalised glide contexts across three regions. N= number, percentage (%) calculated by region for derived and lexicalised contexts

Region	Derived		Lexicalised		Total	
	N	%	N	%	N	%
Île de France	190	14.5	1124	85.5	1314	40.9
Normandie	165	15.7	886	84.2	1052	32.7
Rhône Alpes	104	12.3	744	87.7	848	26.4
Total	**459**	14.3	**2754**	85.7	**3213**	100

The data above show lexicalised contexts are more prevalent than derived in the spontaneous speech samples examined. Among tokens for lexicalised contexts, we have eliminated 202 instances of the adverbial pronoun *y* which is realised near-categorically as the glide [j]. While the adverbial pronoun *y* does vary between [i] and [j], variability is predictable; followed by a vowel, the pronoun *y* is realised as a glide [j]. Otherwise followed by a consonant *y* is realised as [i] (*il s'y trouve* [i] versus *il y a* [j]). The remaining 3213 tokens are broken down as follows: lexicalised glide contexts represent 2754 (85.7%) tokens compared to 459 (14.3%) derived tokens. The same pattern of distribution, approximately 85% lexicalised to 15% derived, is observed across three regions.

We turn now to discussion of distribution of tokens for each high vocoid in lexicalised versus derived contexts. Recall that in an overall summary of tokens, we observe the distributional hierarchy: I ≫ U ≫ Y. When we compare distribution of tokens across three high vocoids for lexicalised (Table 26) and derived (Table 27) contexts some interesting yet consistent patterns are revealed. In the table below, we examine the distribution of lexicalised glide contexts for each high vocoid.

Table 26: Distribution for lexicalised glide contexts across high vocoid /I U Y/: percentages calculated by region for each high vocoid

Region	Vocoid by percentage		
	I	U	Y
Île de France	50.8	31.9	17.3
Normandie	49.1	34.9	16.0
Rhône Alpes	46.5	35.9	17.6
Total by vocoid	**49.1**	**34.0**	**17.0**

The distribution of our data shows that lexicalised /HVV/ contexts pattern similarly for each high vocoid across three regions: tokens for the high vocoid /I/ represent approximately 50%, tokens for the high vocoid /U/ represent approximately 33%, and tokens for /Y/ represent approximately 17%. This pattern is consistent with distributions observed previously (cf. all tokens Table 24). Lexicalised contexts show /I/ occurs most often across three high vocoids averaging 49.1% across three regions. The high vocoid /U/ is second most prevalent accounting for 34%, finally, instances of the high vocoid /Y/ are least prevalent at 17% of all lexicalised tokens. Rates of distribution are constant across regions never varying more than 5% from one region to the other.

Below, we examine the distribution of derived glide contexts /HV+V/ for each high vocoid by region. Interestingly, a slightly different pattern of distribution emerges. The table below shows percentages for derived glide contexts calculated for each region across three high vocoids.

Table 27: Distribution of derived contexts for each high vocoid /I U Y/ calculated by region

Region	Vocoid by percentage		
	I	U	Y
Île de France	77.1	7.2	**15.7**
Normandie	83.7	6.6	**9.0**
Rhône Alpes	75.5	2.1	**22.3**
Total by vocoid	**79.3**	**5.9**	**14.6**

When we compare the distribution of derived contexts across high vocoids, we find that, compared to lexicalised contexts, distribution varies with some consistency such that across all tokens for all regions tokens for /I/ increase to average 79.3%, /Y/ is second most prevalent (14.6%) and instances involving /U/ average 5.9%. Overall, tokens for derived contexts consistently show an inverted distributional hierarchy where the front rounded high vocoid is more frequent than the back rounded: /I/ ≫ /Y/ ≫ /U/ with more pronounced variation in rates of occurrence from region to region. While /I/ is still at the top of the changed hierarchy, derived glide contexts privilege /Y/ over /U/. *Normandie*, for example, shows the highest rate among three regions for /I/ (83.7%) and a low among three regions for /Y/ (9%), while /U/ remains most consistent across regions with an

average of 6.6%. *Rhône Alpes,* however, shows the lowest rate of instances for /I/ (75.5%) and for /U/ with only 2.1% of tokens while tokens for /Y/ at 22.3% show the highest rate for /Y/ across three regions. This variation can be attributed to the nature of spontaneous speech.

Elevated rates of /Y/ contexts may be attributed to inclusion of cross-word-boundary derived contexts in our data since lexical items ending in [y] outnumber those ending in [u]. Juilland (1965) lists more lexical items ending in [y] (283) compared with items ending with [u] (146). Also, cross-word-boundary contexts for /Y/ include 34 (54.8 %) instances involving the pronoun *tu* followed by a vowel initial verb (*tu as*, *tu es*, *tu écrives*, for example); these contexts show considerable variability: high vocoid deletion [ta], dieresis [ty(ɥ)a] and syneresis [tɥa]). When we restrict our data to include only word medial derived contexts, distribution of derived word medial contexts across three high vocoids, while fewer (73 tokens), still pattern as observed in Table 27 above: /I/ ≫ /Y/ ≫ /U/.

Table 28 shows the results for world medial derived glide contexts.

Table 28: Percentage word medial derived /HV+V/ glide contexts for each high vocoid /I U Y/, calculated by region

Region	Vocoid by percentage		
	I	U	Y
Île de France	60.0	15.0	25.0
Normandie	66.7	12.1	21.2
Rhône Alpes	60.0	--	40.0
Total by vocoid	**63.0**	**9.6**	**27.4**

Comparing the overall data for /U/, we find that lexicalised contexts represent 95.7% of all tokens for /U/ and derived contexts represent very few at less than 3%. The disparity between rates of lexicalised and derived /U/ contexts can be attributed to the prevalence of lexicalised [w] particularly instances of the nuclear diphthongs [wa] and [wɛ̃] in the lexicon (Léon 1978, Juilland 1965).

Word medial derived contexts show that /I/ occurs less frequently than was observed in lexicalised contexts, yet tokens involving /I/ consistently represent a strong majority of

tokens (63%). Interestingly, the high vocoid /Y/ follows at 27.4% and least prevalent are contexts involving /U/ (9.6%). Note that data for *Rhône Alpes* is entirely void of word medial derived glide contexts involving /U/, while contexts involving /Y/ are comparatively frequent (40.0%). We will examine data by region in more detail later (cf. section 6.9).

In the following sections, we examine the degree of variability in realisations observed for lexicalised and derived HVV contexts. We begin with an examination of findings for lexicalised contexts (section 6.3.1) then we examine findings for derived glide contexts (section 6.3.2) including an examination of the preceding phonological environments.

6.3 Variability in lexicalised and derived contexts

> An often encountered difficulty in using natural speech as data is that it can be difficult to find a sufficient number of comparable tokens of the possible variants of a word.
>
> Reynolds and Nagy (1997:3)

In the following sections, we examine findings for all lexicalised contexts and all derived contexts coded for potential variability of realisation including three variants: high vocoid deletion, dieresis, and syneresis, with a goal of verifying canonical descriptions while better understanding all aspects of glide phenomena in French.

6.3.1 Variable realisations of lexicalised glide contexts

As discussed in Chapter 1, lexicalised contexts, *bien* [jɛ̃], *emploie* [wa], *suite* [sɥit], for example, are described primarily as invariable, realised exclusively with syneresis, (Schane 1968, Kaye and Lowenstamn 1984, Tranel 1987)[22].

---

[22] Forms such as *emploie* and *envoie* [wa] produced with latent (or analogical) off glide [waj], more evident in infinitive verb forms *employer, envoyer* [waje], are excluded from our analysis. While these forms pattern similarly to the example of *payer* [peje] → *paye* [pej], the presence of yod in stem final position in first person forms and in noun forms, (albeit more common in Canadian French), is ambiguaous and beyond the scope of this analysis.

Note that, unlike our judgement task (Chapter 5), a finding of *uncertain* is not included as an option since our analysis, supported by acoustic cues as discussed earlier (cf. section 6.1.4 above), is intended to be definitive in matters of syneresis versus dieresis. However, some uncertainty with respect to epenthesis and hiatus remains. Therefore HVV tokens are coded for dieresis only.

In Table 29, we examine findings for the realisations of lexicalised glide contexts.

Table 29: Realisations of all lexicalised contexts across three variants: DEL = deleted high vocoid, dieresis, syneresis. N= number, (%) calculated for variant by vocoid

Variant	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	5	0.4	10	1.1	73	15.6	88	3.2
dieresis	29	2.1	1	0.1	1	0.2	31	1.1
syneresis	1318	97.5	924	98.8	393	84.2	2635	95.7
Grand Total	**1352**	**49.1**	**935**	**34.0**	**467**	**17.0**	**2754**	**100**

Table 29 shows findings for variability of realisation for all lexicalised contexts across three high vocoids. Each lexicalised context is coded for realisation with: deleted high vocoid (DEL), dieresis, and syneresis. Across lexicalised contexts high vocoid distribution privileges /I/ at 49.1% of tokens followed by /U/ (34%), and finally /Y/ (17%). Syneresis is the most prevalent realisation amongst lexicalised contexts for all three high vocoids averaging 95.7%; a finding which supports modern descriptions of lexicalised contexts. After syneresis, variants tend to correlate with a particular high vocoid; lexicalised contexts involving /I/ show a prevalence for realisations with dieresis more than deleted high vocoids: 29 (2.1%) compared with 5 (0.4%) of all tokens involving /I/. Contexts involving dieresis are primarily preceded by an OL sequence (51.7%): *client(èle)*, *février,* for example. However, tokens realised with dieresis yet preceded by a single consonant represent 48.3% of tokens for realisations with dieresis and involve proper nouns (*lyon* [li(j)ɔ̃], *Fourvière* [fuʀ.vi.(j)ɛʀ]) or the forms *rien* [ʀi.(j)ɛ̃], *filiale* [fi.li.(j)al], *sixième* [si.zi.(j)ɛm]. Note that descriptions for variable lexicalised forms generally involve the high vocoid /I/ and often in contexts preceded by

[ʀ] as can be found in descriptions of modern French (cf. Léon 1978, Tranel 1987). Variability of lexicalised contexts involving the high vocoids /U/ and /Y/ is predominantly realised with deleted high vocoids. Both /U/ and /Y/ show a single token realised with dieresis (*lui* [ly.(ɥ)i], *Louis* [lu.(w)i]), however, tokens involving the high vocoid /Y/ favour deletion (14.6%) whereas tokens for deleted /U/ represent only 1.1%. Given the high rate of deletion for lexicalised contexts involving /Y/, we observe a dip in the rate for realisations with syneresis (84.2 %) compared with those involving /I/ at 97.5% and /U/ at 98.8%. An elevated rate of high vocoid deletion (14.6%) compared to 0.4% for /I/ and 1.1% for /U/ can be attributed to the frequency of the lexical item *puis* [pɥi] reduced to [pi]: 57 (41%) of tokens, see discussion below. Note that we also find one instance of the word *depuis* [dəpɥi] realised as [dəpi].

The tokens for deleted /Y/ involve 68 (93.2%) tokens for the reduced form: *puis*→ [pi], which can be seen as a lexicalisation of the reduced form. According to Bybee: "Any frequently occurring stretch of speech can be stored in memory and placed into categories with identical and similar units. Categorization occurs at multiple levels. Exemplars of the same word or phrase are mapped onto a single representation" (2003:187). We note also that among tokens for a deleted high vocoid we also find multiple instances for *suis* which show palatalisation with voicing assimilation [ʃɥ̥(i̥)] without reduction in the preceding phrase *je me*. Again according to Bybee, "Tokens of the same construction are similarly mapped onto a representation, and the items in the variable positions of the construction contribute to the formation of categories based on their semantic properties" (2003: 187). In this regard we can confirm that the lexical items *puis* and *suis* show variable degrees of reduction that include deletion at one end of a continuum and at the other end we observe palatalisation [ʃ] and assibilation [p].

These numbers for variability are fairly inconsequential, suggesting at best that /I/, the most frequent high vocoid, is also more susceptible to variability. The number of variable realisations for /I/ is relatively consistent across regions: *Île de France* 13 (1.2%), *Normandie* eight (0.8%), *Rhône Alpes* ten instances (1.2%). Single tokens involving /U/ and /Y/ come from two different speakers in the *Rhône Alpes* region: from Lyon, a male speaker utters the name /*Louis*/ [luwi], and from Grenoble: a female speaker utters the

phrase /*Il lui appris*/ [ilyHa] wherein [i] is lost to vocalic fusion with the following [a] (cf. Bybee 2003, Walker 2001, Ostiguy et al. 1999). Vocalic fusion, commonly described for Quebec French, involves the loss, through fusion or merger, of the first of two vowels in contact at a word boundary.

In sum, we find that, as is most often described for lexicalised contexts (Kaye and Lowenstamn 1984, Tranel 1987, Hall 2006), lexicalised GV sequences privilege syneresis with some variation between a glide and its corresponding high vowel: GV → HV(G)V. Variation in lexicalised contexts occurs most frequently for /I/ with only rare instances involving /U/ and /Y/. The prevalence of variation involving high vocoid /I/ may be attributed to a frequency effect given the prevalence of the front high vocoid (cf. Bybee 2003). Certainly a correspondence between frequency and markedness are exemplified in our in our spontaneous speech sample.

In the following section, we examine the distributional patterns of variants across high vocoids for all derived glide contexts.

6.3.2 Variable realisations of derived glide contexts

In derived glide contexts, recall from section 6.2, we observe considerably fewer tokens (459) for derived /HV+V/ than we do lexicalised contexts /GV/ (2754). In Table 30 below, we show findings for variable realisations for all derived /HV+V/ contexts compared to all lexicalised contexts.

Table 30: Variable realisations of all derived and lexicalised contexts for three variants across all 3213 tokens: DEL = deleted high vocoid, dieresis, syneresis. N= number, (%) calculated by variant for each context

Variant	Derived		Lexicalised	
	N	%	N	%
DEL	50	10.9	88	3.2
dieresis	129	28.1	31	1.1
syneresis	281	61.1	2635	95.7
Grand Total	**459**	**14.3**	**2754**	**85.7**

Variability in the realisation (Del = deleted high vocoid, dieresis, and syneresis) is much more evident in derived contexts as compared with lexicalised contexts. Derived glide contexts also privilege syneresis, but at a rate of 61.1% (compared to 95.7% lexicalised contexts); deleted high vocoids represent 10.9% of tokens compared with 3.2% lexicalised, and 28.1% of tokens are realised with dieresis compared to 1.1% of lexicalised tokens. Increased rates of deleted high vocoids can be attributed to 28 instances of cross-word-boundary /HV+V/ tokens involving the subject pronoun *tu* followed by a vowel initial verb /*tu* + V/ (*tu as*, *tu es*, for example cf. Warnant (1968 section 1.2). These forms: *tu as* → [ta], *tu es* → [te] can be taken to be instances of lexicalised fixed phrases (Bybee 2003). Tokens involving *tu* show 26/34 (76.5%) deleted /Y/. Still, we find seven tokens realised with syneresis and one with dieresis[23]. Our sample shows that realisations in these contexts may vary for a single speaker. Speaker 69asg1, for example, produces both /*tu* + V/ with deleted high vocoid [ta] but also with dieresis [ty.a].[24]

Table 31 below shows realisations of derived contexts for three potential variants across all high vocoids.

Table 31: <u>All derived</u> /HV+V/ contexts across high vocoids for three variants: DEL = deleted high vocoid, dieresis, syneresis. N= number, (%) calculated by variant for each high vocoid

Variant	Vocoid						Total		Distributional hierarchy
	I		U		Y				
	N	%	N	%	N	%	N	%	
DEL	24	7.1	--	--	26	27.1	50	10.9	**Y ≫ I ≫ U**
dieresis	79	23.3	13	52.0	37	38.5	129	28.1	**U≫ Y ≫ I**
syneresis	236	69.6	12	48.0	33	34.4	281	61.1	**I ≫ U ≫ Y**
Total	**339**	**73.7**	**25**	**5.4**	**96**	**20.9**	**459**	**100**	**I ≫ Y ≫ U**

[23] We observe varying degrees of palatalisation in the production of these contexts (*tu es* [tSɥe], for example).

[24] Further study would be necessary to determine conditioning factors.

Table 31 shows the rate at which each potential variant occurs for all derived /HV+V/ contexts. The rate at which each variant occurs across three high vocoids is represented by the distributional hierarchy shown to the right. Derived contexts across high vocoids show the modified hierarchy: /I/ ≫ /Y/ ≫ /U/, noted for the total data above which show that /I/ (73.7%) is most prevalent while /U/ (5.4%) is least prevalent dominated by both /I/ and /Y/ (20.9%) which is second most prevalent. While we might attribute this difference in distributional hierarchy to the inclusion of cross-word-boundary contexts involving the pronoun *tu* (cf. Warnant 1968 section 1.2), when we eliminate these 28 tokens for this cross-word-boundary (WB) /*tu* + V/ context, we find that the distributional hierarchy that privileges /Y/ over /U/: /I/ ≫ /Y/ ≫ /U/ is unchanged.

In the table above and in those to follow, we show distibutional hierarchies for each variant examined in the column at right.

Table 32: All derived /HV+V/ tokens for two environments (excluding 28 /tu +V/): N= number, % calculated for each environment across high vocoids

Evironment	Vocoid						Total		Distributional hierarchy
	I		U		Y				
	N	%	N	%	N	%	N	%	
Across WB	290	84.3	18	64.2	42	67.7	350	81.2	**I ≫ Y ≫ U**
Medial	54	15.7	7	2.8	20	32.3	81	18.8	**Y ≫ ! ≫ U**
Grand Total	**344**	**79.8**	**25**	**5.8**	**62**	**14.4**	**431**	**100**	

Table 36 above shows all derived glide tokens excluding 28 instances of /*tu* + V/ and distinguished by the environment in which they occur: word medial (Medial) and crossword boundary (Across WB). We observe that, even excluding contexts for /*tu* + V/, the distributional hierarchy showing /Y/ dominating /U/ is unchanged.

The following tables (Table 33-36) show data for derived word medial and across-word-boundary data.

Table 33: All derived word medial tokens across three high vocoids for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid						Total		Distributional hierarchy
	/I/		/U/		/Y/				
	N	%	N	%	N	%	N	%	
dieresis	12	22.2	1	12.5	3	15.8	16	19.8	**I ≫ Y ≫ U**
syneresis	42	77.8	7	87.5	16	84.2	65	80.3	**I ≫ Y ≫ U**
Grand total	**54**	**66.7**	**8**	**9.9**	**19**	**23.5**	**81**	**100**	

Table 33 shows findings for all tokens observed in word medial contexts; the word medial environment provides very few tokens accounting for only 17.7% of overall derived data. As observed elsewhere for derived tokens, these tokens exhibit a changed distributional hierarchy across three high vocoids which privileges the front rounded high vocoid over the back rounded. In word medial environment we observe no tokens for a deleted high vocoid and syneresis is strongly privileged by all three high vocoids. Word medial tokens exhibit considerable regularity such that apart from frequency of high vocoids there is little to distinguish between them.

In Table 34 we show our findings for cross-word-boundary tokens.

Table 34: All derived cross-word-boundary tokens across three high vocoids for three variants: DEL = deleted high vocoid, dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid						Total		Distributional hierarchy
	/I/		/U/		/Y/				
	N	%	N	%	N	%	N	%	
DEL	24	8.3	--	--	--	--	24	6.9	**I ONLY**
dieresis	73	25.2	13	72.2	29	69.0	115	32.8	**U ≫ Y ≫ I**
syneresis	193	66.6	6	33.3	13	31.0	212	60.6	**I ≫ U ≫ Y**
Grand Total	**290**	**82.9**	**18**	**5.1**	**42**	**12.0**	**350**	**100**	

Cross-word-boundary tokens shown above represent considerably more tokens (81.2%) than is found for word medial environments (18.8%). We find tokens for deleted high vocoids involving /I/ only (6.9% overall). Interestingly these data show a clear distinction across the three high vocoids. While syneresis is clearly preferred for tokens involving the high vocoid /I/ and for overall data (60.6%), tokens for both rounded high vocoids overwhelmingly privilege realisation with dieresis: /U/ showing 72.2% and /Y/ at 69%. The high frequency of tokens involving /I/ realised with syneresis (193, 66.6%) serves to tip the balance for cross-word-boundary data; despite the rounded high vocoids preferring dieresis overall findings show that syneresis is preferred (60.6%) over dieresis (32.8%) and finally deleted high vocoids.

In a closer examination of cross-word-boundary data, we find 261 tokens (74.9%) involve monosyllabic words bearing a high vocoid followed by a vowel initial word leaving 86 tokens (25.1%). The figure below shows our findings for monosyllabics.

Table 35: Cross-word-boundary tokens for monosyllabics bearing a high vocoid followed by a vowel initial word for three variants: DEL = deleted high vocoid, dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid						Total		Distributional hierarchy
	I		U		Y				
	N	%	N	%	N	%	N	%	
DEL	23	10	--	--	--	--	23	8.8	I ONLY
dieresis	44	19.1	11	64.7	11	78.6	66	25.3	**I ≫ Y ≫ U**
syneresis	163	70.9	6	35.3	3	21.4	172	65.9	**I ≫ U ≫ Y**
Grand Total	**230**	**88.1**	**17**	**6.5**	**14**	**5.4**	**261**	**100**	

Interestingly the distibution across three high vocoids as observed for all other derived contexts shows a change once again. However, the evidence is less than conclusive. Overall these data show yod is most prevalent (88.1%), but the rounded high vocoids are nearly balanced.

High vocoid /I/ represents 88.1% of all cross-word-boundary tokens and privileges syneresis (70.9%), whereas tokens for rounded high vocoids are fewer but privilege dieresis: /U/ represents (6.5%) with 64.7% dieresis and /Y/ represents 5.4% overall tokens with 78.6% dieresis. Among these cross-word-boundary tokens, 131 tokens (56.9%) are for the context /*qui* + V/ that appear to follow the same or similar pattern observed in word medial contexts involving the high vocoid /I/. Cross-word-boundary tokens for /*qui* + V/ contexts privilege glide formation (74%), while also showing some variable realisations: dieresis (9.9%) as well as 15.3% deleted high vocoids. Of the 23 tokens for high vocoid deletion, 20 (15.3%) involve /qui + V/.

Monosyllabic forms containing a high vocoid followed by a vowel initial word prefer realisations with syneresis (65.9%) compared to 25.3% realised with dieresis. However, rates for syneresis versus dieresis are not equal across high vocoids. Both rounded high vocoids in monosyllabics show a preference for dieresis: /U/ realised with (64.7%) and /Y/ (78.6%) dieresis with all remaining tokens realised with syneresis (35.3% and 21.4% repectively). These data appear to follow the margin fitness markedness hierarchy proposed by Bullock (2002, cf. section 4.7: *Mɥ ≫ *Mw ≫ *Mj.

While our data are not coded for position in a phrase, we can identify several contexts where the particular form is necessarily stressed or unstressed. We find, for example, the homophones *ou* and *où* involved in cross-word-boundary contexts '*où est*', and '*ou autres*', for example. Despite their distinct grammatical function in the phrase they behave similarly with respect to realisations observed. Tokens for the coordinating conjunction /*ou* + V/ represent 11.9% of overall tokens for /U/ and show 40% syneresis and 60% dieresis, while tokens for the interrogative *où* represent 23.8% of all tokens for the high vocoid /U/ and also show dieresis is most prevalent at 66.7% with 33.3% realised with syneresis. We find these two contexts behave quite similarly in privileging

dieresis in cross-word-boundary contexts. While these cross-word-boundary contexts are not often acknowledged in the descriptive literature, their frequency and regularity of realisation requires further study. As noted in our literature review the contexts are noted by some authors. According to Lyche (1979), who notes that “the interrogative pronoun où ‘where’ can also be submitted to the gliding rule: *Où as-tu vu ça* [waty], *Où est-ce que c’est* [wɛskœse]; If *où* bears even weak stress gliding is no longer possible” (1979: 328 cf. also Passy 1913 section 2.1.1).

We turn now to the remainder of the cross-word-boundary tokens observed in our study. These tokens necessarily involve a high vocoid final polysyllabic form (*revenue à*, for example) followed by a vowel inital word (preposition or pronoun).

Table 36: Cross-word-boundary tokens for polysyllabics bearing a high vocoid followed by a vowel initial word for three variants: DEL = deleted high vocoid, dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	1	17.5	--	--	--	--	1	1.2
dieresis	29	50.9	2	100	17	63.0	48	55.8
syneresis	27	47.3	--	--	10	37.0	37	43.0
Grand Total	**57**	**66.3**	**2**	**23.3**	**27**	**31.4**	**86**	**100**

As shown in Table 36 above, high vocoid final polysyllabic forms followed by a vowel initial word (*venu avec*, for example) are much fewer than are the monosyllabics bearing a high vowel and followed by a vowel initial. Of the 86 tokens observed, those involving /I/ represent 66.3% followed by /Y/ at 31.4% and then /U/ with only two tokens representing 23.3%. Although these data show the familliar changed distributional hierarchy wherein /Y/ dominates /U/, this is the first instance where syneresis is not privileged (usually overwhelmingly) as a realisation implying that GF is not dispreferred (even for /I/ in this context). Given the distinctly different pattern across variants compared with the monosyllabic data, it appears that polysyllabic forms bearing high vocoids do not paricipate equally in glide formation. We will come back to this finding in Chapter 7.

We turn now to an examination of variable realisations in derived /HV+V/ environments and the possible correlation between variability and preceding phonological environments. In the following section we examine variability and the preceding phonological environment to discern any correlation between realisation of an /HV+V/ context and the phonological environment preceding. Given the prominent stability in lexicalised contexts, we limit our discussion to derived glide data.

6.3.3 Derived glide contexts: variability of /HV+V/ realisations and phonological environments preceding

Recall that our judgement task results (reproduced in Table 37 below) show varying results for the derived stimuli 6 *(fou à) lier* (syneresis is preferred at a rate of 47%) compared with 28 *scier* (dieresis is preferred at 65.3%).

Table 37: Judgement task results, N= number, % calculated by total N for each stimulus

Stimulus	categorical syneresis		categorical dieresis	
	N	%	N	%
6 (*fou à*) *lier*	49	47.1	14	13.5
28 *scier*	12	12.2	64	65.3
30 *mouette*	58	56.3	18	17.5
51 *influence*	1	1.0	51	49.5
63 *miette*	48	71.6	1	1.5
72 (*vous*) *prendriez*	--	--	53	51.0
79 *muette*	4	4.0	50	49.5
81 *trouer*	1	1.0	61	59.8

Given the degree of variable realisations observed in derived contexts (cf. section 6.3.2), we examine preceding phonological environments for all derived contexts to discern if any correlation can be found between variability and a preceding environment. In the analysis that follows, we examine all potential realisations of preceding phonological environment (OL simplification for example) as well as /HV+V/ realisations (high vocoid deletion, dieresis and syneresis) for all derived contexts.

The examples below illustrate the various phonological environments (word boundary, simple onsets, complex onsets) that can precede the derived glide contexts observed beginning with least complex environment (#) immediately preceding the derived context to the most complex environment (OL) with examples taken from our sample:

(122) Observed phonological environments preceding /HV+V/ contexts:

a. #: word boundary – /U/ *où est* /u+#e/, *où on* /u+#ɔ̃/; /Y/ eu a(ssez) /y+#a/;
b. (G) Glide [ɥi]– /I/ *oui on* /wi+ɔ̃/, *oui elle* /wi+ɛl/;
c. (L) Liquid [l;R] – /I/ *(a)lliés* /li+e/, (*se ma*)*riant* /ri+ã/, /Y/*(évo)luer* /ly+e/;
d. (N) [n, m, ɲ] – /I/ (*fi)ni avec* /ni+a/, (*compa)gnie a(erienne)* /ɲi+ a/, / Y/ (*conti)nué* /ny+e/,
e. (O) Obstruent
   i) fricative: /I/ (*négo)ciais* /si+ɛ/, (*chan)gier* /ʒi+e/, /Y/ *vu é*(*normément)* /vy+e/, /U/ *jouer* /ʒu+e/;
   ii) stop: /I/ (*expé)dié* /di+e/, /U/ (*se)couer* /ku+e/, (*sur)tout é(viter)* /tu+e/, /Y/ *tuaient* /tu+ɛ/;
f. (sO) sObstruent – /I/ *c(e) qui é(tait)* /s(ə)+#ki+#e/;
g. (OH) Obstruent, front rounded glide [ɥ] – /I/ (je) *suis a(rrivé)* /sɥi+a/, (*la*) *nuit on* /nɥi+ɔ̃/, *puis a(près)* /p(ɥ)i+a/, *(pro)duit en*(*fin*) /dɥi+ɛ̃/;
h. (OL) Obstruent + Liquid –
   i) Ol - /I/ *(ta)blier* /bl+je/, /Y/ *influer* /fly+e/;
   ii) OR - /I/ *(ou)vrier* [vRi.je], /Y/ *(se serait) cru en* /kRy+#ã/;

Examples shown here are taken from our own data; these do not represent an exhaustive list of possible environments. However, any gaps in example data in (122) are representative of distributional gaps found in our sample. Note, for example, that in 122a., we find no instances of /I/ immediately preceded by a word boundary (#) whereas /I/ is well represented in all other environments. The examples in (122) represent tokens for /HV+V/ (including high vocoid distributions) observed in our sample.

Table 38 below shows totals of all derived tokens for each preceding simple and complex environment observed across three high vocoids.

Table 38: Preceding environments for derived /HV+V/: all simple (#, G, L, N, O) and complex (OH, OL, sO) environments across three high vocoids

Preceding environments	Vocoid			Total
	I	U	Y	
Simple				
# - word boundary		15	6	21
G - glide	8	--	--	8
L – liquid [l, R]	38	2	5	45
N – nasal [n, m, ɲ]	11	--	29	40
O – obstruent (fricatives and stops)	187	7	54	248
Total Simple	244	24	94	362
Complex				
OH – Obstruent[ɥi] + V	73	--	--	73
OL – Ol, OR	14	--	2	16
sO – resyllabation	8	--	--	8
	95	--	2	97
Grand Total	**339**	**24**	**96**	**459**

As mentioned earlier, we have included both word medial as well as contexts derived across word boundaries. Table 38 shows phonological environments preceding all derived tokens. 363 (79.1%) tokens are preceded by a simple phonological environment including a word boundary or a simple consonant. 96 (20.9%) derived tokens are preceded by complex phonological environments (OL) but also including the cross-word-boundary consonant + diphthong [Cɥi] followed by a vowel initial word and /s+k/ found after resyllabation triggered by schwa deletion in the phrases *(tout) ce qui/parce que*.

In the following tables, we examine findings for derived tokens preceded by simple and complex environments (onsets): #, G, L, N, s+O, Oɥ, OL.

Our analysis continues with an examination of /HV+V/ tokens preceded word medially by simple onsets: CHV.V (*lier* [l], *jouer* [ʒ], *tuer* [t], *nier* [n], for example) and tokens of cross-word-boundary glide formation preceded by a simple onset as in (*vé)<u>cu a</u>(vec)* /ky#a/ before turning our attention to tokens preceded by complex onsets (122a-e above).

Among simple phonological environments, we include /HV+V/ contexts preceded by a word boundary as in *où est* /u.e/ and contexts preceded by a single consonant (glide, obstruent (fricative or occlusive), liquid or nasal).

Table 39: Simple phonological environments: derived contexts /HV+V/ preceded by word boundary (#) for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid				Total	
	U		Y			
	N	%	N	%	N	%
dieresis	10	66.7	6	100	16	76.2
syneresis	5	33.3	--	--	5	23.8
Grand Total	**15**	**71.4**	**6**	**28.6**	**21**	**100**

Table 39 shows all tokens for /HV+V/ contexts preceded by a word boundary only: *où est* [uwe], *ou en* [wɛ̃], *(on a) eu a(cheté)* [yɥa], *eu a(ssez)*[yɥa], for example. The absence of tokens for the high vocoid /I/ reflects the absence of words comparable to *où* [u] that consist of the phoneme [i] only. Dieresis is most prevalent in this environment (76.2%). Tokens for /U/ show variation across syneresis and dieresis, while tokens for /Y/ are realised exclusively with dieresis. Five tokens (33.3%) for /U/ are realised with syneresis. These tokens involve the high vocoid /U/ found in essentially identical phonological environments differing only by the following vowel quality: <u>*ou en*</u> [wɛ̃], *où on* [wɔ̃], *où i([ls] faisaisent)* [wi].

Figure 26 below shows a token for *où ils* in a derived cross-word-boundary environment preceded by a word boundary only.

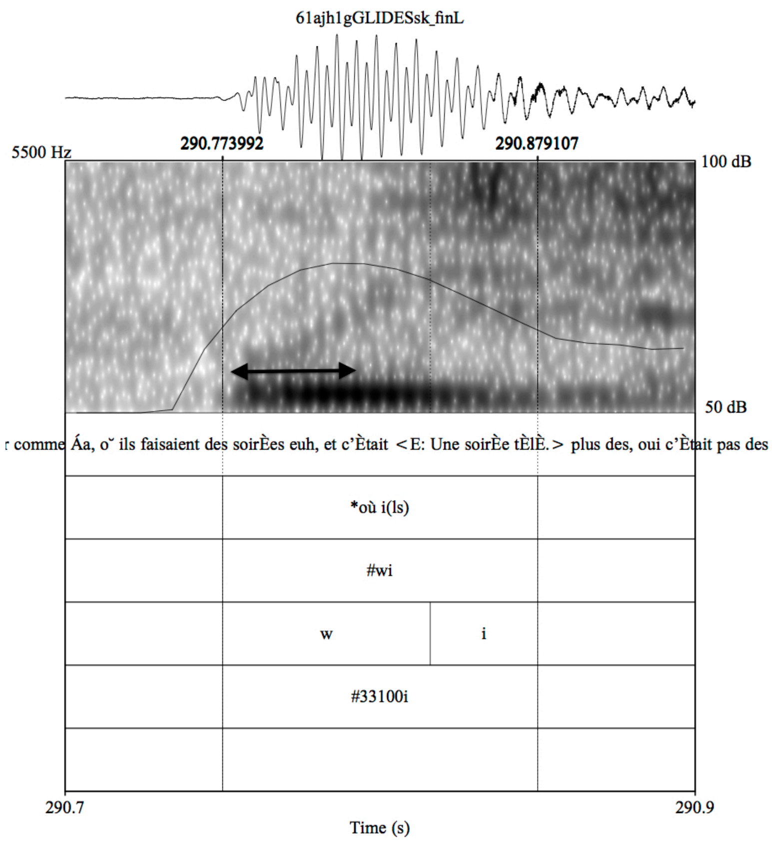


Figure 26: 61 ajh1 '*où i(ls)*' realised with syneresis [wi]: derived across-word-boundary preceded by only a word boundary (#)

This token is realised with syneresis evidenced by a single peak in intensity corroborated by the oscillogram where we observe formant transition immediately upon voicing onset indicating the presence of a glide: *où* [w] followed by [i].

Among tokens preceded by a word boundary only, we observe 16 instances of dieresis involving both /U/ and /Y/, 10 of which occur in identical contexts observed to produce syneresis. The preceding phonological environment is not, therefore, the determining factor in the realisation of this context.

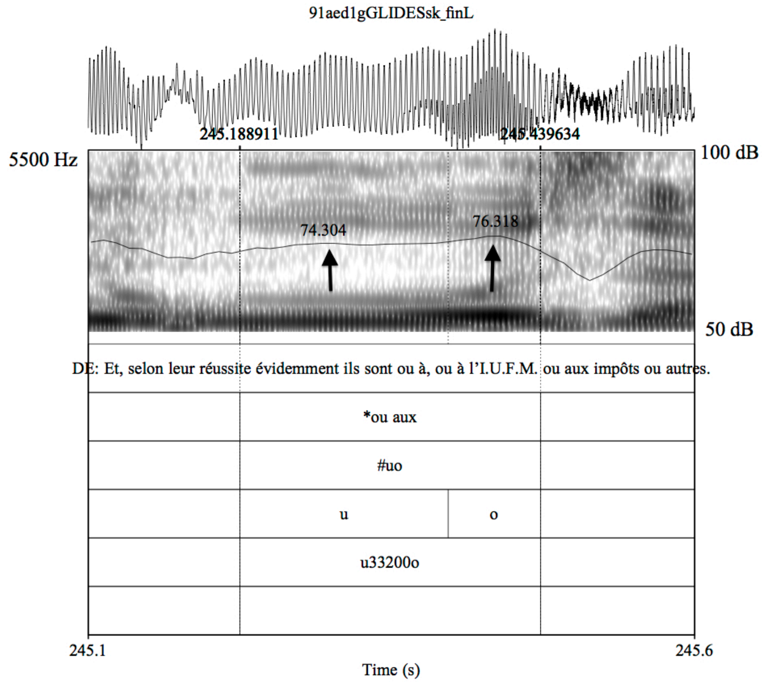


Figure 27: 91aed1 '*ou aux*' realised with dieresis [uo]: derived across-word-boundary preceded by only a word boundary (#)

Figure 27 above shows a token for the phrase *ou aux* /u+o/ judged to be realised with dieresis [uo]. Peaks on the intensity contour are very subtle, however, the intensity listing for this contour gives two peaks, the first attaining 74.304 dB followed by a second at 76.318 dB.

The next environment examined here is that of derived /HV+V/ preceded by a glide (G) only. This is a very particular cross-word-boundary derived /HV+V/ environment which necessarily involves a word with high vowel in final position with G immediately preceding (*oui* [wi], for example) followed by a vowel initial word as in *oui au*, and *oui on*. This subset of derived data necessarily involves the diphthong [ɥi] wherein the high vocoid /I/ is found in word-final position: *nuit* /nɥi+V/ or *pro<u>duit</u>* /dɥi+V/, for example. Tokens number only eight and they are realised with some variability: two tokens (25%) realised with syneresis (*oui on* and *oui elle*), six (75%) are realised with dieresis[25].

---

[25] We observe one realisation with syneresis each from *Île de France* and from *Normandie*.

In Figure 28 below, we show the token *oui on* realised with dieresis.

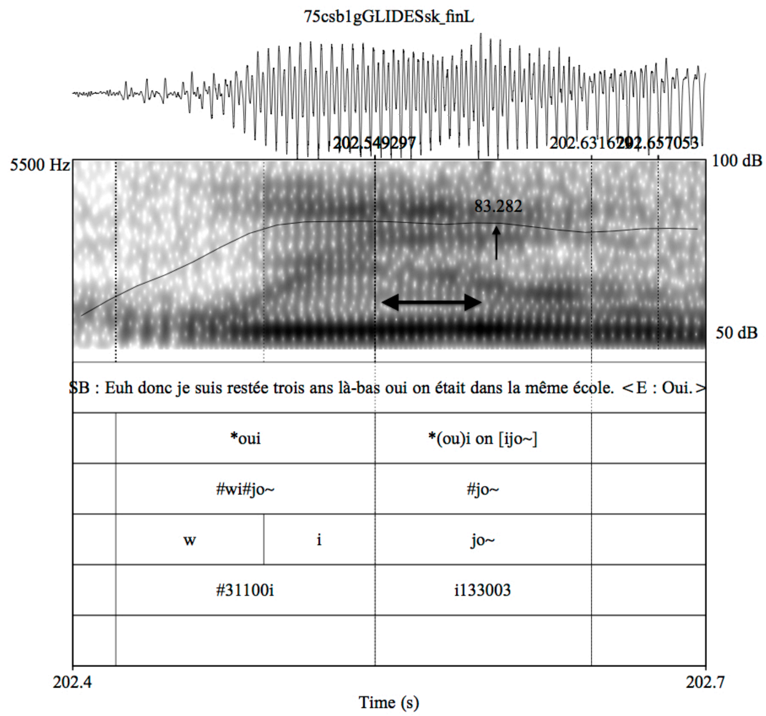


Figure 28: 75csb1 *'oui on'* realised with dieresis [ijɔ̃]: derived across-word-boundary preceded by only a word boundary (G)

The spectrogram shows the formant transition of the utterance initial [w] which is followed by high vowel [i] production: [wi]. We observe a second period of formant transition following [i] wherein the formant strength shows weakening as it transtitions towards the nasal vowel [ɔ̃]. Formants strengthen at the end of this transition for production of [ɔ̃] which corresponds to another peak on the intensity contour.

The preceding data give evidence for variability which is not necessarily conditioned by a phonological environment. This suggests that other factors perhaps linguistic (proximity to stress, for example), perhaps social, play a role. However the small number of tokens observed for the number of participants does not allow us to provide a more rigorous analysis.

We turn now to derived contexts preceded by a simple onset containing a liquid [l, R] only. These tokens are found word medially: *varié, évoluer*, or cross-word-boundary *mari*

avec, for example. Table 40 below shows totals of all instances of /HV+V/ preceded by a liquid [l, R].

Table 40: Derived contexts /HV+V/ preceded by a simple phonological environment – Liquid /l, R/ for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
dieresis	9	23.7	1	50.0	3	60.0	13	28.9
syneresis	29	76.3	1	50.0	2	40.0	32	71.1
Grand Total	**38**	**84.4**	**2**	**4.4**	**5**	**11.1**	**45**	**100**

Table 40 shows 45 tokens for derived /HV+V/ preceded by a simple onset containing a liquid /l, R/ are shown in the table above. Tokens preceded by liquid consonants show syneresis is most prevalent across all three high vocoids overall (71.1%). Closer examination of results for each vocoid brings to light other patterns. Tokens involving /I/ privilege syneresis (76.3%), with only nine tokens (23.7%) realised with dieresis. We find fewer tokens for the rounded high vocoids: five tokens only for /Y/ and two tokens for /U/. The front rounded /Y/ shows 60% are realised with dieresis and 40% of tokens are realised with syneresis, while two tokens involving /U/ are split with one token each for syneresis and dieresis. Overall, syneresis is observed for 71.1% of tokens and dieresis represents 28.9% of tokens in this environment.

In the figure 29 below we show a token for the word *loué* realised with dieresis.

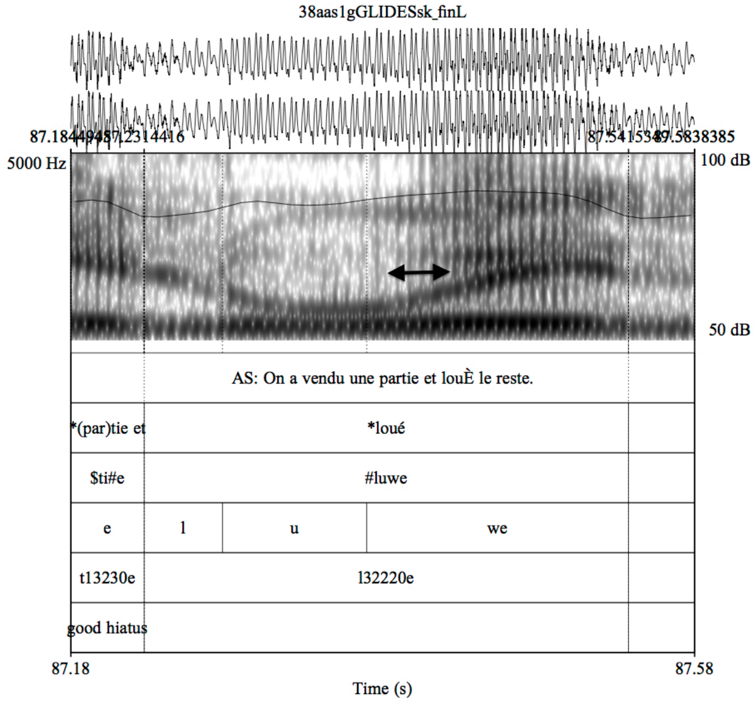


Figure 29: 38aas1 *loué* realised with dieresis [luwe]: derived word medially preceded by (L)

On the spectrogram shown in Figure 29 above we observe a period of vowel production following [l] that corresponds to the high vocoid [u]. Immediately following, we observe strong formant transition which we take to indicate presence of glide before the following mid front vowel [e]. These cues in combination with two peaks on the intensity contour (87.118, 89.895 dB) indicate that this token is realised with dieresis.

In the tables that follow, we examine findings for tokens with preceding liquids separately. Our findings show fewer tokens preceded by /l/: twelve (3.75%), than preceded by /R/: 33 tokens (73.3%). However, the tokens for /l/ show more variability.

Table 41 below shows all /HV+V/ tokens preceded by /l/.

Table 41: Derived contexts /HV+V/ preceded by simple phonological environment – Liquid /l/[26] for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
dieresis	2	40.0	1	50.0	3	60.0	6	50.0
syneresis	3	60.0	1	50.0	2	40.0	6	50.0
Grand Total	**5**	**41.7**	**2**	**16.7**	**5**	**41.7**	**12**	**100**

While we find only twelve tokens preceded by /l/, realisations in syneresis and dieresis observed are split. We find two tokens only for the high vocoid /U/ (one each for realisation with syneresis and realisation with dieresis); the remainder of tokens are evenly distributed across /I/ and /Y/. Syneresis and dieresis are each realised in 50% of tokens with minor variability observed for the particular high vocoid; /I/ privileges syneresis while /Y/ privileges dieresis and /U/ shows parity among variants.

Table 42 below shows all derived tokens preceded by /R/.

Table 42: Derived contexts /HV+V/ preceded by simple phonological environment – Liquid /R/[27] for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	/R/	
	N	%
dieresis	7	21.2
syneresis	26	78.8
Grand Total	**33**	100

[26] Among tokens preceded by /l/, we observe three instances (25%) of /HV+V/ contexts that cross-word-boundary and nine (75%) tokens for word medial. Cross-word-boundary tokens (two /I/, one /U/) are realised with maximal variability (one token each for each potential variant: syneresis, dieresis hiatus and dieresis epenthesis). Word medial tokens: 3 /I/, 1 /U/, and 5 /Y/, are divided between syneresis (55.5%) and dieresis with epenthesis (44.4%).

[27] Tokens preceded by /R/ include 27 (81.8%) instances found in word medial position and six (18.2%) in cross-word-boundary contexts. Cross-word-boundary tokens are realised with dieresis alone while word medial contexts are realised almost exclusively with syneresis.

/HV+V/ contexts preceded by /R/ show less diversity than we observed in the table preceding for tokens preceded by /l/. Tokens preceded by /R/ involve exclusively the high vocoid /I/; however, word medial tokens preceded by /R/ strongly privilege realisations with syneresis.[28] On the other hand, dieresis is categorical when the context is cross-word-boundary. This supports the descriptions from Tranel (1987) and Léon (1978).

In the following table we examine findings for /HV+V/ contexts preceded by the nasals /m/ and /n/.

Table 43: Derived /HV+V/ contexts preceded by simple phonological environment – Nasals /m/ and /n/ for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	vocoid					
	I		Y		Total	
	N	%	N	%	N	%
dieresis	6	51.6	13	44.8	19	47.5
syneresis	5	45.5	16	55.2	21	52.5
Grand Total	**11**	**27.5**	**29**	**72.5**	**40**	**100**

Tokens for /HV+V/ contexts preceded by a nasal consonant involve only two high vocoids favouring the front rounded: /I/ (27.5%) and /Y/ (72.5%); we find no instances of /U/. Syneresis is moderately more prevalent (52.5%) than dieresis (47.5%). Distribution across two high vocoids shows more front rounded /Y/ (72.5% overall) and /I/ at 27.5%. Numbers show relative parity across realisations.

Table 44, shows all tokens for a preceding nasal /m/.

[28] A single token of the word *varié* is produced with dieresis.

Table 44: All derived /HV+V/ contexts preceded by simple phonological environments – Nasal /m/ for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid	
	I	
/m/ cross #	N	%
dieresis	3	75.0
syneresis	1	25.0
Total /m/	4	100

We find only four instances of /HV+V/ contexts preceded by /m/. All tokens for /m/ are cross-word-boundary involving a single high vocoid /I/, *re*mis au realised with syneresis, for example. Cross-word-boundary tokens preceded by /m/ appear to privilege dieresis (75%), however these numbers are too few to make any generalisations.

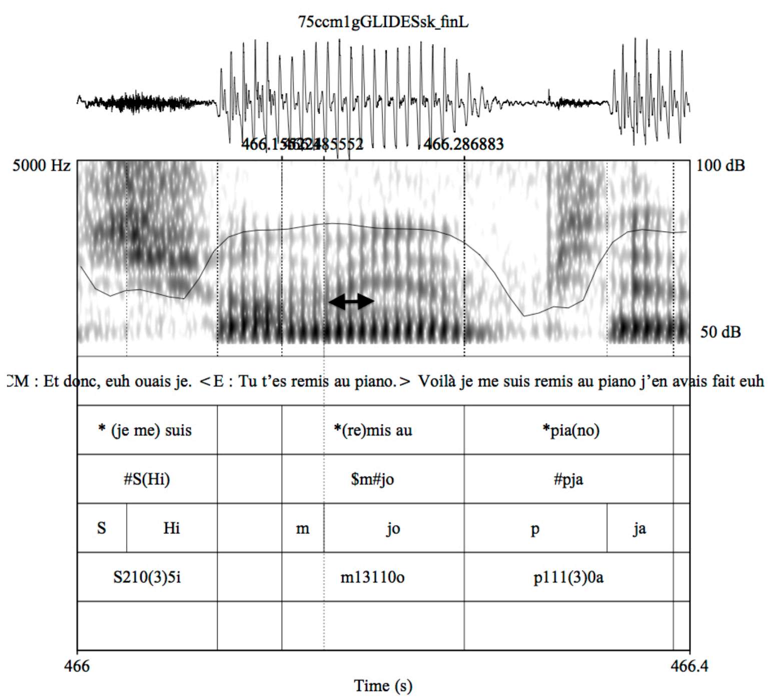


Figure 30: 75ccm1 (*re*)*mis au* realised with syneresis: derived across-word-boundary preceded by [m]

The figure above shows three tokens. We are concerned only with the token for the phrase (re)mis *au.* This token preceded by [m] is realised with syneresis as shown by the single peak of intensity immediately following the production of [m] which corresponds to the high vocoid /I/. Note that formant transition immediately follows [m] towards the rounded mid-high back rounded vowel [o]. The third token here shows *pia*(no) where the [p] shows assibilation which we take to indicate the presence of a devoiced front glide [p'j̥]. The voiced period that follows shows no signs of formant transition which we take to indicate the absence of yod.

Let us look now at our findings for tokens preceded by /n/ (*contin<u>uer</u>*, for example). Table 45 below shows findings for all preceding /n/.

Table 45: All derived /HV+V/ contexts (word medial and cross-word-boundary**) preceded by simple phonological environments – Nasal /n/ for two variants:** dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

	Vocoid					
Variant	I		Y		Total	
/n/ (medial)	N	%	N	%	N	%
dieresis	--	--	3	27.3	3	27.3
syneresis	--	--	8	72.7	8	72.7
Total medial	**--**	**--**	**11**	**100**	**11**	**100**
/n/ cross #	N	%	N	%	N	%
dieresis	3	42.9	10	66.7	13	59.1
syneresis	4	57.1	5	33.3	9	40.9
Total cross #	**7**	**31.8**	**15**	**68.2**	**22**	**100**

We find 33 tokens preceded by /n/; eleven of these are found in word medial contexts involving the high vocoid /Y/ only (*diminuer*, *continuer*, for example). These tokens are too few to make strong generalisations, however they appear to privilege syneresis: eight tokens (72.7%). The remaining 22 tokens cross a word boundary (*four<u>ni à</u>*, for example) and involve the high vocoids /I/ and /Y/. Nine (40.9%) cross-word-boundary tokens are realised with syneresis: four involve /I/, 5 involve /Y/. 13 (59.1%) are realised with dieresis: three /I/, 10 /Y/. These data show contradictory behaviour that may suggest that preference for syneresis versus dieresis is conditioned by the high vocoid involved as well as the environment: medial versus cross-word-boundary. Dieresis is generally

preferred however the gaps in our data make a meaningful analysis difficult. Figure 31 below shows the cross-word-boundary token *venue à.*

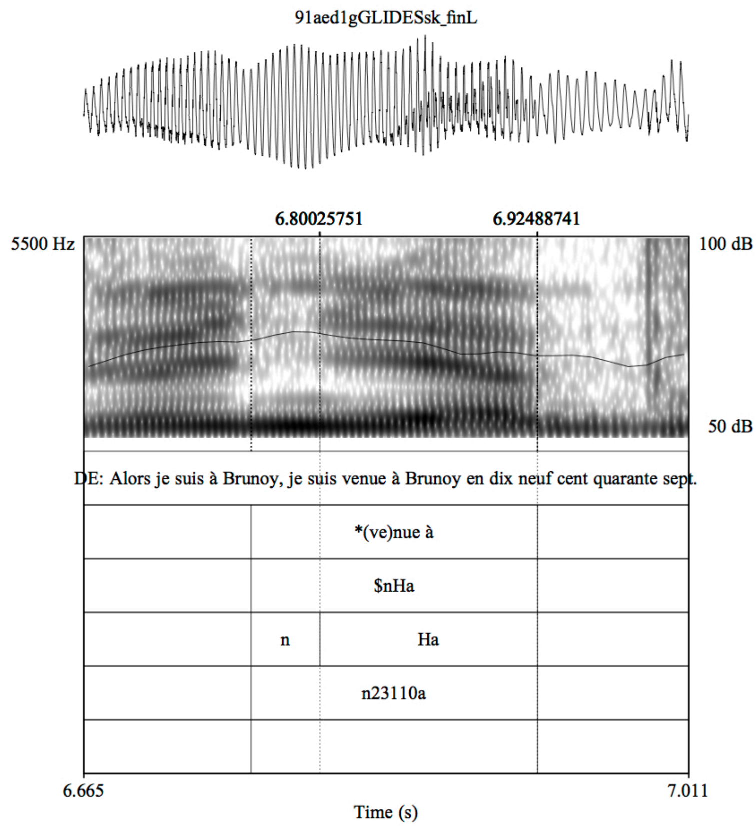


Figure 31: 91aed1 (*ve*)*nue à* realised with syneresis: derived across-word-boundary preceded by [n]

We turn now to an analysis of realisations for tokens of /HV+V/ preceded by fricatives. Table 46 below shows findings for all tokens preceded by the fricatives /f, v, s, z, ʃ, ʒ/.

Table 46: All derived /HV+V/ contexts (word medial and cross-word-boundary**) preceded by a simple phonological environment – Fricatives /f, v, s, z, ʃ, ʒ/ for two variants:** dieresis, syneresis. N= number, (%) calculated for each variant by vocoid.

Variant	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
dieresis	14	30.4	--	--	4	80.0	17	30.9
syneresis	32	69.6	5	100	1	20.0	38	69.1
Grand Total	**46**	**82.1**	**5**	**8.9**	**5**	**8.9**	**56**	**100**

The data shown above include 39 (86.7%) tokens observed in a cross-word-boundary environment preceded by the fricatives /s, v, ʒ/ and 16 (29.1%) tokens in a word medial environment preceded by the fricatives /f, s, ʃ, ʒ/. Overall distribution across high vocoids shows that /I/ is most prevalent: /I/ 81.8%, /U/ 9.1% and /Y/ 9.1%. We note that the high vocoid /U/ shows categorical syneresis in tokens for word medial contexts preceded by /ʒ/: *jouait*, *jouer,* for example, while the high vocoid /Y/ shows more dieresis (80%) when observed in cross-word-boundary contexts preceded exclusively by /v/[29]. Tokens for /I/ include 75.5% cross-word-boundary /HV+V/ contexts and 24.5% word medial contexts. Word medial contexts involving /I/ are realised with categorical syneresis which is in contrast to the judgment task results for the stimulus *28 scier* that show considerable variability (cf. section 5.6), while the cross-word-boundary tokens show 71.1% syneresis compared to 28.9% dieresis.

We turn now to an examination of findings for preceding stops. In Table 47, we examine the results observed for /HV+V/ preceded by the stops: étudiais [djɛ], for example).

[29] All tokens involve the past participle *vu* (*vu arriver*, for example).

Table 47: Derived /HV+V/ contexts preceded by simple phonological environments – Occlusives /p, b, t, d, k, g/ for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid for variant

Variant	Vocoid						Total	
	I		U		Y			
dieresis	18	15.1	1	50.0	6	40.0	25	18.4
syneresis	101	84.9	1	50.0	9	60.0	111	81.6
Grand Total	**119**	**87.5**	**2**	**1.5**	**15**	**11.0**	**136**	**100**

Table 47 shows all derived contexts preceded by occlusives: 136 (29.6%) overall; these represent the largest number of derived tokens by environment. From these data, we have eliminated 56 tokens for high vocoid deletion in cross-word-boundary environment only; these include 34 tokens for /Y/ (*tu as* [ta], for example), and 22 for /I/ (*qui est* [ke], for example). In neither of these environments is deletion of the high vocoid categorical for any one speaker. Tokens for the high vocoid /Y/, wherein *tu as* is realised as *t'as* [ta], represents 50.1% of all derived /Y/ tokens, while tokens for deleted /I/ represent considerably fewer: 6% of overall tokens for /I/[30]. Distribution across high vocoids is otherwise skewed by the frequency of these two contexts; we therefore exclude these from our analysis.

Syneresis is most prevalent across all tokens (79.9%) for dieresis represent 20.2%; high vocoid /I/ shows the highest frequency (82.6%) followed by /Y/ (16%), and finally the high back vocoid /U/ shows only two tokens split between syneresis and dieresis. Across all tokens preceded by an occlusive, syneresis is most prevalent at 79.9%; dieresis occurs in 20.2% of tokens.

These data include seven tokens for word medial environments preceded by [t]: *tuer, habitué,* [k]: *secouer,* and [d]: *expédié, étudiais,* as well as 137 tokens for cross-word-boundary environments preceded by [p, t, k, b, d]; we find no tokens preceded by /g/. Tokens for the high vocoid /I/ include cross-word-boundary contexts such as: *parti au, reparti au*, for example, are realised predominantly with syneresis, representing 84.9%

[30] We find a single token for deleted high vocoid /I/ for the context *parti avec*.

and dieresis (14.1%). Our findings show token distribution across high vocoids as follows: /I/ (82.6%) ≫ /Y/ (16%) ≫ /U/ (1.4%).

In Figure 32 below, we examine a particular token for the cross-word-boundary environment *parti à*.

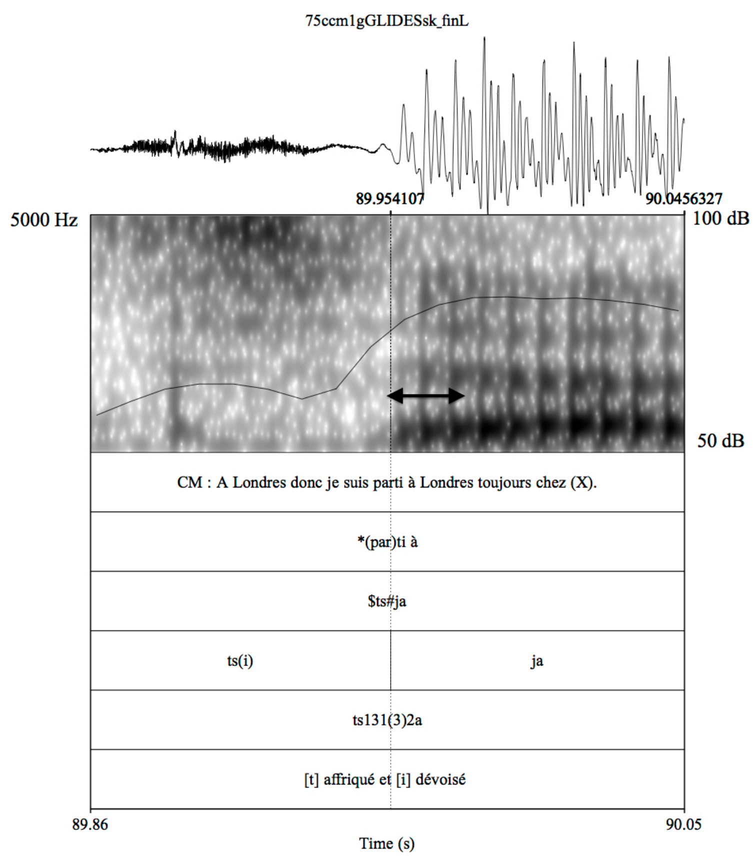


Figure 32: 75ccm1 *parti à* realised with syneresis. Spectrogram with intensity contour, oscillogram and tiers: transcription, token, SAMPA, segments, coding, observations

This token presents a number of interesting aspects for our study of spontaneous speech data. In the final syllable of *par<u>ti</u>*, we observe in the spectrogram increased energy (approximately 4400 Hz) after the burst of [t] which indicates assibilation of the occlusive [ts] followed by a voiceless [i̥] evidenced by subtle formant transition at voicing onset and quite an audible yod.

Unfortunately, the number of tokens observed is too few for a meaningful analysis of the significance of preceding phonological environment in outcomes observed. Adding to the difficulty of low numbers, these tokens may involve one of three high vocoids, which, in turn, are not equally represented across these data. In Table 48 below, data are presented for each preceding occlusive with numbers of tokens only.

Table 48: Cross-word-boundary derived /HV+V/ contexts preceded by simple phonological environments – /p, b, t, d, k/ for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid for variant

Realisation	Vocoid			Total
	I	U	Y	
/b/				
dieresis	--	--	1	1
Total /b/	--	--	**1**	**1**
/d/				
dieresis	2	--	3	5
syneresis	3	--	1	4
Total /d/	**5**	--	**4**	**9**
/p/				
syneresis	1	--	1	2
Total /p/	**1**	**--**	**1**	**2**
/t/				
dieresis	3	1	--	4
syneresis	6	--	--	6
Total /t/	**9**	**1**	**--**	**10**
/k/				
dieresis	13	--	2	15
syneresis	89	--	3	92
Total /k/	**102**	**--**	**5**	**107**
Total medial	**117**	**1**	**11**	**129**

In Table 48, we have isolated each preceding consonant for all word medial tokens preceded by occlusives. With only 25 tokens the disparity among them prevents us from drawing conclusions. We find no instances of /p/ or /g/ in the preceding environment for word medial /HV+V/ tokens; nor do we find instances where all three high vocoids occur with the same preceding occlusive. Overall, we find 8 (32%) realisations with dieresis while we find 17 (68%) realisations with syneresis. Only tokens preceded by /d/ and /k/

involve more than a single high vocoid or a single variant. Tokens preceded by /d/ show a split between realisations with syneresis and dieresis overall; however, when we examine findings for each high vocoid separately /I/ shows more syneresis while /u/ shows more dieresis. When we examine preceding environments for a possible effect for voicing we find that preceding voiceless occlusives show a preference for syneresis: ten of 13 tokens (76.9%) are realised with syneresis whereas voiced environments are split between syneresis and dieresis (54.5% syneresis, 45.5% dieresis). Among the tokens with preceding voiced occlusive, there is a split; the high vocoid /Y/ shows 4:1 dieresis and the high vocoid /I/ shows 6:2 for syneresis. While the number of tokens observed here prevents us from making any conclusions, the effect of preceding environment and the high vocoid involved merits further study.

We turn now to an examination of all /HV+V/ preceded by *complex* environments. These include: consonant+glide (/Cɥ/: *suis obligé* /sɥi#a/, *lui on* /pɥi#ɔ̃/), s+k (/s#k/:*(tout) ce qui était* /s#ki#e/, obstruent+liquid (/OL/: *influer*, for example).

In Table 49 below, we show totals of all tokens of /HV+V/ contexts preceded by a complex phonological environment.

Table 49: Derived /HV+V/ contexts, preceded by complex phonological environments – /Cɥ/, /sk/, /Ol/, and /OR/ for three variants: DEL= deletion, dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid				Total	
	I		Y			
	N	%	N	%	N	%
DEL	2	2.1	--	--	2	2.1
dieresis	31	32.6	1	50.0	32	33.0
syneresis	62	65.3	1	50.0	63	64.9
Grand Total	**95**	**97.9**	**2**	**2.1**	**97**	**100**

These data for the four preceding complex environments (/Cɥ/, /sk/, /Ol/, and /OR/), show no instances of the high vocoid /U/ and only two tokens for /Y/ (one each for word medial and cross-word-boundary). The high vocoid /I/ is most prevalent occurring in 97.8% of derived contexts preceded by complex phonological environments. Realisations

with syneresis are the majority (70.7%) while we observe also 27.1% dieresis and 2.2% deletion.

Among tokens for complex preceding environment (Table 49 above), 79.3% involve a Consonant + Glide sequence (*suis obligé* [sɥ(i)#ɔ], *lui appris* [lɥ(i)#a, for example) which also account for two tokens for a deleted high vocoid. These cross-word-boundary /HV+V/ contexts have an added level of complexity since the high vocoid /I/ and the glide [ɥ] that precedes it comprise the nuclear diphthong [ɥi]. Nonetheless, we observe 73 contexts with varied realisation in these contexts.

Table 50 below, shows all realisations observed for /HV+V/ tokens preceded by CG. The nature of this environment implies that tokens are necessarily cross-word-boundary

Table 50: Derived contexts /HV+V/, preceded by complex phonological environment – CG /Cɥ/ for three variants: DEL= deletion, dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid		Total	
	I			
CG	N	%	N	%
DEL	2	2.7	2	2.7
dieresis	19	26.0	19	26.0
syneresis	52	71.2	52	71.2
Total CG	**73**	100	**73**	100

This is a very specific environment where the high vocoid /I/, part of the nuclear diphthong [ɥi], is found in word-final position, thus the exclusion of all other high vocoids from these data. An /HV+V/ context arises when the diphthong [ɥi] in final position of a word is immediately followed by a vowel initial word: /Cɥi+V/. The high vowel [i] may undergo cross-word-boundary glide formation. The tokens observed here include five different consonants preceding a lexicalised diphthong [ɥi]: *puis a(près)*, *suis a(ller)*, *lui a(ssurer)*, *nuit aux* and *produit enfin*.

While cross-word-boundary /HV+V/ environments are not typically described among contexts for glide formation processes, we find considerable glide formation in this environment. The environment involving a lexicalised diphthong [ɥi] as in *lui*, *suis*, *puis*,

for example, presents the added complexity of potential for two successive glides /CGGV/. This outcome requires a process of resyllabification across a word boundary that permits at least an apparent GG sequence: [C[GHV]#V → [C(G)GV], <u>puis a</u>*(près)* [p(ɥ)ja.pʀe], for example. It is perhaps, as a product of the inherent complexity of this outcome, that a large number of tokens for this environment exhibit coarticulation effects including voicing assimilation (devoicing due to proximity to a voiceless segment /p, s/), palatalisation of a preceding consonant (/p/→ [p'], /s/→[ʃ]), or deletion of the high vocoid.

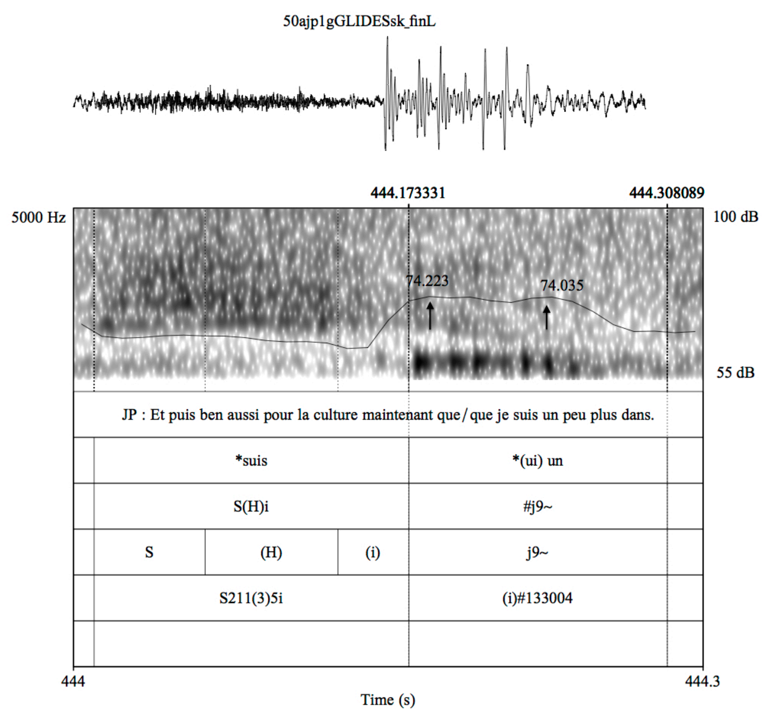


Figure 33: Spectrogram with intensity contour, oscillogram and tiers for 50ajp1 (*suis*) *un* realised with dieresis: derived across-word-boundary preceded by [s(ɥi)]

Figure 33 shows the token *suis un* realised with dieresis showing considerable coarticulatory effects. This recording shows considerable white noise evident on the spectrogram as well as the oscillogram. The signal contains, however, enough articulatory cues such that we are able to diagnose this token. The preceding consonant here is a palatalised [s]→ [ʃ]. Note that the voiceless period shows formants (1st-3rd) before the period of voicing begins. Given the voiceless formant strength followed

immediately by a voiced period showing two peaks of intensity, we find this token to be realised with dieresis. The first peak in intensity occurs early and corresponds to the high vocoid /I/ whereas the second peak occurs late and corresponds to the nasal vowel in final position of this token. Results for /C[ɥi]+V/ contexts show that cross-word-boundary glide formation resulting in syneresis does occur despite inherent complexity. However, in all instances of glide formation resulting in syneresis we observe considerable coarticulation effects or deletion of the glide [ɥ] which we attribute to a propensity for simplification in response to the added complexity. In the instances where no co-articulation effects or simplification strategies are observed, tokens are realised categorically with dieresis. Overall, 74% of tokens are realised with syneresis and 23.3% with dieresis suggesting that glide formation is privileged even at the expense of added complexity.

Figure 34 below shows the token 91aal1 *suis à* realised with syneresis.

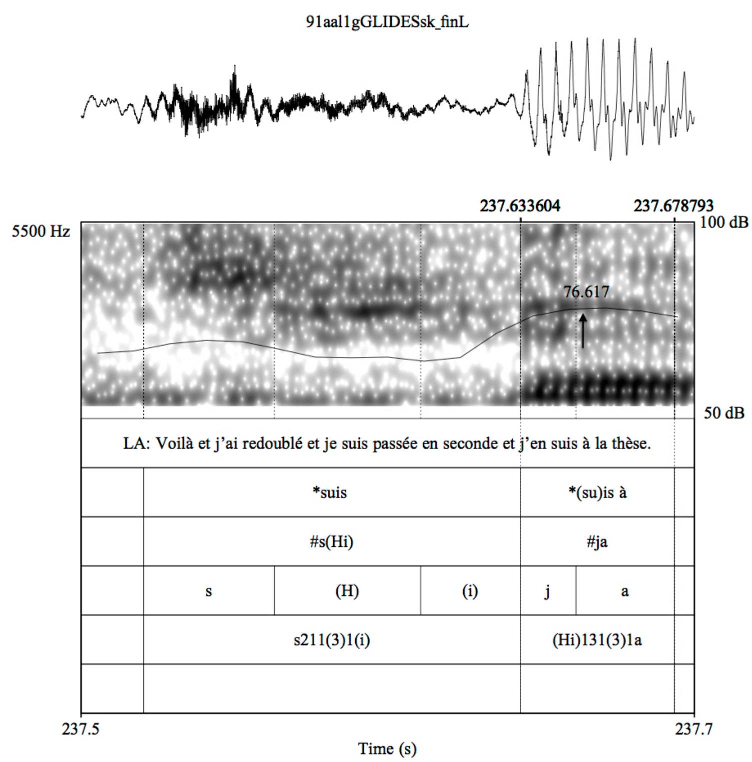


Figure 34: Spectrogram with intensity contour, oscillogram and tiers for 91aal1 (*suis*) *à* realised with syneresis: derived across-word-boundary preceded by [s(ɥi)]

Note that this token for cross-word-boundary glide formation with syneresis is produced with the alveolar fricative [s] indicated by the strength of noise at approximately 4000 Hz which is followed immediately by still voiceless formants (2nd measures 2800 Hz and 3rd measures 3400Hz) corresponding to the high vocoid /I/. These formants continue into the voiced period resulting in a single peak of intensity, which corresponds to the glide [j], followed by transition to [a].

We turn now to findings for tokens of derived /HV+V/ contexts preceded by [sk] resulting from resyllabification: (tout) *ce qui est/était* [(tu)skje]. Table 51 below shows eight tokens for the cross-word-boundary /HV+V/ environment; each token resulting in syneresis contexts involving various forms of the expression (tout) *ce qui est*.

Table 51: Derived contexts /HV+V/ preceded by complex phonological environment – /sk/ for a single variant: syneresis. N= number, (%) calculated for variant by high vocoid

Realisations	Vocoid		Total	
	I			
/sk/	N	%	N	%
syn	8	100	8	100
Total /sk/	**8**	**100**	**8**	**100**

At the core of each of these eight tokens, we observe the same cross-word-boundary /HV+V/ context preceded by /k/ for which we have already observed near categorical results when preceded by a simple phonological environment such as /k/. The eight tokens we observe here involve /k/ in the preceding environment. This context is particular to a spontaneous speech sample and involves the processes of schwa deletion and resyllabification resulting in several possible realisations: *ce qui est* [səkie] → [skie; skije; skje], for example. Nonetheless, the more complex environment of /sk/ does not impede glide formation resulting in syneresis. Figure 35 below shows the token (*tout) ce qui é(tait)* [tu.skje] a cross-word-boundary derived environment realised with syneresis.

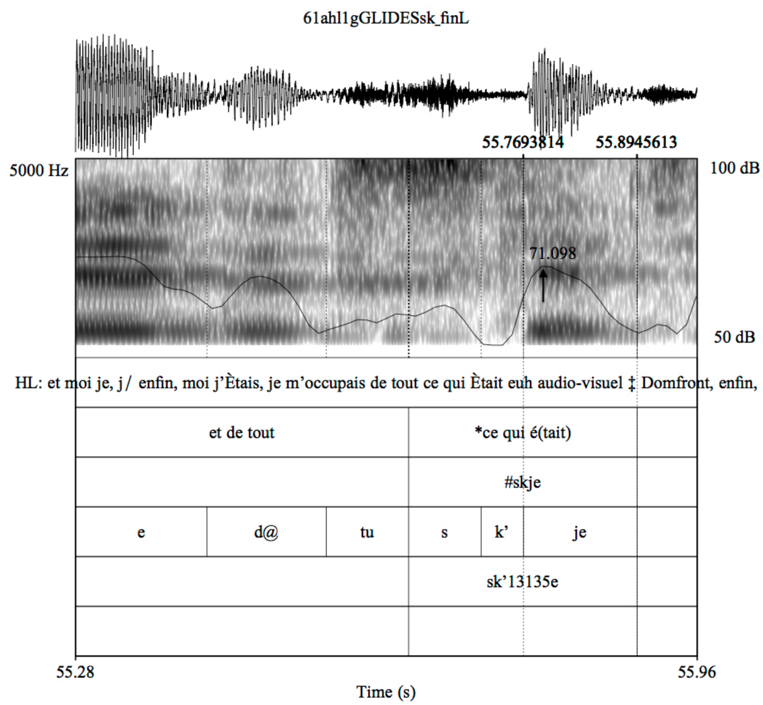


Figure 35: Spectrogram with intensity contour, oscillogram and tiers for 61ah1 *qui é(tait)* realised with syneresis [kje]: derived across-word-boundary preceded by [sk']

We note that the alveolar stop [t] preceding the high vocoid /U/ shows clear signs of assibilation which is observed by the formants observed throughout the production but also by the noise of assibilation observed at 4300-5000 Hz on the spectrogram. Formants are visible throughout [s] and while the velar stop [k] is identified through the subtle burst at its release, we observe that [k] also shows the noise of assibilation throughout its production: [k] → [k']. Classic signs of a velar pinch are not evident perhaps because of the assibilation observed.

We turn now to contexts preceded by OL (Ol and OR) sequences, of which, we find sixteen tokens overall: two preceded by Ol and fourteen preceded by OR. The table below shows totals of all tokens for /HV+V/ preceded by an OL sequence. It must be noted that these contexts do not involve exclusively derived contexts. We include here some lexicalised contexts because their variable realisations are of interest to the examination here.

Table 52: /HV+V/ contexts preceded by complex phonological environment – /OL/ for two variants: dieresis, syneresis. N= number, (%) calculated for each variant by vocoid

Variant	Vocoid				Total	
	I		Y			
OL	N	%	N	%	N	%
dieresis	12	85.7	1	50.0	13	81.3
syneresis	2	14.3	1	50.0	3	18.7
Total OL	**14**	87.5	**2**	**12.5**	**16**	100

Interestingly, we find three tokens preceded by OL realised with syneresis. These infinitive forms involve *influer* /ɛ̃fly+e/ preceded by /Ol/, and two tokens for *trier* preceded by /OR/. These word medial contexts exhibit simplification of the OL sequence (see figures shown below). Syneresis is unexpected since a preceding /Ol/ sequence is expected to block glide formation (Gougenheim 1932, Fouché 1959, Kaye and Lowenstamn 1984) yet these tokens are realised with syneresis based on visual cues from formants and intensity contour. In the following section, we will examine figures for each of these tokens in order to examine cues and the simplification observed.

The figure below shows the token *influer* realised with syneresis. This derived context is typically preceded by an /Ol/ sequence. Here we observe, however, that the liquid (and possibly the high vocoid following) have undergone considerable simplification (cf. Morin 1976).

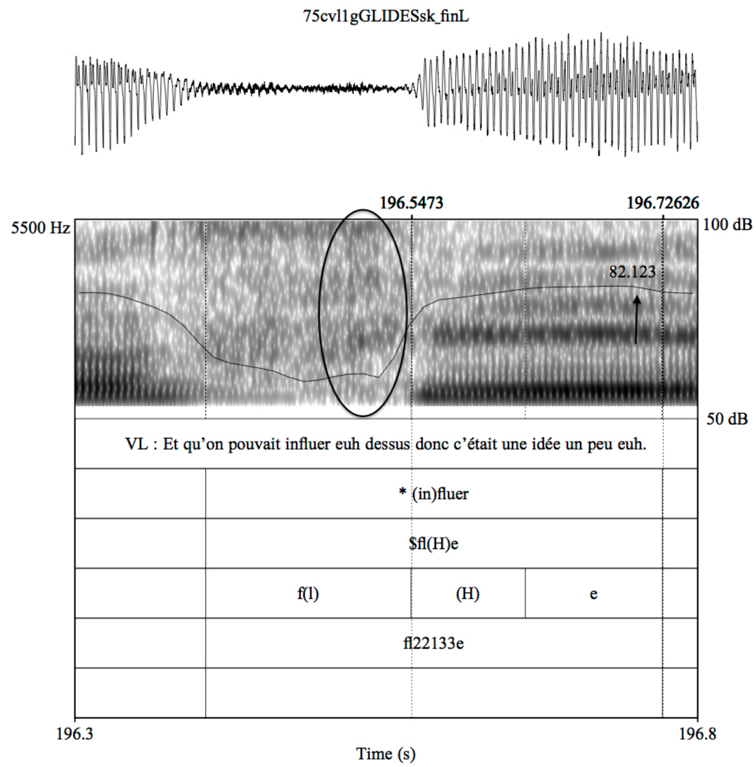


Figure 36: Spectrogram with intensity contour and oscillogram for 75cvl1 *influer* realised with syneresis with simplification of the preceding OL sequence

The token *influer* is realised with syneresis and shows considerable simplification in the preceding environment. During the vocalic period we observe very clearly the formants of the mid-front vowel [e] in a vocalic period realised with syneresis corroborated by a single peak of intensity. However, evidence of the high vowel [y] is not obvious. We can identify production of [f] during the voiceless period that precedes voicing, however, presence of [l] following is not obvious. If [l] is present it is clearly devoiced (note the sudden then gradual ascent of the intensity contour). The shape of the intensity contour, coupled with the appearance of weak formants observed during this voiceless period, indicates the presence of a voiceless periodic segment. Given the nature of spontaneous speech, it is difficult to assess this token. At the very least, we see that the liquid preceding is devoiced and the high vowel is assimilated. At most the liquid and high vocoid are both deleted.

Simplification of OL is not restricted to tokens realised with syneresis. Among tokens for lexicalised contexts preceded by an OL sequence, we find also a single instance of the word *tablier /table+ie/* which shows OL simplification. This token is unique among those mentioned thus far because there is no high vowel in root final position in the morpheme; the realisation described for *tablier* in standard French is a realisation with dieresis. The suffix *–ier* is realised with dieresis because of the OL preceding. Lyche (1979) describes individual variability, while Durand and Lyche (1999) describe grammatical variation particularly noted for langue d'Oc varieties, which they attribute to the position of the high vocoid with respect to the morpheme boundary. According to Lyche (1979), suffix variability is to be expected primarily for the high vocoid /I/ which can be attributed to language acquisition:

> From the point of view of language acquisition, the differences described here are easily explained: *tablier* never alternates and is learned as such; *oublier* can alternate, but the root vowel is never affected by any rule; in *sembliez*, the vowel [i] shows up in a few isolated forms and a simplification is to be expected. (Lyche 1979: 327).

Figure 37 below shows a spectrogram with intensity contour and oscillogram for the token *tablier* realised with dieresis. The context preceded by OL shows the liquid and the high vocoid have undergone considerable simplification.

We observe in the figure below an instance of the word *tablier* which, like *sembliez* described by Lyche above, undergoes simplification through liquid deletion or liquid vocalisation. It is not clear as to the state of the liquid preceding, but it is our position that the token shows simplification.

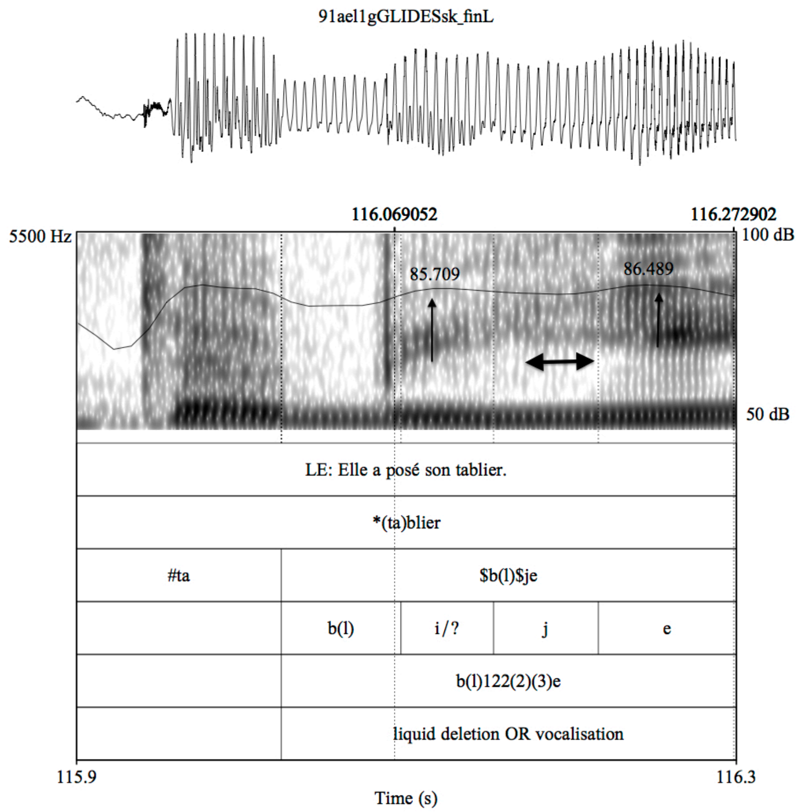


Figure 37: Spectrogram with intensity contour and oscillogram for 91ael1 *tablier* realised with dieresis with simplification of the preceding OL sequence

On the spectrogram in Figure 37, we observe formant transition with strong energy at voicing onset after the burst of [b]. The [l] of the OL sequence preceding appears to be vocalised and formants transition to the high vowel /i/. We note that energy weakens while the 2nd and 3rd formants transition to [e]. This token is realised with dieresis which is corroborated by two peaks in intensity.

In Figure 38 below, we examine a token for the word *quatrième* realised with syneresis and simplification of the preceding OL sequence.

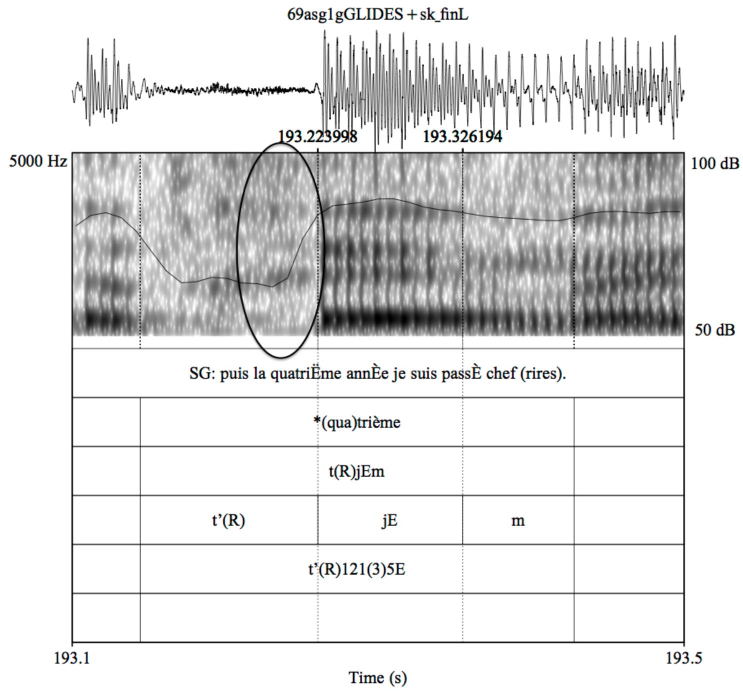


Figure 38: Spectrogram with intensity contour and oscillogram for 69asg1 *quatrième* realised with syneresis with simplification of the preceding OL sequence

Similar to the preceding token, this OL sequence shows simplification of the preceding OL sequence through the devoicing of [ʀ]. Note also the presence of weak formants, which show signs of transition (circled), during the voiceless period following the burst of [t] and the peak of intensity which we take to indicate the presence of voiceless [ʀ]. This token shows a single peak on the intensity contour indicating that it is realised with syneresis.

To this point, we have observed only simplification of OL sequences through the devoicing or vocalisation of the preceding liquid. Also among the word medial tokens preceded by OL, the token *voudriez* is realised with schwa epenthesis between the obstruent and liquid thereby simplifying the complexity that an OL sequence may present [vudəʀje]. Recall that this phenomenon is described by Morin (1976) as one strategy employed to relax the tension of an OL sequence.

Figure 39 below shows schwa epenthesis observed in a token for *vous voudriez*.

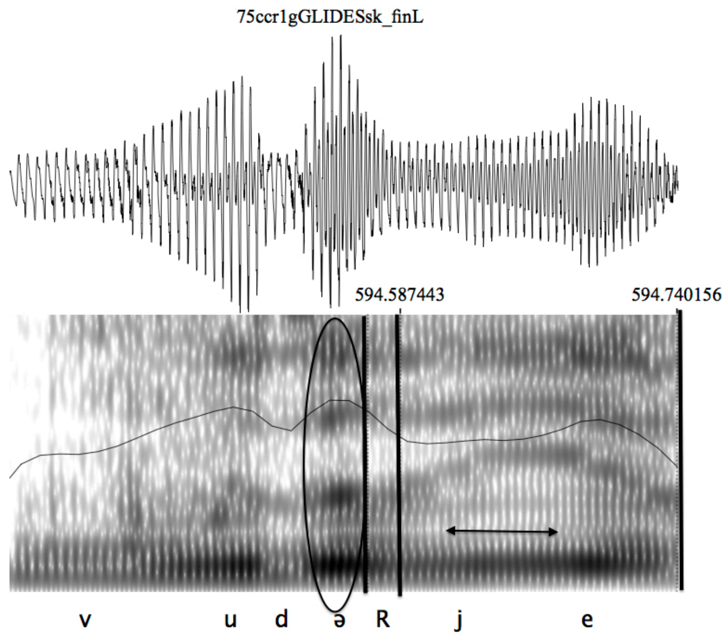


Figure 39: Spectrogram with intensity contour and oscillogram for 75ccr1 *voudriez* realised with schwa epenthesis and syneresis.

We observe that schwa intervenes between the voiced alveolar [d] and the following liquid [R] as evidenced by three strong formants on the spectrogram (circled). The oscillogram and spectrogram show the energy of formants for schwa followed by glide transition before [e]. This token is in keeping with the findings of Chow and Poiré who in their study of consonant plus liquid sequences in Windsor French find that "[...] schwa epenthesis tends to occur much more often after voiced than unvoiced consonants" (2007: 99). In the preceding tokens, we have observed various instances of OL simplification in word medial contexts of /HV+V/ which may influence the realisation (syneresis or dieresis) of a token. We note also that we observe several instances of schwa metathesis that triggers GF when a word initial [ʀ] is preceded by a high vowel in *je suis retraité* [ʒəsɥijəʀtʀɛte].

We turn now to findings for the cross-word-boundary tokens preceded by OL sequences.

Tokens for cross-word-boundary contexts are realised categorically with dieresis. These include three instances of the past participle *pris*: *pris un*, *pris une,* and *pris en* as well as one instance of *Hongrie en* and a single instance of a past participle *cru* plus preposition: *cru en*. In Figure 40 below we examine a token for *pris une.*

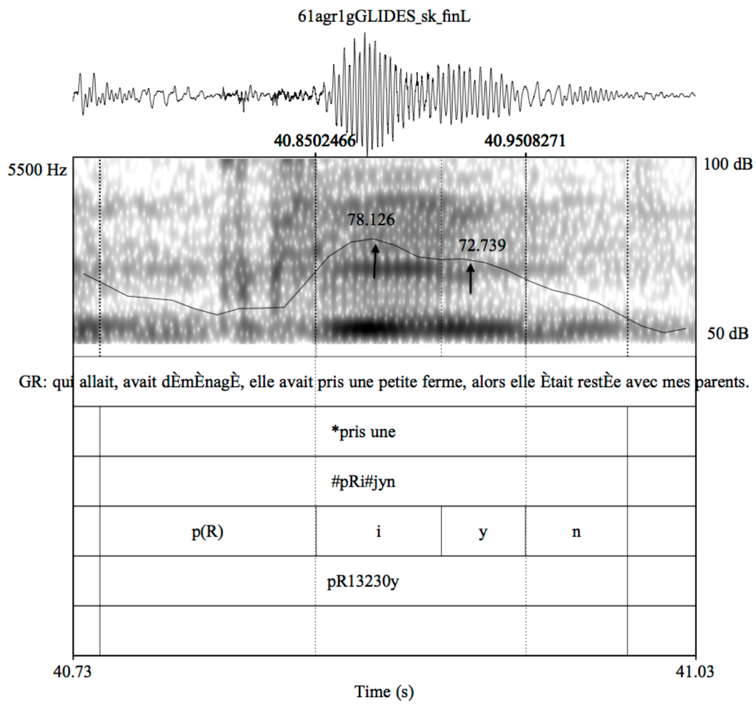


Figure 40: Spectrogram with intensity contour, oscillogram and tiers for 61agr1 *pris une* realised with dieresis: derived across-word-boundary preceded by [pR]

Figure 40 shows the token *pris une* realised with dieresis; we find no signs of formant transition, however we observe a clear change in the 2nd formant that signals the production of [i] (strongest formant strength) to [y] (weaker formant strength). The intensity contour shows two peaks that correspond to these two high vowels corroborated by the wave amplitude of the oscillogram. We observe two periods of strengthened formant contour in F1 separated by a brief period of intensity loss which we attribute to transition between vocalic periods. The noted absence of formant transition points to a realisation of dieresis with hiatus.

6.3.4 Observations: variability of lexicalised and derived contexts

Our findings show that lexicalised and derived contexts are not equal with respect to distribution and variability. Lexicalised contexts are realised most commonly with syneresis and show less variability. In fact, where variability is present, we observe a possible correlation with the high vocoid involved.

We also observe that OL sequences may undergo processes of simplification (schwa epenthesis or liquid deletion cf. section 3.4.1, Morin 1976) which may also trigger voicing assimilation between a preceding consonant and the following high vocoid. Overall data shows the expected distributional hierarchy discussed in section 4.3.2: I $\gg$ U $\gg$ Y, which is evident also in overall lexicalised data. Distribution across high vocoids in derived contexts shows that tokens involving /I/ are consistently most prevalent, but in derived contexts tokens involving the high vocoid /Y/ outnumber those involving /U/. An examination of each variant shows that high vocoid deletion shows the inverse. The high vocoid /Y/ is deleted most frequently followed by /U/ and finally /I/ whereas tokens for realisations with dieresis privilege /I/ with the rounded high vocoids / U, Y/ showing equal rates of deletion.

Lexicalised and derived glide contexts also distinguish themselves on the basis of variability. Derived contexts, although fewer, show more variability than is observed for lexicalised contexts. Across three potential variants both lexicalised and derived contexts show strong preference for syneresis yet they can be distinguished by the varying degrees of variability exhibited. Lexicalised contexts show 95.7% syneresis and ultimately less variability. Although dieresis does occur (31 tokens 1.1%), the contexts in which dieresis is observed are largely predictable (following OL sequences, for example). Note also that deletion of a high vocoid, while more common in derived contexts, occurs in both lexicalised and derived contexts. Derived contexts, with only 61.1% syneresis, show a higher degree of variablility: 28.1% dieresis and 10.9% high vocoid deletion (DEL), as compared with lexicalised contexts.

Across the various preceding environments, contexts preceded by a word boundary (#), *où est*, for example, clearly privilege dieresis and liquids, fricatives and stops privilege syneresis however finding are not always clear. Nasals stops /m/ and /n/, for example, are nearly balanced between syneresis (52.5%) and dieresis (47.5%). In fact when we isolate tokens by the preceding environment we find that each preceding segment may be a separate factor in the realisations observed, but findings are not always clear; the number of tokens prevent us from making conclusions.

In the following sections, we examine variable realisations of tokens in lexicalised and derived contexts across three high vocoids, for each region by survey.

6.4 Data analysis by region: Île de France; Rhône Alpes; Normandie

In the following sections we detail our findings for each of the surveys in our sample. We begin with distributions of lexicalised GV and derived /HV+V/ contexts for each region (cf. Table 54 below) then we breakdown these distributions by survey (cf. Table 55 below). Before looking at the findings for each region we will discuss the overall demographic information for all of the speakers included in our study.

Table 54 below shows the distribution of data (lexicalised and derived contexts) across three regions

Table 53: Derived versus lexicalised contexts, N= number, percentages (%) calculated by region

Region	Derived		Lexicalised		Total	
	N	%	N	%	N	%
Île de France	190	14.5	1123	85.5	1313	40.9
Normandie	165	15.7	886	84.2	1052	32.7
Rhône Alpes	104	12.3	744	87.7	848	26.4
Grand Total	**459**	**14.3**	**2753**	**85.7**	**3212**	**100**

Our data, taken from spontaneous speech samples and selected based on signal quality, are not balanced across regions; 40.9% of tokens are drawn from *Île de France*, 32.7% of

tokens from *Normandie* and 26.4% from *Rhône Alpes*. Table 54 below shows distribution of lexicalised and derived contexts across each of six PFC surveys included in our study.

Table 54: All lexicalised and Derived tokens by survey for each region, PFC alpha-numeric correspondences: 75c *Paris, Centre ville*, 91a *Brunoy*, 50a *Brécey*, 61a *Domfrontais*, 38a *Grenoble* and 69a *Lyon,* percentages at bottom calculated on the total number of tokens for each survey

Region	*Île de France*		*Normandie*		*Rhône Alpes*		Total
PFC Survey	75c	91a	50a	61a	38a	69a	
	%	%	%	%	%	%	%
Lexicalised	83.8	87.7	84.8	83.8	86.5	89.8	85.7
Derived	16.2	12.3	15.2	16.2	13.5	10.2	14.3
Grand Total	**22.5**	**18.4**	**17.0**	**15.7**	**16.6**	**9.7**	**100**

As noted, the number of tokens observed varies considerably from survey to survey. Despite the imbalance, the distribution of contexts is relatively constant: lexicalised tokens consistently represent 85% of tokens and derived approximately 15% of tokens.

In Table 56 below, we examine the demographic information: gender and age, for all speakers in our study by region for each survey.

Table 55: Demographics for each speaker by region/survey grouped by age and gender.

	Age				
	20-30	31-59	60+	F	M
Region/Survey					
Île de France					
91a Brunoy	20-30	31-59	60+	F	M
Female	1	--	3	4	
Male	1	1	2		4
75c Paris Centre	20-30	31-59	60+		
Female	2	1	1	4	
Male		1			1
Total *Île de France*	4	3	6	8	5
Normandie					
50a Brécey	20-30	31-59	60+	F	M
Female		1	1	2	
Male	1	1			2
61a Domfrontais	20-30	31-59	60+	F	M
Female		2		2	
Male	1		1		2
Total *Normandie*	2	4	2	4	4
Rhône Alpes					
38a Grenoble	20-30	31-59	60+	F	M
Female	1	1	2	4	
Male	1		1		2
69a Lyon	20-30	31-59	60+	F	M
Female		1	1	2	
Male	1	1			2
Total *Rhône Alpes*	3	3	4	6	4
Grand Total	**9**	**10**	**12**	**18**	**13**

The small number of speakers retained here does not allow for surveys balanced for age and gender.

In the following sections, we discuss distribution of lexicalised and derived contexts across speakers and across high vocoids for each survey by region followed by an examination of the variability of realisation of tokens for each context across each of two surveys from three regions: 6.5 *Île de France*, 6.6 *Normandie* and finally 6.7 *Rhône Alpes*.

6.5 Île de France

Île de France is the most populated region of France (11,786,234)[31] and it represents 40.9% of our data (1314 tokens). These data are from the PFC surveys 75c *Paris Centre ville* (5 speakers retained) and 91a *Brunoy* (8 speakers retained). In this section, we discuss demographic information for each speaker as well as examine the distribution of tokens for lexicalised and derived contexts for each survey across speakers and across high vocoids.

Table 57 below shows findings for lexicalised and derived data from each of two *Île de France* surveys.

Table 56: *Île de France* all tokens observed across two PFC surveys. N= number, percentages (%), data for each of two surveys in the region

Île de France						
Survey	91a Brunoy		75c Paris		Total	
	N	%	N	%	N	%
Derived	73	12.3	139	18.7	190	14.4
Lexicalised	518	87.7	605	81.3	1123	85.6
Grand Total	**591**	**45.0**	**722**	**54.9**	**1313**	**100**

As noted, the selection of speakers based on acoustic quality of recordings results in an imperfect distribution. Tokens are not well balanced for social factors among speakers which when coupled with the uncontrolled nature of spontaneous speech leads to further incongruities in data distribution; Paris (75c), for example, provides 722 tokens (54.9% of the region) from five speakers while the Brunoy (91a) data shows 591 (45.1%) tokens from eight speakers.

Recall from our previous discussion that the inherently variable nature of spontaneous speech data exacerbates issues of balance across glide contexts. As a result, we are unable to examine the implications for age and sex as factors in the variable glide realisation. In

[31] Source: http://www.insee.fr/

the discussion to follow, we examine the demographic information for each participant in two *Île de France* surveys. The table below shows demographic information for five speakers retained from the PFC survey 75c *Paris, Centre ville*.

Table 57: *Île de France*, 75c *Paris, Centre ville*, age (approximate), sex and N= number of tokens, % (calculated by speaker) for each speaker

PFC speaker code	**Age**	Total tokens	
Female		N	%
75cab1 (1935)	70	163	22.6
75ccr1 (1970)	35	75	10.4
75csb1 (1979)	26	157	21.7
75cvl1 (1979)	26	142	19.7
Total F		**537**	**74.4**
Male		N	%
75ccm1 (1971)	34	185	100
Total M		**185**	**25.6**
Total retained *Paris* (75c)		**722**	**54.9**
Île de France		**1313**	**100**

Five speakers from the Paris (75c) survey: four female, one male, range in age from a single speaker of 71 years to four other speakers who are each below the age of 35. From these speakers, we have identified 722 tokens representing 54.9% of HVV contexts examined in Île de France.

Next, we examine the demographic information for eight speakers retained for analysis from the Brunoy (91a) survey. These speakers represent a balanced group with respect to sex; however, speakers are not balanced for age.

Table 58: ***Île de France*****,** ***Brunoy*** **(91a) survey,** age, sex and N= number of tokens, % calculated by speaker

PFC speaker code	Age	Total tokens	
Female		N	%
91aal1	27	63	10.7
91aed1	60	80	13.5
91ael1	73	68	11.5
91asl1	64	88	14.9
Total F		**299**	**50.6**
Male			
91acs2	22	77	13.0
91adb1	54	56	9.5
91ajc1	63	90	15.2
91amb1	62	69	11.7
Total M		**293**	**49.6**
Total *Brunoy*		**591**	**45.0**
Île de France		**1313**	**100**

As shown in Table 59, we retain 30% more speakers from *Brunoy* than from *Paris Centre ville* yet the numbers of tokens identified are fewer (45.1% of region). Speakers are balanced for sex: four male, four female, but not for age; we observe a full range of ages (20-73 years), yet, only two speakers are below 30 and six are above 50 years of age.

Let us turn now to the distribution of lexicalised and derived contexts for each speaker for each of two *Île de France* surveys. We begin with findings for 722 tokens from *Paris Centre ville* (75c). Table 60 below shows derived and lexicalised tokens for each of the five speakers.

Table 59: ***Île de France*****, 75c** ***Paris, Centre ville*****, all tokens** N= number of tokens, % calculated by speaker for derived and lexicalised contexts

	Speaker										Total	
	75cab1		75ccm1		75ccr1		75csb1		75cvl1			
	N	%	N	%	N	%	N	%	N	%	N	%
Derived	13	8.0	40	21.6	7	9.3	25	15.9	32	22.5	117	16.2
Lexicalised	150	92.0	145	78.4	68	90.7	132	84.1	110	77.5	605	83.8
Grand Total	**163**	**122.6**	**185**	**25.6**	**75**	**10.4**	**157**	**21.7**	**142**	**19.7**	**722**	**100**

Overall distribution of lexicalised versus derived data for the *Paris* survey resembles patterns observed for the overall data shown in Table 25 above (85.7% versus 14.3%);

lexicalised GV tokens (83.8%) are much more prevalent than derived tokens (16.2%). While numbers do vary across five speakers, derived tokens represent no greater than 22.5% of tokens, while lexicalised are never less than 77.5% of tokens. Inter speaker variation in rates of derived versus lexicalised must be expected given the nature of spontaneous speech; here, lexicalised data ranges 92% to 77.5% and derived data ranges 22.5% to 8% (14.5%).

In Table 61, we examine the distribution of 722 tokens for lexicalised and derived contexts across three high vocoids.

Table 60: *Île de France*, *Paris* (75c), distribution of all tokens across three high vocoids /I U Y/, N= number of tokens, % calculated by vocoid

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
Derived	76	19.5	7	3.6	34	24.3	117	16.2
Lexicalised	313	80.5	186	96.4	106	75.7	605	83.8
Grand Total	**389**	53.9	**193**	26.7	**140**	19.4	**722**	100

Of note, derived data show a changed pattern of distribution across all high vocoids similar to the pattern observed for overall derived data examined earlier (cf. also Table 22): /I/ 65% ≫ /Y/ 29% ≫ /U/ 6%, whereas lexicalised data follow the more commonly described distributional hierarchy: /I/ ≫ /U/ ≫ /Y/ (cf. section 4.3.2). We must note that derived data include both word medial as well as cross-word-boundary contexts. 75c Paris data include 22 instances of cross-word-boundary contexts involving the pronoun *tu*, for example. While we may wish to attribute the changed distribution of tokens across high vocoids to the inclusion of cross-word-boundary contexts, when these contexts are excluded from the derived data, the changed hierarchy persists; among all derived /HV+V/ contexts, we find 12 tokens in a word medial environment: five involve /I/ (41.7%), four /Y/ (33.3%) and three /U/ (25%). We attribute this change in distribution across high vocoids to the rates at which word final high vocoids occur in the lexicon (Juilland 1967).

We turn now to an examination of data from the second *Île de France* survey: *Brunoy* (91a). For ease of presentation of data for eight speakers, we will examine results for female speakers followed by the male speakers. The table below shows the distribution of derived and lexicalised contexts across four female speakers from *Brunoy*.

Table 61: *Île de France*, 91a *Brunoy*, derived and lexicalised contexts, for female speakers, N= number of tokens, % calculated by speaker

	Speaker				Total	
	91aal1	91aed1	91ael1	91asl1		
	%	%	%	%	N	%
Derived	15.9	21.2	13.2	8.0	41	13.7
Lexicalised	84.1	78.8	86.8	92.0	258	86.3
Grand Total	21.1	26.7	22.7	29.4	**299**	100

The overall distribution for derived (13.7%) and lexicalised (86.3%) contexts is consistent with patterns observed thus far (cf. tables 26 and 27 above). Table 62 shows derived contexts as low as 8% (speaker sl1) and as high as 21.2% (speaker ed1) across four speakers. Variability in the distribution of contexts across speakers is attributed to the random nature of spontaneous speech.

Let us look now at the distribution of contexts for the male speakers of *Brunoy*. Table 63 below shows distribution of derived and lexicalised contexts for four male speakers.

Table 62: *Île de France*, 91a *Brunoy*, derived and lexicalised contexts for male speakers, N= number of tokens, % (calculated by speaker)

	Speaker				Total	
	91acs2	91adb1	91ajc1	91amb1		
	%	%	%	%	N	%
Derived	7.8	16.1	11.0	10.1	32	10.9
Lexicalised	92.2	83.9	89.0	89.9	261	89.1
Grand Total	**26.3**	**19.1**	**31.0**	**23.6**	**292**	**100**

Overall distribution of derived and lexicalised contexts across male speakers shows a slightly decreased rate for overall derived data (10.9%) compared with the overall average for *Brunoy* (12.7%) which necessarily results in an increased rate for lexicalised

data (89.1%) compared to 86.3% overall. Fluctuations of this sort must be taken as incidental given the nature of spontaneous speech.

We look now at the distribution of contexts across three high vocoids. Table 64 below shows the distribution for lexicalised and derived data from *Brunoy* across three high vocoids.

Table 63: *Île de France*, *Brunoy* (91a), distribution of all tokens across three high vocoids /I U Y/, N= number of tokens, % calculated by vocoid

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
Derived	54	17.5	5	2.8	16	15.4	75	12.7
Lexicalised	255	82.5	173	97.2	88	84.6	516	87.3
Grand Total	**309**	**52.3**	**178**	**30.1**	**104**	**17.6**	**591**	**100**

While overall distribution of lexicalised and derived contexts across three high vocoids differ slightly between *Paris* (16.2% derived -83.8% lexicalised) and *Brunoy*, the distribution of tokens across glides shows a trend that is consistent across surveys; derived contexts average approximately 15% with lexicalised contexts averaging approximately 75% (cf. Table 25 above).

In the sections that follow, we examine the other surveys for two other regions in our study. For each region we will discuss the demographic information for speakers and the distribution of lexicalised and derived contexts followed by discussion of realisations observed across speakers and across high vocoids. We begin by an examination of realisation of tokens for lexicalised and derived contexts across speakers and across high vocoids for the *Île de France* region.

6.5.1 *Île de France*: variable realisations across lexicalised and derived data

Table 65 shows the variability observed across all tokens for *Paris* (75c).

Table 64: *Île de France*, *Paris* (75c), distribution of all tokens across three high vocoids for all potential variants, N= number of tokens, % calculated by vocoid for each variant: deleted high vocoid (DEL), dieresis, and syneresis

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	3	0.8	--	--	28	20.0	31	4.3
dieresis	19	4.9	2	1.0	9	6.4	30	4.2
syneresis	367	94.3	191	99.0	103	73.6	661	91.6
Grand Total	**389**	**53.9**	**193**	**26.7**	**140**	**19.4**	**722**	**100**

Tokens from the Paris region show a strong preference for realisations in syneresis (91.6%). Dieresis (4.2%) and deletion (4.3%) are infrequent but show a strong correlation to the high vocoid involved; both are most evident in tokens involving /Y/ whereas the high vocoid shows no deletion and only 1% (2 tokens) for dieresis.

Let us turn now to variable realisations observed in the data from the Paris survey. Table 66 below shows findings for realisations of all lexicalised data.

Table 65: *Île de France*, *Paris* (75c), Lexicalised contexts, N= number of tokens, % (calculated by vocoid) for two variants ((DEL) = deleted high vocoid, syneresis)

	Vocoid						Total	
	I		U		Y			
Variant	N	%	N	%	N	%	N	%
DEL	1	0.03	--	--	12	11.3	13	2.2
syneresis	312	**99.7**	186	100	94	88.7	592	97.7
Grand Total	**313**	**51.7**	**186**	**30.7**	**106**	**17.5**	**605**	**100**

Data from the Paris survey show that at 97.7% syneresis is highly favoured. Deleted high vocoid is the only other variant observed and is most evident for tokens involving /Y/ in words such as *puis* [pi], *je suis* [ʃi̥]. We find a single token of *science* [sɛ̃s] realised with a deleted high vocoid /I/. Tokens showing reduced forms with high vocoid deletion are not uncommon for the words *puis* and *suis*, we can account for their regularity through lexicalisation of frequently occuring stretches of speech (cf. Bybee 2003). Deleted /I/ in *science* is quite unusual which we attribute to the particularities of spontaneous speech.

Table 67 below shows all tokens for lexicalised GV contexts across five speakers retained from the *Centre ville* survey.

Table 66: *Île de France*, *Paris* (75c), Lexicalised contexts, N= number of tokens, % (calculated by speaker) for lexicalised variants (deleted high vocoid-DEL and syneresis-Syn.)

	Speaker										Total	
	75cab1		75ccm1		75ccr1		75csb1		75cvl1			
	N	%	N	%	N	%	N	%	N	%	N	%
DEL	-	--	8	6.2	-	--	-	--	5	4.5	14	2.3
Syneresis	150	100	136	93.8	68	100	132	100	106	95.5	592	97.7
Grand Total	**150**	**24.8**	**145**	**23.9**	**68**	**11.2**	**132**	**21.8**	**111**	**18.3**	**605**	**100**

As observed earlier, lexicalised GV contexts privilege syneresis at a rate of 95.7% overall. Parisian data do not deviate from this pattern; syneresis is privileged in lexicalised GV contexts (97.5%) while the balance of tokens (2.3%) are realised with a deleted high vocoid (DEL). Three of five speakers show categorical syneresis while two speakers: 75ccm1 and 75cvl1 show deletion of the high vocoid (2.5%). Perhaps, not surprisingly, the majority of tokens for deleted high vocoids (86.7%) involve instances of the words *puis* (realised as [pi]) and *suis* (realised as [ʃ]). These well-noted examples may be attributed to coarticulation effects and/or lexicalisation (cf. Bybee 2002). The remaining two tokens are less expected, however, including one instance of a deleted /j/ (*science* [sãs]) and one instance of deleted /w/ in the word *droit* [dRa]. Excluding the token for *science*, all other instances of a deleted high vocoid involve a lexicalised diphthong: [ɥi], [wa]. The phrase *je suis* is commonly reduced to [ʃɥi] or even [ʃ]. Note, we observe approximately 40 such instances of this phrase reduction: *je suis* → [ʃ].

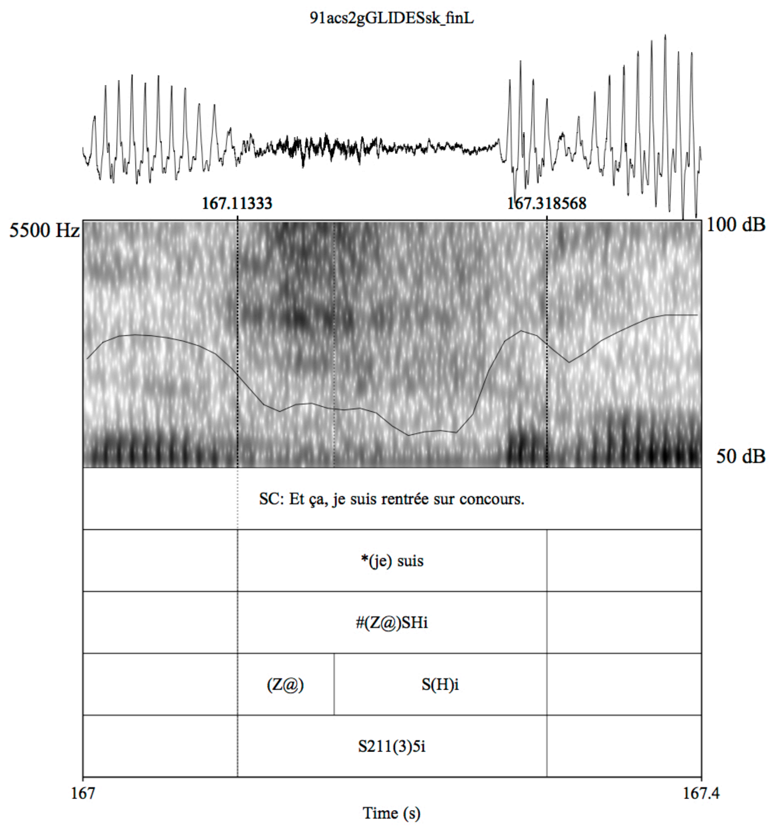


Figure 41: 91acs2 *je suis* [ʒəʃ(ɥ)i] reduced

In Figure 41 we show the token (*suis* [ʃ]) with reduction of the diphthong [ɥi] such that the brief voiced period following the voiceless palatal fricative [ʃ], contains only the vowel [i]. We observe evidence of formant strength during production of the voiceless palatal fricative corroborated by the intensity contour which may indicate presence of a voiceless high vocoid. The implication here is that the glide [ɥ] part of the diphthong [ɥi] is devoiced if not deleted in the reduced form realised as [ʃ(i)] (cf. Cedergren, H. J. and L. Simoneau 1985 for high vowel deletion in Montreal French), although further analysis would be required to determine the presence of devoiced glide.

Lexicalised data in the Paris (75c) survey suggest that realisations with a deleted high vocoid may be conditioned by an individual speaker. The reductions and deletions observed here involve only two speakers: ccm1 (34), cvl1 (36), who are among the youngest participants examined. Speaker cab1, the oldest speaker in our sample, does not

show the same tendencies for reduction (assimilation) and deletion, while the two speakers who do show reduction in these contexts are each below the age of 35.

We turn now to the examination of variable realisations in derived data from the Paris (75c) survey.

Table 67: *Île de France*, *Paris* (75c), all derived /HV+V/ context realisations, N= number of tokens, calculated for each variant: deleted high vocoids (DEL), dieresis, and syneresis) across three high vocoids

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	2	2.6	-	--	16	47.1	18	15.4
Dieresis	19	25.0	2	28.6	9	26.4	30	25.6
Syneresis	55	72.4	5	71.4	9	26.5	69	59.0
Grand Total	**76**	**64.9**	**7**	**6.0**	**34**	**29.1**	**117**	**100**

Table 68 shows the extent of variation in the realisations for all derived glide contexts across three high vocoids; variants observed include high vocoid deletion (14.4%), dieresis (25.6%) and syneresis, which is most prevalent at 59% across all derived tokens.

Among the 117 derived /HV+V/ tokens, we find twelve tokens (10.3%) for word medial contexts and 105 tokens (94.9%) for cross-word-boundary derived contexts. Word medial tokens include verbs such as *jouer, louer, influer*, for example, which are realised with near categorical syneresis (91.7%) across all three high vocoids. A single token shows dieresis for the infinitive form of the verb *continuer* [nyɥe] which is found in the verb phrase: *je peux continuer essayer,* and exhibits vocalic fusion between the final syllable of *continuer* and the vowel initial *essayer* [kɔ̃tiny**ɥe**sej(e)]. This speaker is using an intonation that could be described as a singing intonation to which we attribute the production of dieresis. Complexity of the phrasal structure might also contribute to the realisation with dieresis in this context; however, further study of the effects for this kind of complexity would be required.

Another word medial token observed involves the verb *influer* which also presents an /HV+V/ environment preceded by an OL sequence which is realised with considerable

reduction. The inherent complexity of a preceding OL sequence is simplified through the deletion of the preceding liquid (shown below).

We turn now to an examination of 105 tokens for derived cross-word-boundary contexts. Table 69 below shows findings for all derived /HV+V/ tokens observed in cross-word-boundary context.

Table 68: *Île de France*, *Paris* all tokens for cross-word-boundary derived /HV+V/ contexts, N= number, calculated for each variant: deleted high vocoid (DEL), dieresis, and syneresis, across speakers

	Speaker										Total	
	75cab1		75ccm1		75ccr1		75csb1		75cvl1			
	N	%	N	%	N	%	N	%	N	%	N	%
DEL	--	--	9	25.0	1	20.0	--	--	8	26.7	18	17.1
dieresis	**7**	**63.6**	8	22.2	1	20.0	**10**	**43.5**	3	10.0	29	27.6
syneresis	4	36.4	19	52.8	3	60.0	**13**	**56.5**	19	63.3	58	55.2
Grand Total	**11**	**10.5**	**36**	**34.3**	**5**	**4.8**	**23**	**21.9**	**30**	**28.6**	**105**	**100**

The *Paris Centre ville* (75c) data show that, across all five speakers, syneresis is most prevalent (55.2%) followed by dieresis (27.6%) then high vocoid deletion (17.1%). Individual patterns of variability are observed across speakers; cab1 privileges dieresis (63.6%) while csb1 uses both forms almost equally (dieresis 43.5% and syneresis 56.5%).

Variability in token realisations may also be conditioned by the particular high vocoid involved in a given context. We examine next the distribution of contexts across each of three high vocoids. In Table 70, we show findings for all derived cross-word-boundary contexts across high vocoids for each variant observed.

Table 69: *Île de France*, *Paris* all derived cross-word-boundary contexts for three variants: DEL, dieresis, syneresis. N= number, % calculated for each variant: deleted high vocoid (DEL), dieresis, and syneresis, across speakers

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	2	2.8	--	--	16	53.3	18	17.1
dieresis	19	26.8	2	0.5	8	26.7	29	27.6
syneresis	50	70.4	2	0.5	6	20.0	58	55.2
Grand Total	**71**	**67.6**	**4**	**3.8**	**30**	**28.6**	**105**	**100**

We note that distribution of cross-word-boundary derived tokens across each of three high vocoids shows that /U/ is least prevalent (4.8%), followed by /Y/ at 9.6%, while /I/ is most prevalent at 85.5%. Across variants, deletion is least prevalent (17.1%), however, /Y/ is deleted most (53.3%) compared to zero tokens for /U/ and only 2.8% of tokens for /I/. Interestingly, where /Y/ is not deleted tokens show slightly more dieresis than syneresis. The vocoid /U/ shows no deletions and parity for realisations with syneresis and dieresis. Finally the high vocoid /I/ shows only two tokens realised with deletion of the high vocoid and where the vocoid is maintained syneresis (70.4%) is preferred with 26.8% realised with dieresis.

We note that tokens above include 22 instances of the pronoun *tu* followed by a vowel initial verb: *tu as*, *tu étais*, *tu entres,* for example, as well as 83 tokens for all other cross-word-boundary contexts (*suis aprés* [ʃ(ɥ)ja], *vécu à* [kɥa], *qui était* [kje], for example). In Table 71 we show findings for 22 tokens of the pronoun *tu* followed by a vowel initial verb are shown by variant observed for each of five speakers.

Table 70: 75c *Paris*, Derived /tu+V/ cross-word-boundary contexts, N= number of tokens**,** % calculated by speaker for each variant: deleted high vocoids DEL, dieresis, and syneresis

	Speaker										Total	
	75cab1		75ccm1		75ccr1		75csb1		75cvl1			
	N	%	N	%	N	%	N	%	N	%	N	%
DEL	--	--	8	80	--	--	--	--	8	80	16	72.7
dieresis	2	100	1	10	--	--	--	--	1	10	3	13.6
syneresis	--	--	1	10	--	--	--	--	1	10	3	13.6
Grand total	**2**	**9.1**	**10**	**45.5**	**--**	**--**	**--**	**--**	**10**	**45.5**	**22**	**100**

Of these, 16 (72.7%) are realised with a deleted high vocoid (DEL). Only two speakers exhibit a deleted high vocoid /Y/ in this environment. These two speakers are the youngest speakers retained for our study: cm1 (M, 34), cvl1 (F, 26), while the oldest speaker retained in our sample of the Paris survey produces two tokens: tu es, tu entres, realised with dieresis only. Speakers cm1 and vl1 produce ten tokens each realised with variability: majority deleted and one token each realised with syneresis and dieresis. All other speakers (cr1 and sb1) show no instances of the pronoun tu followed by a vowel initial verb. Speaker cm1 produces ten tokens for this context of which eight are deleted, one of two tokens for tu es shows syneresis and a single token is produced with dieresis (tu écrives). Speaker vl1 produces two of six tokens for tu as with variability: one with syneresis, one with dieresis; all other tokens are produced with deleted /Y/. In summary, our oldest speaker cab1 produces only dieresis in the context of the pronoun tu followed by a vowel initial verb while the younger speakers show a tendency to delete the high vocoid but, when not deleted these contexts are produced primarily with syneresis. Realisation of the /tu + V/ may be accompanied by a degree of assibilation which effects which we attribute to the environment in spontaneous speech. We observe assibilation in tokens from these two speakers which is not evident when the high vocoid is deleted; tu as, for example, is realised as [ta] not [t'a;tsa] which suggests that high vocoid deletion occurs before assibilation. Figures 42-43 below show two such instances from each of these two speakers.

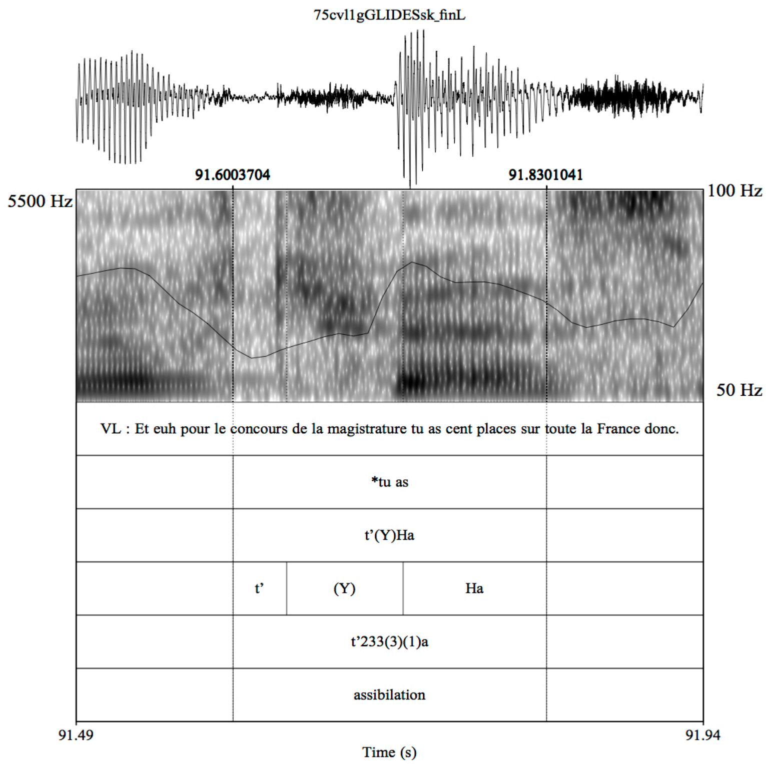


Figure 42: 75cvl1 *tu as* realised with dieresis (epenthesis), assibilation, and high vocoid devoicing

Figure 42 shows two peaks on the intensity contour, indicating dieresis. On the spectrogram we observe two strong formants during the voiceless period preceding voicing onset which we take to indicate presence of a voiceless vowel (cf. Cedergren and Simoneau 1985). Also we note that formant transition during the voiced period of vowel productionis not obvious, however, if we consider the voiceless period preceding with the voiced period that follows, formant transition is more evident. Transition here corresponds to the greatest peak of intensity on the contour as well as on the oscillogram. We note also that the preceding [t] shows considerable assibilation marked by the energy observed at approximately 5000 Hz.

In Figure 43 below, we examine a token for *tu es* realised with syneresis.

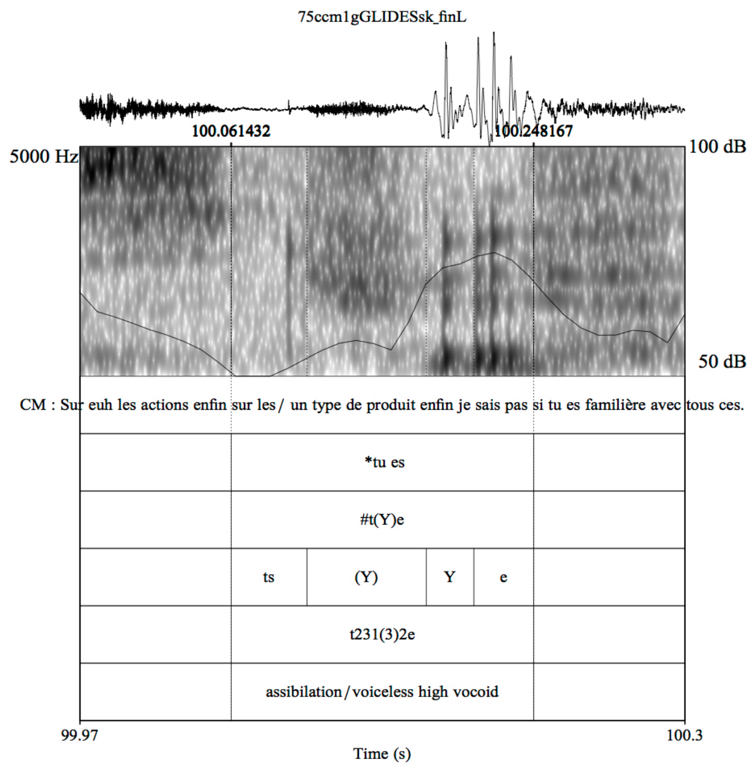


Figure 43: 75ccm1 *tu es* realised with syneresis, assibilation, and high vocoid devoicing

Syneresis as evinced by the single peak on the intensity contour, is corroborated by a single period of energy observed on the oscillogram. We note also that the voiceless period following [t] shows evidence of formants, albeit weak, as well as noise associated with assibilation at approximately 4500 Hz.

We now examine findings for all other tokens of derived cross-word-boundary contexts. Table 72 below shows variants observed for the remaining 83 cross-word-boundary contexts across five speakers.

Table 71: 75c *Paris*, Derived cross-word-boundary contexts: N= number of tokens, % calculated by speaker for each variant: deleted high vocoids DEL, dieresis, and syneresis

	Speaker										Total	
	75cab1		75ccm1		75ccr1		75csb1		75cvl1			
	N	%	N	%	N	%	N	%	N	%	N	%
DEL	--	--	1	3.8	1	0.2	--	--	--	--	**2**	**2.4**
dieresis	5	55.5	7	26.9	1	0.2	10	43.4	3	15.0	**26**	**31.3**
syneresis	4	44.4	18	69.2	3	0.6	13	56.5	17	85.0	**55**	**66.3**
Grand total	**9**	**10.9**	**26**	**31.3**	**5**	**6.0**	**23**	**100**	**20**	**27.7**	**83**	**100**

For all remaining cross-word-boundary tokens, syneresis is most prevalent overall (66.3%) with 31.3% of tokens realised with dieresis and 2.4% show a deleted high vocoid produced by two speakers only.

The data from the *Paris, Centre ville* (75c) survey show that variability is most evident for derived data; where lexicalised data exhibit variation of form, these involve exclusively instances of high vocoid deletion. Derived tokens show a much greater degree of variability which may be conditioned by an individual speaker or the high vocoid involved. Further study with a balanced data set would be required in order to say for sure.

We turn now, to results for *Brunoy* (91a), the second survey from *Île de France*. We begin our discussion with an examination of variability for all tokens across three high vocoids shown in Table 73 below.

Table 72: *Île de France*, *Brunoy* (91a), distribution of all tokens across three high vocoids for all potential variants. N= number of tokens, % calculated by vocoid for each variant: deleted high vocoid (DEL), dieresis, and syneresis

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	2	0.7	3	1.7	12	11.5	17	2.9
dieresis	28	9.1	4	2.3	8	7.7	40	6.8
syneresis	279	90.3	171	96.1	84	80.8	534	90.4
Grand Total	**309**	**52.3**	**178**	**30.1**	**104**	**17.6**	**591**	**100**

Across all tokens and for all high vocoids, we observe that syneresis is favoured at 90.4% with 6.8% dieresis and 2.9% of tokens showing a deleted high vocoid. Patterns of distribution across high vocoids are consistent with patterns observed for *Paris*: 53.9% /I/ followed by 26.7% /U/ and finally 19.4% /Y/. All three high vocoids show deletion although /Y/ deletion is most prevalent: 11.5% /Y/ compared to 1.7% /U/ and 0.7% /I/.

We turn now the distribution of tokens for lexicalised contexts. For ease of presentation, the table below shows token realisations for each of four female speakers.

Table 73: *Île de France*, *Brunoy* (91a), distrbution of lexicalised tokens for all potential variants across four female speakers. N= number of tokens, % calculated by vocoid for each variant: deleted high vocoids (DEL), dieresis, and syneresis

	Female speaker								Total	
	91aal1		91aed1		91ael1		91asl1			
	N	%	N	%	N	%	N	%	N	%
DEL	1	1.9	2	3.2	2	3.4	2	2.5	7	2.7
dieresis	--	--	2	3.2	2	3.4	--	--	4	1.6
syneresis	52	98.1	59	93.7	55	93.2	79	97.5	245	95.7
Grand Total	**53**	**20.7**	**63**	**24.6**	**59**	**23.0**	**81**	**31.3**	**256**	**100**

Table 74 shows that 256 lexicalised tokens for female speakers privilege syneresis (95.7%). These data are only slightly more variable than is observed for Paris where syneresis is observed in 97.7% of tokens and the remaining tokens are realised with a deleted high vocoid (2.3%). Lexicalised data in Brunoy show dieresis (1.6%) and high vocoid deletion (2.7%) much more evident than in Paris and they involve all three high vocoids.

In Table 75, we examine variability of tokens for lexicalised contexts across four male speakers.

Table 74: *Île de France*, *Brunoy* (91a), distribution of lexicalised tokens for all potential variants across four male speakers, N= number of tokens, % calculated by vocoid for each variant: deleted high vocoids (DEL), dieresis, and syneresis

	Male Speakers								Total	
	91acs2		91adb1		91ajc1		91amb1			
	N	%	N	%	N	%	N	%	N	%
DEL	--	--	2	4.3	2	2.5	4	6.6	8	3.1
dieresis	--	--	--	--	1	1.2	5	8.2	6	2.3
syn	71	100	45	95.7	78	96.3	52	85.2	246	94.6
Grand Total	**71**	**27.3**	**47**	**18.1**	**81**	**31.2**	**61**	**23.5**	**260**	**100**

Distribution of tokens for lexicalised contexts pattern similarly to tokens observed for female speakers shown above. Syneresis is privileged (94.5%) followed by deleted high vocoid (3.1%) and then dieresis (2.3%). Speaker cs2 (M, 22 yrs.) shows categorical syneresis, speaker db1 (M, 54 yrs.) privileges syneresis (95.7%) with the remaining two tokens (4.3%) realised with deleted high vocoid. Speakers jc1 (M, 63 yrs.) and mb1 (M, 62 yrs.) show more propensity towards variability. Thus supporting the observation that the potential for variable realisation of lexicalised tokens is related to the individual speaker. This does not preclude, however, phonological environment as a factor. An examination of overall lexicalised tokens for all speakers across high vocoids shows that variability of lexicalised contexts is potentially conditioned by the high vocoid involved. Table 76 below shows findings for all lexicalised contexts across high vocoids.

Table 75: *Île de France*, *Brunoy* (91a), distribution of all lexicalised tokens across three high vocoids for all variants, N= number of tokens, % calculated by vocoid for each variant

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	1	0.4	3	1.7	11	12.5	15	2.9
Dieresis	10	3.9	--	--	--	--	10	1.9
Syneresis	244	95.7	170	98.3	77	87.5	491	95.0
Grand Total	**255**	**49.4**	**173**	**33.5**	**88**	**17.0**	**516**	**100**

As we observed above, lexicalised tokens from Brunoy show slightly more variability than is observed for *Paris* (97.9% syneresis, 2.1% deletion, cf. Table 66 above). Data

from *Brunoy* show ten instances (2.1%) of dieresis involving exclusively the high vocoid /I/ and 15 instances of deleted high vocoid: /Y/ (12.5%. Distribution across high vocoids is consistent with data from Paris: /I/ 49.4%, /U/ 33.5%, /Y/ 17.0%. While all high vocoids privilege syneresis and only the high vocoid /I/ shows realisations for all variants; tokens involving other high vocoids /U/ and /Y/ are limited to deletion and syneresis. Two tokens for word medial glide contexts preceded by an OL sequence, show variable realisations: *qua<u>trième</u>* with dieresis [tʀijɛm], and *ta<u>blier</u>* with syneresis (after simplification of the OL sequence through liquid deletion [bje]). Note that both of these contexts involve a stem final OL thus we observe variability of a suffix: *-ième* [jɛm; ijɛm], *-ier* [ije; je].

We turn now to findings for derived contexts from *Brunoy*. Table 77 shows results for all derived /HV+V/ contexts.

Table 76: *Île de France*, 91a *Brunoy*, all derived /HV+V/ data: N= number of tokens, % calculated by vocoid for each variant, deleted high vocoids (DEL), dieresis with epenthesis, dieresis with hiatus and syneresis

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	1	1.9	--	--	1	6.3	2	2.7
dieresis	18	35.2	4	80.0	8	50.0	30	40.0
syneresis	35	67.3	1	20.0	7	43.8	43	57.3
Grand Total	**54**	**72.0**	**5**	**6.7**	**16**	**21.3**	**75**	**100**

Across high vocoids, /I/ is most prevalent of three high vocoids at 72% followed by /Y/ at 21.3% and /U/ at 6.7%. While syneresis is realised most often (57.3%), dieresis is well represented (40%). The high vocoid /I/ shows more syneresis (67.3%) with dieresis realised at 35.2%. High vocoids /Y/ and /U/ privilege dieresis over syneresis: /Y/ shows 50% dieresis and 23.8% syneresis and /U/ shows 80% dieresis over 20% syneresis.

Derived data include nine (12%) word medial tokens and 64 (85.3%) across-word-boundary tokens. Only two of the cross-word-boundary tokens (3.1%) involve the pronoun *tu* followed by a vowel initial verb. The nine word medial tokens show seven

realised with syneresis and two with dieresis. Eight of these contexts involve /I/ and a single context shows /Y/: *diminuée* is realised with syneresis [nɥe].

Next, we look at cross-word-boundary derived data. Table 78 below shows cross-word-boundary data for four female speakers from Brunoy.

Table 77: *Île de France*, *Brunoy* (91a), variants for cross-word-boundary derived /HV+V/ contexts across four female speakers, N= number of tokens, % calculated by vocoid for each variant, deleted high vocoids (DEL), dieresis, and syneresis

	Female Speakers (F)								Total	
	91aal1		91aed1		91ael1		91asl1			
VARIANT	N	%	N	%	N	%	N	%	N	%
dieresis	6	60.0	9	52.9	6	66.7	--	--	21	48.8
syneresis	4	40.0	8	47.1	3	33.3	7	100	22	51.2
Grand Total	**10**	**23.3**	**17**	**39.5**	**9**	**20.9**	**7**	**16.3**	**43**	**100**

Female speakers from *Brunoy* produce 43 tokens of cross-word-boundary derived contexts none of which involve the pronoun *tu*. Realisations with syneresis (51.2%) and dieresis (48.8%) are equally represented amongst female speakers with the exception of speaker sl1 (F, 64 yrs.) who produces categorical syneresis; all other speakers show a preference for dieresis which ranges from a slight preference (52.9%) to a clear majority 60-66.7%).

Next, we examine data for derived cross-word-boundary contexts for four male speakers from Brunoy. In the Table 79 below we show findings for variable realisations of cross-word-boundary derived contexts for each male speaker in *Brunoy* (91a).

Table 78: *Île de France*, *Brunoy* (91a), variants for cross-word-boundary derived /HV+V/ contexts across four male speakers, N= number of tokens, % calculated by vocoid for each variant: deleted high vocoids (DEL), dieresis, and syneresis

	Male Speakers (M)								Total	
	91acs2		91adb1		91ajc1		91amb1			
VARIANT	N	%	N	%	N	%	N	%	N	%
DEL	--	--	1	11.1	1	10.0	--	--	2	6.3
dieresis	--	--	2	22.2	2	20.0	5	71.4	9	28.2
syneresis	6	100	6	66.7	7	70.0	2	28.6	21	65.6
Grand Total	**6**	**18.8**	**9**	**28.1**	**10**	**31.3**	**7**	**21.9**	**32**	**100**

The data above include only two tokens for the context of a pronoun *tu* followed by a vowel initial verb. Overall these data show a pattern in realisations consistent with what we have seen in the *Paris* (75c) data where syneresis is privileged followed by dieresis then deleted high vocoid. These data differ considerably from what we have just seen for the female speakers of *Brunoy* who show a preference for dieresis (cf. table 78 above).

Across all contexts observed in *Île de France*, we have found mostly consistent patterns for realisations of HVV tokens. Lexicalised contexts show syneresis is most prevalent: *Brunoy* 95.0%, *Paris* 97.9%. All other variants occur only marginally: deleted high vocoids occur in 2.7% of tokens (*Paris*) and 2.3% of tokens (*Brunoy*), while dieresis is realised in 0% and1.6% of tokens. Derived contexts are equally consistent until we examine data from *Brunoy* across gender; female speakers show a preference individually for dieresis in realisations of cross-word-boundary derived tokens (cf. Table 69 above). Distributions across high vocoids vary depending on context. In all lexicalised glide contexts we find that /I/ is most prevalent (approximately 50%) followed by /U/ (approximately 30%) and finally /Y/ (17%). The derived tokens consistently show however that contexts involving the high vocoid /Y/ are consistently more frequent than contexts involving /U/.

6.5.2 *Île de France*: observations for two surveys

Data from *Île de France*, collected from 75c *Paris centre ville* (722 tokens) and 91a *Brunoy* (591 tokens) provide 1313 tokens overall across 13 speakers: eight females and

five males. Surveys are not balanced for numbers of speakers nor social factors. We observe considerable inter-speaker variability for each of these surveys such that derived data represent as low as 8% and as high as 22.5%. Nonetheless, 1313 tokens overall from *Île de France* show patterns of distribution across lexicalised and derived contexts (85% versus 15%) consistent with distributions observed across all data. Across each of three high vocoids, tokens for the high vocoid /I/ consistently represent the greatest proportion of data. Although lexicalised tokens show the expected distribution /I/ ≫ /U/ ≫ /Y/, we find consistently that derived tokens exhibit a different distributional hierarchy that privileges front high vocoids: /I/ ≫ /Y/ ≫ /U/.

When we examine these data for variable realisations, variability of lexicalised contexts is almost inexistent; slightly more prevalent in Brunoy: 95% syneresis followed by 1.9% dieresis and 2.9% deleted high vocoid, while Parisian data shows little to no variability: 97.7% syneresis, 7.3% deleted high vocoid. Derived contexts provide slightly more variable realisations: *Paris* shows 59% syneresis, 25.6% dieresis and 15.4% deleted high vocoids, and *Brunoy* 75 derived tokens 57.3% syneresis, 40% dieresis and 2.7% deleted high vocoids.

Variability may be related to the particular high vocoid involved, to the particular environment in which a token is found, or to the individual speaker. Unfortunately, the imbalance between speakers across surveys and the uncontrolled nature of spontaneous speech prevents us from making stronger conclusions. Nonetheless, we do observe that lexicalised contexts privilege syneresis: Paris 97.7%, Brunoy 90.4% while we find derived contexts somewhat more variable.

Across the various environments examined in derived data (medial, across-word-boundary) as well as preceding phonological environments (simple versus complex onset), we observe variability, however, we are once again limited by the nature of our spontaneous speech data. Word medial derived tokens are too few to make any significant observations (*Paris* 12, *Brunoy* nine). The medial tokens observed privilege syneresis, while some tokens show other variants as a possible realisation (deleted high vocoid, and dieresis). More study, giving consideration to social factors as well as

phonological environments (preceding onset, phrase structure, and intonation patterns), is required. Glide formation across word boundaries, while acknowledged in some descriptions, is not widely discussed. We recognise that many of the environments presented (/*tu* + V/, for example) may in fact present some grammaticalised forms (*tu* reduced to *t'a*, for example) for an individual speaker and this may vary across speakers. Further study would be required to understand the factors involved in such variability.

In the following sections, we examine data for the region of *Normandie* in two PFC surveys: 50a *Brécey* and 61a *Domfrontais*. In section 6.7, we examine the overall distribution of tokens as well as demographic information for speakers of each survey.

6.6 Normandie

The region of *Normandie* is located in the northwest of France and has a population of a little more than 5% of the country. Data from two PFC surveys: 50a *Brécey* and 61a *Domfrontais,* represent 30.8% of all tokens in our study of glide contexts in spontaneous speech. These rural areas of lower Normandy can be contrasted to the urban centres of Île de France and Rhône Alpes. Brécey is a town (*commune*) of just more than 2,000 inhabitants while Domfront is a *commune* in the Department of Orne with a population of just under 4,000 at the time of this survey[32]. In this section we will examine the distribution of glide contexts across surveys, by speakers and across vocoids. The table below shows distribution of derived and lexicalised glide tokens across the two surveys from *Normandie*.

[32] Source: http://www.insee.fr/

Table 79: *Normandie*, all tokens, N= number, % calculated by context for each survey

	PFC Survey				Total	
	50a *Brécey*		61a *Domfrontais*			
	N	%	N	%	N	%
Derived	86	15.8	82	16.2	168	16.0
Lexicalised	460	84.2	423	83.8	883	84.0
Grand Total	**546**	**52.0**	**505**	**48.0**	**1051**	**100**

Distribution for lexicalised and derived data is consistent with patterns observed for *Île de France* (lexicalised 84% compared to derived at 16%). Numbers of tokens across two surveys is largely balanced (Brécey 52% and Domfrontais 48%).

Four speakers are retained for each of the two surveys. Table 81 below shows demographic information for each speaker from *Brécey*.

Table 80: *Normandie*, *Brécey* (50a), age, sex, N= number of tokens, % calculated by speaker

PFC speaker code	**Age**	Total tokens	
Female (F)		N	%
50alb1	55	153	52.4
50app1	69	139	47.6
Total F		**292**	**53.5**
Male (M)		N	%
50ajp1	22	70	27.6
50apb1	54	184	72.4
Total M		**254**	**46.5**
Total *Brecey*		**546**	**52.0**
Normandie		**1051**	**100**

The four speakers retained for *Brécey* (50a) are balanced for sex including two female and two male speakers. However, age is not balanced across both sexes; female speakers are clustered at the older end of the spectrum while male speakers represent young (20 yrs.) and middle aged (69 yrs.). Grouping all speakers shows a range in ages from 22 years to 69 years.

In Table 82 below, we show distribution of lexicalised and derived contexts across three high vocoids.

Table 81: ***Normandie,* 50a *Brécey*, all tokens,** N = number, % calculated by context across three high vocoids

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
Derived	71	23.0	6	4.1	9	10.1	86	15.8
Lexicalised	238	77.0	142	95.9	80	89.9	460	84.7
Grand Total	**309**	**56.6**	**148**	**27.1**	**89**	**16.3**	**546**	**100**

Token distribution across high vocoids shows /I/ ≫ /Y/ ≫ /U/ for derived contexts but /I/ ≫ /U/ ≫ /Y/ for lexicalised contexts.

The table below shows the distribution of lexicalised and derived contexts for all tokens across all speakers retained from Brécey.

Table 82: ***Normandie, 50a Brécey* all tokens for derived and lexicalised contexts** N= number of tokens, % calculated by speaker

	Speaker								Total	
	50ajp1		50alb1		50apb1		50app1			
	N	%	N	%	N	%	N	%	N	%
Derived	7	10.0	27	17.6	32	17.4	20	14.4	86	15.8
Lexicalised	63	90.0	126	82.4	152	82.6	119	85.6	460	84.2
Grand Total	**70**	**12.8**	**153**	**28.0**	**184**	**33.7**	**139**	**25.5**	**546**	**100**

Overall, distribution for derived and lexicalised contexts in Brécey is consistent with patterns observed for surveys in *Île de France* (cf. tables 77 and 78, above): 84.2% lexicalised and 15.8% derived. Distribution varies moderately from speaker to speaker such that we observe derived tokens representing as little as 10% and as much as 17.6%.

Table 84 below shows demographic information for four speakers retained for the 61a *Domfrontais* survey.

Table 83: ***Normandie*, 61a *Domfrontais*,** age, sex and N= number of tokens, % calculated by speaker

PFC speaker code	Age	Total tokens	
Female (F)		N	%
61agr1	55	126	40.0
61ajh1	48	189	60.0
Total Female		**315**	**62.4**
Male (M)		N	%
61ahd1	83	91	47.9
61ahl1	20	99	52.1
Total Male		**190**	**37.6**
Total *Domfrontais*		**505**	**48.0**
Normandie		**1051**	**100**

While ages of speakers range from 20 years to 83 years across participants, as we observed for Brécey, only four speakers do not allow for balance across surveys; two female participants are middle aged (below 60), while two male participants represent two extremes of the spectrum: under 30 and older than 80. Tokens across gender are not well balanced: female speakers contribute 62.4% of tokens while male speakers contribute 37.6%.

In the table below, we include all four speakers marked for sex (F/M).

Table 84: *Normandie, 61a Domfrontais* all tokens for derived and lexicalised contexts across four speakers, N= number of tokens, % calculated by speaker

	Speaker								Total	
	61agr1 (F)		61ahd1 (F)		61ahl1 (M)		61ajh1 (M)			
	N	%	N	%	N	%	N	%	N	%
Derived	38	30.2	8	8.8	12	12.1	24	12.7	82	16.2
Lexicalised	88	69.8	83	91.2	87	87.9	165	87.3	423	83.8
Grand Total	**126**	**25.0**	**91**	**18.0**	**99**	**19.6**	**189**	**37.4**	**505**	**100**

Table 85 above shows that while overall distribution conforms to the patterns observed across all other surveys to this point across, individual speakers vary. Derived distributions range as high as 30.2% (gr1) to as low as 8.8% (hd1) while the other speakers are more constant (12.1% and 12.7%).

Table 86 below shows distribution of lexicalised and derived tokens across three high vocoids. Again, we observe that distribution of tokens across high vocoids varies between contexts: lexicalised I ≫ U ≫ Y versus derived I ≫ Y ≫ U.

Table 85: *Normandie,* 61a *Domfrontais* all tokens across high vocoids for lexicalised and derived contexts, N = number, calculated for each context (derived and lexicalised) across three high vocoids

	Vocoid						Total	
	I		U		Y			
Context	N	%	N	%	N	%	N	%
Derived	71	26.8	5	2.9	6	8.8	82	16.2
Lexicalised	194	73.2	167	97.1	62	91.2	423	83.8
Grand Total	**265**	**52.5**	**172**	**34.1**	**68**	**13.5**	**505**	**100**

The data show a familiar pattern of distribution across lexicalised (83.8%) and derived (16.2%) glide contexts consistent with patterns exhibited in *Île de France*. The high unrounded vocoid /I/ is privileged in both lexicalised and derived glide contexts but the two rounded vocoids appear to be in competition such that /U/ is privileged over /Y/ across lexicalised contexts, however, /Y/ is privileged over /U/ across derived contexts.

In the following section we examine the variability observed for realisations of lexicalised and derived tokens for two surveys in *Normandie*.

6.6.1 *Normandie*: variable realisations across lexicalised and derived data

In this section we examine in lexicalised contexts and derived contexts all variants: deleted high vocoid (DEL), dieresis and syneresis, across individual speakers and across high vocoids.

Table 87 below shows all realisations for tokens in lexicalised contexts across three high vocoids for the survey 50a *Brécey*.

Table 86: *Normandie*, *Brécey* (50 a), all lexicalised GV contexts across three high vocoids for each variant (DEL= deleted high vocoid, Dieresis and Syneresis) N= number of tokens, % calculated by variant across high vocoid

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	--	--	1	0.7	5	6.3	6	1.3
dieresis	4	1.7	--	--	--	--	4	0.9
syneresis	234	98.3	141	99.3	75	93.7	450	97.8
Grand Total	**238**	**51.7**	**142**	**30.9**	**80**	**17.4**	**460**	**100**

Lexicalised data from *Brécey* show that syneresis is privileged (97.8%). The remaining 2.2% is split between realisations with a deleted high vocoid (1.3%) and realisations with dieresis (0.9%). Tokens for a deleted high vocoid show five instances of deleted /Y/ involving the word *puis* [pɥi] reduced to [pi] while the single token for /U/ is less expected and involves the word *armoire* [aRmwaR] reduced to [aRmaR]. Note that all other variation involves the high vocoid /I/ only: four tokens realised with dieresis, 234 with syneresis. Since we have retained only four speakers (two male and two female) in the *Brécey* sample, we will examine their production together in a single table. The sex of each speaker is identified with speaker identification. In Table 88 we examine findings for all lexicalised contexts across four speakers.

Table 87: *Normandie*, *Brécey* (50a), realisations for all lexicalised contexts across four speakers for each variant (DEL= deleted high vocoid, Dieresis and Syneresis)**,** N= number of tokens, % calculated by variant across speaker

	Speaker (Gender)								Total	
	50ajp1 (M)		50alb1 (F)		50apb1 (M)		50app1 (F)			
	N	%	N	%	N	%	N	%	N	%
DEL	--	--	3	2.4	2	1.3	1	0.8	6	1.3
dieresis	--	--	1	0.8	1	0.7	2	1.7	4	0.9
syneresis	63	100	122	96.8	149	98.0	116	97.5	450	97.8
Grand Total	**63**	**13.7**	**126**	**27.4**	**152**	**33.0**	**119**	**25.9**	**460**	**100**

The data show that variability of lexicalised contexts is minimal at 2.2% showing a slight preference for deleted high vocoids over dieresis. Syneresis is most prevalent at 97.8%. Three tokens realised with dieresis are instances of the word *clientèle* [klijɑ̃tɛl] from two

different speakers (50apb1 M 54 yrs., 50app1 F 69 yrs.). Dieresis is expected in this context given the presence of OL in the preceding environment. Finally, we find a single token for the word *sixième* realised with dieresis [sizijɛm]. This finding suggests more variability of the suffix *–ième* [jɛm: i(j)ɛm] than is described.

Figure 44 below shows the token *sixième* realised with dieresis. Our analysis concerns the realisation of the final portion (suffix) of the token. Analysis of the intensity contour is improved with analysis of the oscillogram above.

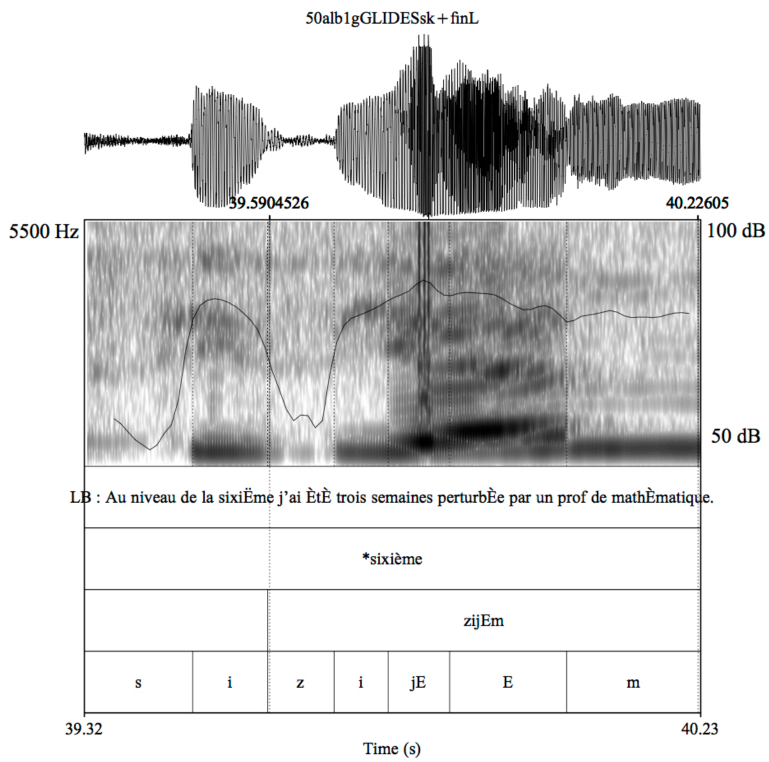


Figure 44: 50alb1 *sixième* realised with dieresis with four tiers: transcription, token, SAMPA, segmentation.

The two peaks of the intensity contour are corroborated by two peaks observed on the oscillogram. The spectrogram leaves no doubt as to the production of two distinct vowels; it is less clear whether this token is realised with hiatus (no glide formation) or epenthesis (glide formation). This speaker is clearly emphasising the final syllable which

may affect perception of epenthetic glide which might be the object of further study in this area.

We turn now to an examination of derived contexts in the data from 50a *Brécey*. Among these derived data, we find 19 instances of word medial derived contexts that show all three high vocoids: /I/ in the forms *mariage, ouvrier,* /Y/ in the participles *continué,* and *habitué* and /U/ in the verb *jouer/ait.* Excluding two tokens for the noun *ouvrier(s),* which present a stem final OL sequence and are realised as with dieresis, word medial derived tokens for all high vocoids are realised with categorical syneresis. In the following table, we examine findings for all derived cross-word-boundary contexts.

Table 88 shows realisations for derived contexts across four speakers.

Table 88: *Normandie*, *Brécey* (50a), realisations for all derived cross-word boundary contexts across four speakers, N= number of tokens, % calculated by variant (DEL= deleted high vocoid, Dieresis and Syneresis), across speakers

	Speaker (Gender)								Total	
	50ajp1 (M)		50alb1 (F)		50apb1 (M)		50app1 (F)			
	N	%	N	%	N	%	N	%	N	%
DEL	--	--	1	4.3	3	15.0	1	5.6	5	7.5
dieresis	3	50.0	10	43.5	2	10.0	9	50.0	24	35.8
syneresis	3	50.0	12	52.2	15	75.0	8	44.4	38	56.7
Grand Total	**6**	**9.0**	**23**	**34.3**	**20**	**29.9**	**18**	**26.9**	**67**	100

As we have observed for other surveys, variability of derived contexts patterns differently than for lexicalised contexts. While syneresis is most prevalent it represents a smaller proportion of realisations and cross speaker variation is more evident. Overall, syneresis represents 56.7% of cross-word-boundary derived token realisations however individual speakers range from as low as 44% to as high as 75%. Dieresis is more prevalent in these data averaging 35.8% while deleted high vocoids represent 7.5%. With the exception of a single speaker (pb1, M, 54 yrs.), rates of syneresis versus dieresis are nearly balanced: dieresis ranges from 43.5-50% and syneresis ranges from 44.4-52.2%.

Table 90 shows realisations for all derived contexts across high vocoids.

Table 89: ***Normandie,*** **50a** ***Brécey*****, realisations for all derived /HV+V/ contexts,** N= number of tokens, % calculated by high vocoid for each variant (DEL= deleted high vocoid, Dieresis and Syneresis), across high vocoids

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	5	7.0	--	--	--	--	5	5.8
dieresis	19	26.8	3	50.0	4	44.4	26	30.2
syneresis	47	66.2	3	50.0	5	55.6	55	64.0
Grand Total	**71**	**82.6**	**6**	**7.0**	**9**	**10.5**	**86**	**100**

Derived tokens show a familiar pattern of distribution across three high vocoids: /I/ (82.6%) is most prevalent followed by /Y/ (10.5%) then /U/ (7.0%). Contexts involving /I/ deletion are nearly identical cross-word-boundary contexts: *qui est* [ke], *qui était* [ketɛ], *qui ont* [kɔ̃], produced by three different speakers (two female, one male). Only the high vocoid /I/ shows deletion whereas the other high vocoids /U/ and /Y/ show near parity of realisations.

We turn now to an examination of token realisation for the *Domfrontais* (61a) survey. We begin with lexicalised data across speaker then we examine the distribution of lexicalised contexts across high vocoids. Table 91 below shows realisations observed for all lexicalised (GV) contexts for four speakers. Since we have only four speakers (two male and two female) we examine their production together in a single table. The sex of each speaker is identified with speaker identification.

Table 90: ***Normandie*****,** ***Domfrontais*** **(61a), all lexicalised contexts across four speakers**: N= number of tokens, % calculated by variant (DEL= deleted high vocoid, Dieresis and Syneresis) across speaker

	Speaker									
	61agr1 (F)		61ahd1 (M)		61ahl1 (M)		61ajh1 (F)		Total	
	N	%	N	%	N	%	N	%	N	%
DEL	18	20.5	3	3.6	4	4.6	2	1.2	27	6.4
dieresis	--	--	--	--	--	--	1	0.6	1	0.2
syneresis	70	79.5	80	96.4	83	95.4	162	98.2	395	93.4
Grand Total	**88**	**20.8**	**83**	**19.6**	**87**	**20.6**	**165**	**39.0**	**423**	**100**

Lexicalised contexts show a clear predilection for realisation with syneresis found in 93.4% of tokens. Variation of lexicalised contexts is minimal (6.6%) involving primarily deleted high vocoids with a single token of dieresis for the token *clients* [klijã] which, with the preceding OL sequence, is entirely expected. 27 instances of deleted high vocoid is notably high which can be attributed to a single speaker who produces 15 tokens of the word *puis* [pɥi] reduced to [pi].

We turn now to the distribution of tokens for all lexicalised contexts across high vocoids. In Table 92, we show distribution of tokens across three high vocoids for all variants. We will discuss deletion further in the following section.

Table 91: *Normandie*, *Domfrontais* (61a), all lexicalised (GV) contexts across three high vocoids: N= number of tokens, % calculated by high vocoid for each variant (DEL= deleted high vocoid, Dieresis and Syneresis)

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	--	--	2	1.2	25	40.3	27	6.4
dieresis	1	0.5	--	--	--	--	1	0.2
syneresis	193	99.5	165	98.8	37	59.7	395	93.4
Grand Total	**194**	**45.9**	**167**	**39.5**	**62**	**14.7**	**423**	**100**

As we have seen for other surveys, lexicalised contexts show a familiar pattern of distribution which is evident in these data: contexts involving /I/ are most prevalent (45.9%) followed by /U/ (39.5%) and finally /Y/ (14.7%). Of 27 instances of deleted high vocoid, 25 involve reduction of the word *puis* [pɥi] to [pi]. A single instance of deletion in the word *suis* (*je me suis* [si], for example), and two instances of deleted high vocoid /U/ involve the reduction of the lexicalised diphthong [wa] when preceded by OL sequences: *froid*, *exploitation* are produced by one speaker: gr1, F, 74.

We turn now to a discussion of all tokens for derived contexts. We begin with a look at all derived contexts. Table 93 shows the realisations observed for all derived contexts across three high vocoids.

Table 92: ***Normandie*****,** ***Domfrontais*** **(50a), all derived /HV+V/ contexts across three high vocoids:** N= number of tokens, % calculated for each variant, (DEL= deleted high vocoid, Dieresis and Syneresis)

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	12	16.9	--	--	--	--	12	14.6
Dieresis	18	25.4	2	40.0	3	50.0	23	28
Syneresis	41	57.7	3	60.0	3	50.0	47	57.3
Grand Total	**71**	**86.6**	**5**	**6.1**	**6**	**7.3**	**82**	**100**

Derived contexts consist of 84.2% cross-word-boundary contexts and 15.8% word medial contexts. Word medial contexts show categorical syneresis across all three high vocoids observed in the forms *marié(e)* (five tokens), *mariage* (three tokens) and single tokens for each of *expédié, tuaient,* and *secouer*. Overall, distribution across three high vocoids shows 69.2% /I/, 23.1% /Y/ and 7.7% /U/. Syneresis is most prevalent (57.3%) but these findings are not nearly as systematic as is observed for lexicalised contexts. Across all contexts we find that 42.7% of tokens realised as one of the other variants: 28% dieresis and 14.6% high vocoid deletion. Twelve tokens show deleted high vocoid /I/ in near identical cross-word-boundary contexts: *qui est* [ke], *qui ont* [kɔ̃], *qui aurait* [ko], *qui a(llions)* [ka].

We examine, next, all derived cross-word-boundary contexts across speakers. In Table 94 we show findings for this context across four speakers.

Table 93: ***Normandie*****,** ***Domfrontais*** **(61a), all cross-word-boundary derived contexts across four speakers:** N= number of tokens, % calculated by speaker for each variant, (DEL= deleted high vocoid, Dieresis and Syneresis)

	Speaker								Total	
	61agr1 (F)		61ahd1 (M)		61ahl1 (M)		61ajh1 (F)			
	N	%	N	%	N	%	N	%	N	%
DEL	7	24.1	2	40.0	--	--	3	12.5	12	17.4
dieresis	8	27.6	1	20.0	3	27.3	11	45.9	23	33.3
syneresis	14	48.3	2	40.0	8	72.7	10	41.7	34	49.3
Grand Total	**29**	42.0	**5**	**7.2**	**11**	**15.9**	**24**	**34.8**	**69**	**100**

As we examine findings for derived cross-word-boundary contexts, we observe that in cross-word-boundary contexts syneresis, though privileged (49.3%), does not show the same strong preference for syneresis that is observed in other environments. Across speakers the rate of realisation varies considerably: syneresis dips as low as 40% and while speaker jh1 shows near parity between syneresis (41.7%) and dieresis (45.9%), this is the only case where dieresis dominates. Most interesting, findings for speaker hl1, who shows no tokens for high vocoid deletion, exhibits 72.7% syneresis and 27.3% dieresis. We observe no instances of the pronoun *tu* followed by a vowel initial verb among these data. Table 94 shows distribution of cross-word-boundary derived tokens across three high vocoids.

Table 94: *Normandie*, *Domfrontais* (50a), variants for derived cross-word-boundary contexts, N= number of tokens, % calculated for each variant, (DEL) deleted high vocoids, dieresis, and syneresis

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	12	19.4	--	--	--	--	12	17.4
dieresis	18	29	2	50.0	3	100	23	33.3
syneresis	32	51.6	2	50.0	--	--	34	49.3
Grand Total	**62**	89.9	**4**	5.8	**3**	4.3	**69**	100

The data show that 69 derived cross-word-boundary contexts involve /I/ predominantly (89.9%). Tokens for /I/ exhibit all possible variants while the other two high vocoids involved in the cross-word-boundary context show dieresis more prominently; /U/ shows parity between syneresis and dieresis while /Y/ shows categorical dieresis. Four tokens for /U/ include *où on* realised with dieresis by jh1 then we find the tokens *où ils* and *où on* realised with syneresis by speakers jh1 and hd1. Three tokens for /Y/ realised with dieresis are produced by a single speaker jh1 (who produces five of seven tokens for rounded high vocoids).

We turn now to observations for data from the two surveys from the Normandy region. In the following section we provide a summary of findings across the two surveys in our sample.

6.6.2 *Normandie*: observations across two surveys

Data from the region of Normandy comes from a sample of two PFC surveys: 50a Beécey (546) and Domfrontais (505 tokens), providing 1051 (30.8%) tokens overall across eight speakers (four women and four men). While these two surveys are balanced for age and sex, four speakers from each survey are too few to allow for any generalised observations. Despite a smaller sample size, patterns in the distribution of lexicalised and derived contexts (roughly 85% to 15%) and high vocoids (/I/ ≫ /U/ ≫ /Y/) are consistent with findings for *Île de France*. For each of the two surveys from *Normandie*, we find very comparable data follow similar patterns of distribution. As elsewhere, lexicalised data are most prevalent at 84% and shows a familiar pattern of distribution across high vocoids: /I/ ≫ /U/ ≫ /Y/ compared to derived tokens (16%) which exhibit a slightly different pattern: /I/ ≫ /Y/ ≫ /U/ (cf. section 6.3.2).

Variability of realisations (syneresis versus dieresis and high vocoid deletion) also exhibits distinct patterns of behavior across lexicalised and derived contexts. Syneresis is privileged in both lexicalised and derived contexts, however, the proportion of tokens realised with syneresis differs across the two contexts. Derived contexts average 60.7% syneresis with 29.2% dieresis while lexicalised contexts show 95.6% syneresis and only 0.6% dieresis. Lexicalised tokens (1.7%) that exhibit dieresis are preceded by an OL cluster (*clientèle*, for example), while a single token for the word *sixième* is realised with dieresis demonstrating variability for the suffix *–ième*. High vocoid deletion is most prevalent in lexicalised contexts involving only the two rounded high vocoids which are seen as part of the true diphthongs: [ɥi] or [wa]. Evidence for deleted high vocoids amongst lexicalised data is variable across speakers pointing to the possibility that high vocoid deletion may be conditioned by the individual speaker and the phonological environment presented. Some tokens may be a result of a generalised reduction for an individual speaker (*puis* reduced to *pi*, for example), while other isolated tokens such as the deletion of [w] in a single token of *armoire* may be isolated instances attributed to spontaneous speech. In any case, deletions in lexicalised contexts favour the high vocoid /Y/. In derived contexts, tokens for syneresis are consistently preferred but at a lower rate: approximately 55% across surveys. Dieresis is much more common among derived

tokens than is observed across lexicalised tokens. Finally the rate of deleted high vocoids is considerably higher among derived cross-word-boundary and favours the high vocoid /I/ (*qui est, for example*).

We turn now to an examination of data for the *Rhône Alpes* region our third region of study.

6.7 Rhône Alpes

The *Rhône Alpes*, a region located to the south of Paris on the eastern border of France, is the second most populous region next to *Île de France*. The two PFC surveys sampled here are from two major urban centres in the region: 38a *Grenoble* (560,453 inhabitants) and 69a *Lyon* (1,798,395 inhabitants) in the region[33]. In this regard these survey locations are quite comparable to *Île de France* while quite distinct from *Normandie*.

Table 96 below shows results for each survey with total region.

Table 95: *Rhône Alpes*, all tokens for across two surveys (38a *Grenoble*, 69a *Lyon*), N= number of tokens, % calculated by context for each survey

	Rhône Alpes				Total Region	
	38a *Grenoble*		69a *Lyon*			
	N	%	N	%	N	%
Lexicalised	463	86.5	280	89.5	743	87.6
Derived	72	13.5	33	10.5	105	12.4
Total survey	**535**	**63.1**	**313**	**36.9**	**848**	**100**

[33] Source: http://www.insee.fr/

We find 848 tokens for ten speakers across two surveys for the *Rhône Alpes* region which represents 26.4% overall tokens. Consistent with patterns observed across regions, tokens for lexicalised contexts represent 87.7% while tokens for derived contexts show only 12.3% for the region which is relatively constant across regions. Demographic information for four speakers retained from the *Lyon* survey is shown below.

Table 96: *Rhône Alpes*, 69a *Lyon*, age, sex and N= number of tokens, % calculated by speaker

PFC speaker code	**Age**	Total tokens	
Female		N	
69ajd1	46	97	61.8
69all1	74	60	38.2
Total F		**157**	**50.2**
Male		N	
69aag1	51	73	46.8
69asg1	25	83	53.2
Total M		**156**	**49.8**
Total *Lyon*		**313**	**37.0**
Rhône Alpes		**848**	**100**

As we saw for the surveys in *Normandie*, four speakers, balanced for sex, are not balanced for age. Nonetheless, a full range of ages is observed across speakers (25 to 74 yrs.).

In Table 98 below, we examine distribution for all tokens of lexicalised and derived contexts across high vocoids.

Table 97: *Rhône Alpes*, 69a *Lyon*, all tokens by context across three high vocoids, N= number of tokens, % calculated by vocoid for lexicalised and derived contexts

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
Derived	17	11.4	--	--	16	24.2	33	10.5
Lexicalised	132	88.6	98	100	50	75.8	280	89.5
Grand Total	**149**	**47.6**	**98**	**31.3**	**66**	**21.1**	**313**	**100**

Perhaps the result of the low N, data from *Lyon* show a distribution that highly favours lexicalised contexts (89.5%) over derived contexts (10.5%). Nonetheless, distribution across high vocoids shows the same patterns observed elsewhere wherein overall distribution of contexts show /I/ is most prevalent (49.2%) followed by /U/ (32.3%) and then /Y/ (18.5%). Tokens for derived contexts show the same pattern of distribution across high vocoids that privileges the front rounded high vocoid /Y/ over the back rounded such that with such low N we find no tokens for the high vocoid /U/.

Let us turn now to the demographic information for speakers retained from the PFC 38a *Grenoble* survey. Table 99 shows information for six speakers retained for our study.

Table 98: *Rhône Alpes*, 38a *Grenoble*, age, sex and N= number of tokens, % calculated by speaker

PFC speaker code	**Age**	Total tokens	
Female		N	%
38aca1	53	94	28.5
38aep1	33	72	21.8
38agp1	66	52	15.8
38amb1	22	111	33.7
Total F		**329**	**61.6**
Male		N	%
38aas1	68	124	60.5
38aym1	28	81	39.5
Total M		**205**	**38.4**
Total *Grenoble*		**534**	**61.9**
Rhône Alpes		**863**	**100**

We have retained four female and two male speakers that combined represent three age groups; ages range from 22 to 68 years of age. Amongst our male speakers we have included one speaker aged 28 yrs. and another aged 68 yrs.; female speakers straddle three age groups: one speaker under 30, and two over 30 but under 60 and one over 60.

In Table 100 we examine distribution of tokens for all contexts across three high vocoids in *Grenoble*.

Table 99: *Rhône Alpes*, 38a *Grenoble* all tokens for lexicalised and derived contexts across three high vocoids, N= number of tokens, % calculated for derived and lexicalised contexts across three high vocoids

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
Derived	55	20.5	2	1.2	15	15.8	72	13.5
Lexicalised	213	79.5	169	98.8	80	84.2	462	86.5
Grand Total	**268**	**50.2**	**171**	**32.0**	**95**	**17.8**	**534**	**100**

Distribution for lexicalised (86.5%) and derived (13.5%) contexts show the same patterns we have seen across all other surveys. The high vocoid /I/ is most prevalent for both derived (76.3) and lexicalised (46.1) contexts. The rounded high vocoids (front and back) show the pattern /Y/ ≫ /U for derived data but /U/ ≫ /Y/ for lexicalised. We discuss these patterns for distribution of High Vocoids in the next section.

In the following section we examine all realisations observed in lexicalised and derived contexts for each *Rhône Alpes* survey.

6.7.1 *Rhône Alpes*: variation across lexicalised and derived data

In this section, we examine the distribution of tokens for various realisations: deleted high vocoid (DEL), dieresis and syneresis, across speakers for each of two surveys from the Rhône Alpes region. We begin our discussion with data from the PFC survey for 69a Lyon. In Table 101 below we show realisations for all tokens of lexicalised contexts across three high vocoids from *Lyon*.

Table 100: *Rhône Alpes*, 69a *Lyon*, realisations for all lexicalised contexts across high vocoids for three variant: N= number of tokens, % calculated for each variant, (DEL= deleted high vocoid, Dieresis and Syneresis), across three high vocoids

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	1	0.8	--	--	1	2.0	2	0.7
Dieresis	5	3.8	1	1.0	--	--	6	2.2
Syneresis	126	95.5	97	99.0	49	98.0	272	97.1
Grand Total	**132**	**39.8**	**98**	**35.0**	**50**	**17.9**	**280**	**100**

The data show that lexicalised contexts privilege syneresis (97.1%) alternate realisations include dieresis 2.2% and deleted high vocoids 0.7% overall. While variability is minimal (2.5% overall) all three high vocoids are involved. /I/ is most prevalent 75% among non syneresis variants with /U/ and /Y/ representing only 25% each. Two tokens for deleted high vocoids include deleted /I/ in *gracieuse* [søz] and deleted /Y/ *puis* [pi]. Dieresis is observed in the proper nouns *Lyon*, *Fourvière* /I/ and *Louis* /U/.

In Table 102 we examine distribution of realisations for all lexicalised contexts across four speakers.

Table 101: *Rhône Alpes*, *Lyon* (69a), realisations for all lexicalised contexts across four speakers, N= number of tokens, % calculated by variant across speaker

	Speaker								Total	
	69aag1 (M)		69ajd1 (F)		69all1 (F)		69asg1 (M)			
	N	%	N	%	N	%	N	%	N	%
DEL	--	--	--	--	--	--	2	2.9	2	0.7
dieresis	2	3.0	3	3.3	1	1.9	--	--	6	2.1
syneresis	65	97.0	87	96.7	53	98.1	67	97.1	272	97.1
Grand Total	**67**	**23.9**	**90**	**32.1**	**54**	**19.3**	**69**	**24.6**	**280**	**100**

Our findings show that across four speakers syneresis is highly privileged for lexicalised contexts. Only one speaker shows high vocoid deletion (2 tokens only) for the tokens *puis* [pi] and *gracieuse* [søz], while dieresis is marginal. Interesting to note, three tokens for dieresis involve the place name *Lyon*. Other tokens realised with dieresis include *Fourvière*, *Louis* and *sixième*.

We find only a handful of word medial derived contexts. Five word medial tokens show 80% syneresis (4 tokens) and 20% dieresis (one token only of the word *pluriemploi*). Three tokens realised with syneresis show three instances of /HV+V/ preceded by a simple onset while a single token for the word *quatrième* presents an OL sequence reduced through liquid deletion: [kat(R)jɛm].

Table 103 shows realisations observed for 33 derived contexts. These include five word medial and 28 cross-word-boundary tokens.

Table 102: *Rhône Alpes*, *Lyon* (69a), tokens for all derived cross-word-boundary contexts across four speakers for each variant, N= number of tokens, % calculated by variant (DEL= deleted high vocoid, Dieresis and Syneresis) across speaker

	Speaker								Total	
	69aag1 (M)		69ajd1 (F)		69all1(F)		69asg1 (M)			
	N	%	N	%	N	%	N	%	N	%
DEL	--	--	--	--	--	--	9	64.3	9	27.3
dieresis	1	16.7	2	28.6	2	33.3	1	7.1	6	18.2
syneresis	5	83.3	5	71.4	4	66.7	4	28.6	18	54.5
Grand Total	**6**	**26.1**	**7**	**30.4**	**6**	**26.1**	**14**	**42.4**	**33**	**100**

Among these derived contexts, ten tokens for the cross-word-boundary environment involving the pronoun *tu* followed by a vowel initial verb produced by single speaker (sg1). With the exception of a single realisation with dieresis (*tu enlèves* [tsy(ɥ)ɑ̃]); all other high vocoids (nine) are deleted in this context.

Finally, we examine the distribution of these derived contexts across high vocoids. In Table 104 we show the distribution of all derived contexts across two high vocoids.

Table 103: *Rhône Alpes, Lyon,* realisations for all derived /HV+V/ contexts across high vocoids for three variants, N= number of tokens, % calculated by high vocoid for each variant (DEL= deleted high vocoid, dieresis and syneresis) across high vocoid

	Vocoid				Total	
	I		Y			
Variant	N	%	N	%	N	%
DEL	--	--	9	56.3	9	27.3
dieresis	2	11.8	4	25.0	6	18.2
syneresis	15	88.2	3	18.8	18	54.5
Grand Total	**17**	51.5	**16**	48.5	**33**	100

We note that 33 derived tokens show no instances involving the back high vocoid /U/. Interestingly, tokens are at near parity for each remaining high vocoid. Tokens for /Y/ show more diversity of variants: DEL 56.3%, dieresis 25%, syneresis 18.8%, compared to tokens for /I/ which strongly privilege syneresis 88.2% and 11.8% dieresis. Overall, 54.5% of derived tokens show syneresis followed next by deleted high vocoids at 27.3%; finally, dieresis is minimal at 18.2% overall. 29 tokens for cross-word-boundary contexts, which include four instances only of dieresis: tokens for /Y/include *venu à, revenue en* and *devenu un*) and for /I/ (*vie à*).

At this point, we would like to discuss a particular context of cross-word-boundary syneresis observed among these data. The spectrogram with oscillogram for the token *suis retraité* is shown below.

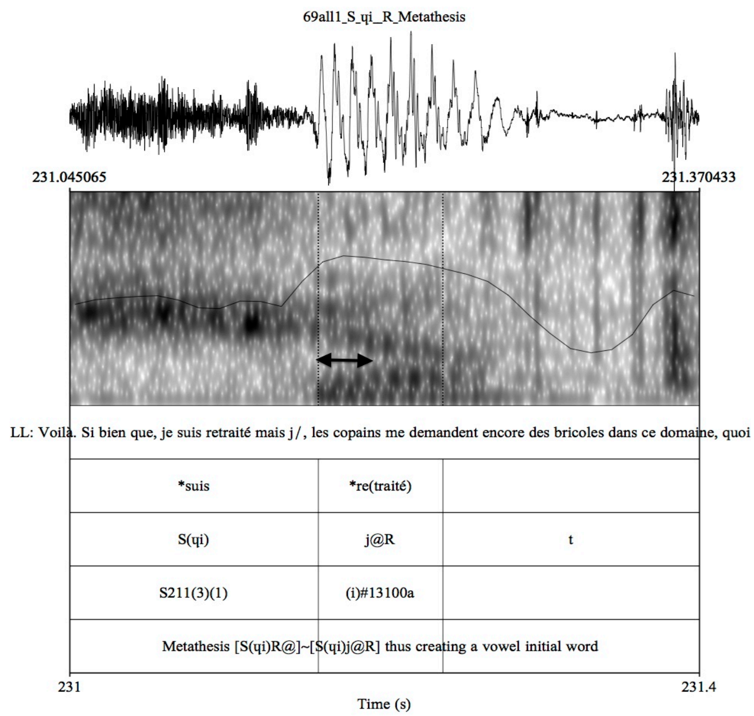


Figure 45: *Lyon* 69all1 *suis retraité* [ʃ(ɥi)jəʀ]realised with metathesis. Spectrogram with intensity contour, oscillogram and coding tiers: orthographic transcription, (*) token, SAMPA transcription (of token), coding, and observations

Figure 45 shows the phrase *je suis retraité* [ʃ(ɥi)jəʀ], in which, we observe assimilation between the pronoun and verb: *je suis*, which results in the voiceless palatal fricative [ʃ] followed by a devoiced glide + HV sequence [ʃ(ɥi)]. This token involves schwa metathesis which results in the verb *suis* [ʃ(ɥi)] followed by a vowel initial participle and a cross-word-boundary derived glide ensues: *je suis retraité* [ʃ(ɥi)jəʀ(tRɛte)]. The voiced production immediately following [ʃ] shows signs of transition which we take to indicate presence of the glide [j]. The presence of stable formant contours (not shown here) during the voiceless period preceding indicates the presence of a voiceless high vowel [i] rather than syncope (cf. Cedergren and Simoneau 1985). Note the low F2 at voicing onset and the transition shown by F2 and F3 indicated by the double arrow on the spectrogram. Across the vocalic period (between two vertical lines), the intensity contour shows a single peak occurring in the onset of voicing while the oscillogram shows a single peak at the middle of the waveform. This token is realised with syneresis: [ʃɥ̥i̯jəʀ]

We turn now to results for the survey from *Grenoble* (38a). *Grenoble* shows 535 tokens representing 63.1% of tokens for the region. In Table 105 below, we show results for all realisations of lexicalised tokens across three high vocoids.

Table 104: *Rhône Alpes*, 38a *Grenoble*, realisations for all lexicalised contexts across high vocoids for three variants, N= number of tokens, % calculated for each variant (DEL= deleted high vocoid, dieresis and syneresis), across three high vocoids

	Vocoids						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	1	0.5	2	1.2	19	23.5	22	4.8
Dieresis	4	1.9	--	--	--	--	4	1.0
Syneresis	208	97.7	167	98.8	61	75.3	436	94.2
Grand Total	**213**	**46.1**	**169**	**36.6**	**80**	**17.3**	**462**	**100**

86.5% of tokens observed for *Grenoble* are from lexicalised contexts. Of these, 94.1% are realised with syneresis the remaining 5.9% for other variants include: 4.8% deleted high vocoids, 1.1% dieresis. The high vocoid /I/ is most prevalent 46.1% then /U/ 36.6% and finally /Y/ is observed in 17.3% of all tokens. Tokens for deletion involve primarily the high vocoid /Y/: 13 instances of deleted /Y/ in the word *puis* (reduced to [pi]) as well as a single instance of *depuis* reduced to *depis* [dəpi], other deletions include *soignions* [soɲɔ̃] (with a change in the quality of the remaining vowel) and *voyage* [vajaʒ]. We also find instances involving /U and /I/. We find only four tokens realised with dieresis involving exclusively the high vocoid /I/ including two instances from the same speaker of the word *Clio* preceded by an OL sequence. Other instances involve the high vocoid /I/ preceded by a simple onaet (C) in the words *rien* [ʀijɛ̃] and *Parisienne* [ijɛn].

We turn now to an examination of all realisations for derived tokens observed for *Grenoble* in Table 106 below.

Table 105: *Rhône Alpes*, 38a *Grenoble*, realisations for all derived contexts across high vocoids for three variants N= number of tokens, % calculated for each variant (DEL= deleted high vocoid, Dieresis and Syneresis), across three high vocoids

	Vocoid						Total	
	I		U		Y			
	N	%	N	%	N	%	N	%
DEL	4	7.3	--	--	--	--	4	5.6
dieresis	8	14.5	2	100	9	60.0	19	26.4
syneresis	43	78.2	--	--	6	40.0	49	68.1
Grand Total	**55**	**76.4**	**2**	**2.8**	**15**	**20.8**	**72**	**100**

Derived tokens from *Grenoble* show syneresis to be most prevalent (68.1%); all other realisations (dieresis and deletion (DEL)) account for 31.9% of tokens. The high vocoid /I/ is most prevalent (76.4%), followed by /Y/ (20.8% of tokens) and finally we find /U/ involved in only 2.8% of tokens. Derived contexts realised with syneresis show no tokens for the high vocoid /U/. In Table 107 below we examine findings for cross-word-boundary contexts.

Table 106: *Rhône Alpes*, 38a *Grenoble*, all cross-word-boundary tokens across high vocoids for three variants N= number of tokens, % calculated for each variant (DEL= deleted high vocoid, Dieresis and Syneresis), across three high vocoids

	Vocoid							
Variant	I		U		Y		Total	
	N	%	N	%	N	%	N	%
DEL	4	8.7	--	--	--	--	4	7.1
dieresis	7	15.2	2	100	6	75	15	26.8
syneresis	35	76.1	--	--	2	25	37	66.1
Grand Total	**46**	**82.1**	**2**	**3.6**	**8**	**14.3**	**56**	**100**

Instances of deleted high vocoids involve the high vocoid /I/ only in cross-word-boundary contexts (*qui est, lui appris*, for example). Tokens realised with dieresis are primarily observed in cross-word-boundary contexts.

6.7.2 *Rhône Alpes*: observations across two surveys

Although this region is the second most populous in our study, 848 tokens across 10 speakers (six female, four male) represent only 26.4% of our sample or third most productive across surveys. Nonetheless, familiar patterns in the distribution of data are observed. We find a slightly higher number of lexicalised tokens overall which we can attribute to the distribution observed in Lyon (89.5% lexicalised). Still these numbers are in line with the overall distribution observed for *Île de France* and *Normandie*. Syneresis is privileged with much higher rates across lexicalised tokens as compared to derived contexts though derived tokens show more diversity among variants that diverge across high vocoids. With respect to high vocoid distribution we observe that the high vocoids again show a change distributional hierarchy: /I/ ≫ /Y/ ≫ /U/, such that in the derived data for Lyon we find no tokens for the back rounded /U/.

6.8 Realisations of /HVV/ sequences in Spontaneous Speech: Observations

Results from our study of lexicalised and derived glide contexts in the spontaneous speech of Hexagonal French speakers from the regions of Île de France, Normandy and Rhône Alpes show a consistent pattern of distibution. Overall, distribution across lexicalised and derived contexts shows that lexicalised contexts represent roughly 85% of all glide tokens, while derived contexts represent roughly 15%. Although distribution of lexicalised and derived contexts show minor variance (± 5%) across individual speakers, cross-speaker averages show this overall pattern maintains for all survey and regions: 85% lexicalised compared to 15% derived.

Interestingly, high vocoid distribution varies across these two contexts. All lexicalised contexts show the well described (and expected) distributional hierarchy across three high vocoids: I ≫ U ≫ Y, while distribution differs for derived contexts. The distributional hierarchy observed for lexicalised glides is not unexpected since we have seen this hierarchy described for French glides in markedness discussions in Chapter 4 (cf. section 4.5.2). The distribution of lexicalised contexts is in keeping with glide

typologies (cf. Greenberg 1966, Maddieson 1984). The typologically exceptional three-glide inventory of French can be seen to correspond to a markedness hierarchy evinced also in the evolution of the language. We know that yod /j/ occurs early and is seen as a tool to avoid hiatus in HVV contexts. Much later, the two rounded glides emerge in environments resulting in diphthongisation (cf. section 1.3).

Contexts involving the high vocoid /I/ show more realisations with dieresis (29 tokens, 2.1%) compared to one token each for /U/ and /Y/. Some tokens for the high vocoid /I/ realised with dieresis are predictable (preceded by an OL sequence as in the words *clientèle, février*, for example). We also observe multiple tokens preceded by a single consonant (*sixième, supérieur, filiale*, for example). These tokens show the consonants [z, l, ʀ, v] (all voiced, fricatives [v, z] or approximants [l, ʀ]).

Examination of variable realisations among lexicalised data supports descriptions for lexicalised glide contexts in French as (mostly) invariable. Lexicalised contexts privilege syneresis at a rate of 95.7% overall and across all three high vocoids (97.5% /I/, 98.8 /U/ and 84.2% /Y/). Nevertheless, some variable realisations, though infrequent, are observed amongst lexicalised data. These variable realisations show a possible correlation to the particular high vocoid: deleted high vocoid, and dieresis favour tokens involving /Y/, for example. /Y/ shows frequent high vocoid deletion (14.6%) whereas the high vocoids /U/ (1.1%) and /I/ (0.4%) disprefer deletion. Of these tokens: *voyage* [vəjaʒ], *science* [sɛ̃s], for example, high vocoid deletion is irregular and must be attributed to the nature of spontaneous speech (false starts and irregular articulations).

Derived tokens consistently show a distinct distributional hierarchy that privileges the front high vocoids over the back one: I ≫ Y ≫ U. We must attribute this finding for derived contexts to lexical distribution; morpheme final and word-final lexical entries privilege the high vowels /I/ and /Y/ over /U/ (cf. Juilland 1087). Juilland (1965) shows more lexical forms ending with -i [i] (1,777) followed by –u [y] (283) and finally -ou [u] 156 in word-final position (including past participles). As for verb forms that show word medial HV+V Juilland lists the following: 244 [ije; je], 79 [ye; ɥe], 60 [ue; we] (cf. Juilland 1965). Given the facts in this study, we propose that a markedness hierarchy

reflects an unordered relationship of dominance (/U, /Y/) between the two rounded high vocoids.

On the other hand, tokens for deleted /Y/ are quite regular. Among tokens for high vocoid deletion, we find 90.9% for the reduced forms: *puis*→ [pi], for example. We also find many tokens for reduced or assimilated forms as in (*je me) suis* [ʃɥi] which we attribute to the lexicalisation of the reduced forms. According to Bybee (2003) the frequency of a form (word or sequence of words) and any recurring patterns in usage can be seen influence the mental representation of that form (even the phonetic form). Among tokens for lexicalised reduced forms, we find multiple instances for *suis* which show palatalisation with voicing assimilation [ʃ ɥ̥(i̥)] that may co-occur with a non-reduced form [sɥi] from a single speaker. These instances may occur with or without reduction of the preceding elements: *je me*, for example. Reduced forms through high vocoid deletion are observed across all three regions (and six surveys) in varying number. An examination of these findings by speaker shows that presence of deleted /Y/ can be correlated to an individual speaker. These individual speakers span all ages of speaker (20-83). Reduced forms involving high vocoid deletion are in evidence across all three regions and appear to be conditioned by the individual speaker. This reduction can be seen as similar to the stop assibilation that is observed for tokens involving /tu+V/.

Chapter 7

7 Analysis and Conclusions

> [T]he interplay between syneresis and dieresis is quite variable and dependent on such things as regional variation, affective pronunciation and speech rate.
>
> Bullock (2002:13)

In this chapter, we discuss the implications of variable lexicalised and derived glide data and discuss accounts for exceptional dieresis, and by implication possible glide epenthesis, as well as exceptional syneresis exhibited in true diphthongs. We propose an account for instances of exceptional dieresis as lexical exceptions in a multilinear model based on discussions in Guerssel (1986) and Encrevé (1988, cf. 4.2-4.3 above) wherein dieresis in lexicalised and derived contexts can be accounted for through the preassociation of high vowels to a rime head. For the more broadly distributed variants observed in our derived glide data, realisations are represented in an Optimality Theoretical account. In an OT account, we show that these variable realisations may correspond to slightly different grammars i.e. different constraint rankings for derived glide contexts.

In the following section we discuss variable forms for lexicalised contexts.

7.1 Lexicalised glide contexts: analysis

Across 3415 tokens in our study of /HV(+)V/ contexts (cf. Chapter 6.2), tokens for all lexicalised contexts represent a large majority 80.6% of tokens. Among these data we observe lexicalised contexts show great consistency in realisations. Tokens realised with syneresis, the expected variant (based on desciptions of these contexts in the literature), represent 95.7% of overall glide data, while variable realisations are infrequent. High vocoid deletion is observed in 3.2% of data from lexicalised contexts and for all three high vocoids. Deleted tokens show an inverted distributional hierarchy compared to the general pattern of glide distribution in lexicalised contexts: /I/ ≫ /U/ ≫ /Y/, which corresponds to a markedness hierarchy (MH, cf. (83) in section 4.4.2). Among tokens for deleted high vocoids in lexicalised contexts, 15.6% of overall data involve /Y/, 1.1% of

all tokens involve /U/, and 0.4% involve /I/. This inverted hierarchy shown by data for deleted high vocoids suggests that high vocoid deletion may be conditioned by markedness: *M[ɥ] ≫*M[w], *M[j] (Bullock 2002 cf. discussion of margin fitness section 4.7). We show the back rounded and front unrounded glides as unranked because their small numbers are insufficiant. Across all lexicalised data 1.1% of tokens are realised with dieresis. Dieresis is observed in exactly 31 tokens of which 29 (93.5%) involve /i/ with a single token each for /u/ (*Louis* [lu.wi]) and /y/ (*lui* [ly.ɥi]). These findings provide further evidence to support a general markedness hierarchy for high vocoids: /I/ ≫ /U/, /Y/ (cf. 4.4.2 above). The high vocoid /I/, least marked of three high vocoids, is most prevalent in our data which therefore dominates distribution of these exceptional data. Exceptional data across the rounded high vocoids show a similar low frequency further supporting the unordered status given here.

29 realisations with dieresis involving the high vocoid /I/ include 15 (58.6%) expected realisations of dieresis given that they are preceded by an OL sequence (*clientèle, février, quatrième*, for example). Dieresis is, however, observed in 14 instances of /HVV/ preceded by a simple onset (C) in forms like *rien*, *Lyon, Parisienne, sixième*. Note that included in these are instances of morphemes beginning with high vocoid /I/ that we take to be lexicalised. Therefore variability (syneresis verus dieresis) is unexpected.

These forms with unexpected dieresis are best accounted for as lexical exceptions. Recall that according to Durand and Lyche, the greatest variability between syneresis and dieresis can be observed in monomorphemic forms:

> Même si nos locuteurs de FM prononcent les mots ci-dessus [(à savoir [bjoloʒi], [djabɛtə], [pjɔŋ], etc.)] avec synérèse, on ne peut exclure que d'autres locuteurs préfèrent la diérèse dans certains de ces mots. Et, pour nos locuteurs, la diérèse est habituelle là où l'usage standard favorise la synérèse : *Lia* [lija] (prénom), *liane* [lijanə], *liasse* [lijasə], *lion* = *Lyon* [lijɔŋ], *miette* [mijɛtə], *via* [vija], *mouette* [muɛtə], *luette* [lyɛtə], *nuage* [nyaʒə]. Un marquage lexical semble s'imposer.
>
> (Durand and Lyche 1999: 43)

How is this lexical marking to be implemented? The exceptional forms observed in our data are not all monomorphemic. Rather, some result from a morpheme realised variably with dieresis (105def.) below. Lyche (1979, cf. also section 3.4.2) discusses in some

detail the variability of surface forms observed in lexicalised forms: *tablier* /tabl- +-(i)je/, in which a morpheme shows variability compared to inflected forms: *oubliez* /ubli+e/ with a stem final /i/ as well as *sembliez* with a stem final OL (*sembl*+-*(i)*ez):

> From the point of view of language acquisition, the differences described here are easily explained: *tablier* [tablije] never alternates and is learned as such; *oublier* can alternate, but the root vowel is never affected by any rule; in *sembliez*, the vowel [i] shows up in a few isolated forms and a simplification is to be expected.
>
> (Lyche 1979: 327 transcriptions provided by Lyche added).

The implication here is that different forms of a single morpheme are lexicalised and are realised with variabilty depending on the context: *tablier* [tablije], *oubliez* [ublije] and *sembliez* [sã.bl(ij)e].

Table (107) below lists all instances from our lexicalised data of /HVV/ preceded by a simple onset (C) and realised with exceptional dieresis.

Table 107: Lexical items preceded by simple onsets /C/ realised with dieresis

Lexical items	N
a. *rien*	4
b. *Lyon*	4
c. *Fourvière*	1
Morphemes	
d. *Paris<u>iènne</u>* [iɛn]	1
e. *f<u>iliale</u>* [ial]	1
f. *six<u>ième</u>* [iɛm]	2
Total	13

In many instances, we may observe the same lexical item realised with syneresis by the same speaker. It appears that these data support a *'marquage lexical'* as suggested by Lyche and Durand (above) but how is this implemented? Following Guerssel's account (1986, cf. section 4.2 above) of glide formation in Berber, we propose an adapted multilinear analysis based also on the model for variability and latent consonants in French from Encrevé (1988, cf. section 4.3 above). In this model, for a limited number of lexical items (proper nouns and a few other frequently occurring words), some speakers may have two forms in their lexicon that alternate according to register, affective speech, or for other sylistic reasons. We count these instances of exceptional dieresis as lexical exceptions. We note that all realisations of lexically preassociated lexical items could

include epenthesis. Before discussing the representation of exceptional dieresis, we discuss a model for epenthesis that is complementary to our model for lexical preassociation.

Epenthesis is most often discussed in the context of /HV+V/ preceded by an OL sequence because dieresis is expected. As a result dieresis with or without epenthesis may occur but it is most commonly described for the front high vocoid /I/ [ij] excluding /Y/ [yɥ] and /U/ [uw]. Gaatone (1976), and Kaye and Lowenstamm (1984) represent glide epenhesis as a possible outcome for all of the high vocoids in French. Following Gaatone, we accept that epenthesis is a possible outcome for any high vocoid in hiatus. Thus all examples in this account show epenthesis in the outcomes though we remain agnostic on the subject of presence or absence of epenthesis[34].

> Cependant l'observation attentive de la prononciation de mots de cette espèce par des informants francophones ne laisse aucun doute sur la présence d'une semi-voyelle de transition entre les deux voyelles. Il est vrai que la perceptibilité, sur le plan acoustique, de [ɥ] et [w], est moindre que celle du yod, ce qui peut sans doute expliquer leur omission dans les transcriptions phonétiques.
>
> (Gaatone 1976: 323)

This account, based on a multilinear model of French liaison and elision (Encrevé 1988, cf. section 4.3), represents epenthesis as a process of the spreading of segmental material motivated by a dispreference for hiatus in the grammar (cf. discussion of Rosenthall in 4.5.1).

We turn now to a model of lexical exceptions observed in our data. Forms realised with dieresis when preceded by simple onsets (C) are unexpected, and are therefore marked. These exceptional lexical items can be best represented as lexical exceptions with the preassociation of the high vocoid to a rime head, as proposed by Guerssel (1986). The preassociation serves to prevent a glide formation rule from operating on this segment as shown in (123) below for the input /ʀIɛ̃/.

---

[34] In conducting our judgement task, judges did mention that it was difficult to make such fine-grained assessments of their perception (cf. section 5.3 above).

(123) Lexical preassociation (exceptional dieresis): *rien*

a. /ʀIɛ̃/ ⟶ [ʀi.jɛ̃]
  |
  x
  |
  R
(adapted from Guerssel 1986)

In the examples below, we illustrate the various phonological rule operations followed by the processes of syllabification based on Kaye and Lowenstamm (1984) for the the word *rien* with a preassociated rime (124 adapted from Guerssel 1986) and a form without preassociation (125) for the word *rien* that can account for surface realisations with syneresis versus dieresis.

In (124) below we show the phonological processes and rules of syllabification that will apply before the form with dieresis is realised in a surface form.

(124) GF is blocked by preassociation of the high vocoid.

a. /ʀIɛ̃/ ⟶ /ʀiɛ̃/
  | | | |
  x x x x
  |
  R

In (124a.) the preassociated rime prevents the glide formation rule from operating thus a surface realisation with syneresis (privileged in this environment) is blocked and we observe a realisation with dieresis. The high vocoid /I/ is realised as the high vowel [i] which is necessarily designated [+ voc.].

b. /ʀIɛ̃/ ⟶ /ʀi_ɛ̃/
  | | | | |
  x x x x x
  | | | | |
  R O R O R

All segmental material (as defined by features) is associated to a syllable constituent based on a CV template that is privileged in the grammar. In this instance the rime node is preassociated (124b.).

```
c. /ʀIɛ̃/ →  /ʀi jɛ̃/   →  [ʀi.jɛ̃]
     |         | |/| |
     x         x x x x
     |         | | | |
     R         O RO R
```

As a result of segmental association an empty onset following the high vowel [i] remains. In a grammar that privileges a CV syllable structure epenthesis (high vocoid spreading to a following empty onset) can be seen as a repair strategy. The surface form is realised with epenthesis and dieresis (124c.).

In (125) below we show the same operations for the word *rien* without rime node preassociation. The high vocoid that is not preassociated and will undergo a rule of glide formation and therefore surfaces with syneresis.

(125) No preassociation of the high vocoid

```
a. /ʀIɛ̃/ →  /ʀjɛ̃/
     |         | ||
     x         x x x
```

Following Kaye and Lowenstamm who stipulate “[...] les semi-voyelles et les nasales soient spécifiées, exactement comme les liquides, comme [+voc, +cons].” (1984: 131), segmental material is associated to a syllable constituent based on the CV template that is privileged in the grammar (125b.).

```
b. /ʀIɛ̃/ →  /ʀjɛ̃/
     |         | | |
     x         x x x
               | | |
               CC V
```

In this way, the glide corresponding to the high vocoid /I/ is designated [+ cons.] (125b.) and is syllabified to the onset position that precedes it.

```
c. /ʀIɛ̃/ →  /ʀjɛ̃/
     |         |/  |
     x         xx  x
     |         |   |
     R         O   R
```

(adapted from Guerssel 1986 based on Kaye and Lowenstamm 1984)

In (125a.) above, the form has a lexically preassociated rime for /I/ thus preventing a rule such as glide formation from operating on the high vocoid. Therefore, the surface form shows dieresis. Expected forms with no preassociation are subject to all processes such as GF and are realised with syneresis (124b.). Unlike the contrastive forms in Berber, which are the object of Guerssel's account (cf. 68ab. below), the example here models lexical exceptions. Words with a lexically preassociated rime will not undergo phonological rules (such as GF). Instead, they pass unaltered to operations of syllabification where we may observe a process of epenthesis motivated by a grammatical imperative which privileges CV surface. These forms may then be subject to post-lexical rules such as voicing assimilation and epenthesis (125c.).

A model of lexical preassociation can also represent instances of exceptional syneresis where dieresis is expected. In our spontaneous speech data we observe a single token for exceptional syneresis for the morpheme *–ième* when added to a stem ending in OL *quatr-* [katʀ]. The preassociation of the HVV sequence to a rime node can account for the exceptional realisation of GF following a complex onset OL as shown in (126) below.

(126) Lexical preassociation (exceptional syneresis): *quatrième*

a. /-Iɛm/ → [ka.tʀjɛm]
   xx
   |
   R

No preassociation (dieresis): *quatrième*

b. /-Iɛm/ → [ka.tʀi.jɛm]
   |
   x

(adapted from Guerssel 1986)

In (126a.) the preassociation of a /HVV/ sequence to a rime results in exceptional syneresis (mirroring exceptional dieresis above). In this instance the morpheme *–ième* is exceptionally realised with syneresis [jɛm]. In this instance, we propose the lexical preassociation of the morpheme internal structure to a rime node resulting in a diphthong (126a.). Without lexical preassociation the morpheme is subject to all rules and

constraints (*OLG) of the grammar such that following an OL sequence dieresis is expected as shown in (126b.).

In (127) below we show the phonological processes and rules of syllabification that will apply before to the morpheme with dieresis is realised in a surface form. In (127a.) the preassociated nucleus triggers a rule of glide formation that operates on the morpheme-internal HVV sequence which can be seen to create the diphthong [je].

(127) Lexical preassociation of the nucleus

```
a.  /-IEm/    →  /-jem/
      V            V |
     xx            xx  x
      |
      N
```

In (127b.) the vowel undergoes a rule to adjust the aperture/quality of the closed mid vowel in a closed syllable (/e/→ [ɛ]) resulting in the surface realisation with [jɛ]. The morpheme is realised with syneresis which is highly unexpected when preceded by an OL sequence as in *quatrième*

b. Vowel adjustment

```
/-jem/     → /katʀ+jɛm/  →   [katʀjɛm]
  V            | | | |  V |
  xx           x xx x  xx x
  |              |      |
  N              N      N
  |              |      |
  R              R      R
```

In 127b.) this diphthong internal morpheme is appended to the stem *quatr-* and satisfies the rules of syllable wellformedness as in the lexicalised monomorpheme *trois*.

In (128a.) we show the same morpheme with no preassociation. We can assume that the same vowel adjustement rule: /e/ → [ɛ], operates in the lexicon before derivation.

(128) No preassociation

```
a.  /-IEm/    →  /-iɛm/
      | |           | |
      x x           x x
```

When appended to the stem *quatr*- below, GF is blocked by the preceding OL sequence and does not apply. As a result this morpheme bearing an HVV sequence is realised with dieresis (128b.).

```
b. /-IEm/ → /ka.tʀi_ɛm/
     | |          || |||  ||
     x x          xx xxx  xx
                 || V  |  V
                 OR O  R_R
```

As before, all segmental material is associated to an appropriate segmental class. Based on a CV template in the grammar, the rules of syllabification operate maximising onsets before syllabifying consonants to a coda. As in (124c.) above, the resulting empty onset preceded by a high vowel triggers glide epenthesis (128c.).

```
c. [ʀi.ɛm/   →  [ka.tʀi.jɛm]
     |
     x
```

In (128c.), the surface form is realised with epenthesis and dieresis in the morpheme.

In the following discussion, we assume the various processes illustrated above: glide formation and glide epenthesis, to determine surface realisations as illustrated in (124-125 *rien* and 127-128 *quatrième*) above.

In (129a.) below, we illustrate a model of lexical preassociation based on the example of exceptional dieresis in *rien* in (124) above to account for the exceptional dieresis observed in our data for the toponym *Lyon.*

(129) Lexical preassociation (exceptional dieresis): *Lyon*

```
a. /lIɔ̃/  →  [li.jɔ̃]
     |
     x
     |
     R
```

Without preassociation (syneresis): *Lyon*

b. /lIɔ̃/ → [ljɔ̃]
  |
  x

(adapted from Guerssel 1986)

We turn now to instances observed for variable realisations of morphemes beginning with the high vocoid /I/. Examples below are based on tokens observed in our data for vowel-initial morphemes realised with exceptional dieresis. We follow the model of preassociation illustrated in (125) above to account for exceptional dieresis in the morphemes *–ial* as in *filiale*, *–ien* as in *Parisienne* and *–ième* as in *sixième.* In (130a.) below, we show a lexically preassociated rime for the first high vocoid in the morpheme *–ial* which can account for the unexpected realisation with dieresis observed in our data.

(130) Lexical preassociation (exceptional dieresis): *filiale*

a. /-Ial/ → [fi.li.jal]
  |
  x
  |
  R

No preassociation (syneresis): *filiale*

b. /-Ial/ → [fi.ljal]
  |
  x

(adapted from Guerssel 1986)

Forms with no preassociation as in (130b.) undergo rules such as GF producing a surface form with syneresis.

In the examples below, we model lexical preassociation of the rime node and absence of preassociation for the morpheme *–ienne* [jɛ̃; jɛn] and *–ième* [jɛm]. In (131a.-132a.) the rime is preassociated to the high vocoid /I/. Thus, this form surfaces with dieresis. Forms without preassociation are subject to rules including GF rendering the surface form with syneresis (131b.-132b.).

(131) Lexical preassociation (exceptional dieresis): *Parisienne*

a. /-Iɛn/ [pa.ʀi.zi.jɛn]
   |
   x
   |
   R

No preassociation (syneresis): *Parisienne*

b. /-Iɛn/ [pa.ʀi.zjɛn]
   |
   x (adapted from Guerssel 1986)

(132) Lexical preassociation (exceptional dieresis): *sixième*

a. /-Iɛm/ → [si.zi.jɛm]
   |
   x
   |
   R

No preassociation (syneresis): *sixième*

b. /-Iɛm / →[si.zjɛm]
   |
   x (adapted from Guerssel 1986)

Forms shown in (130-132) show lexicalised tokens for HVV sequences realised with dieresis accounted for with the preassociation of rime nodes to the high vocoid of the morpheme. These can be seen as lexical exceptions to the GF rule motivated by a grammar that privileges CGV (syneresis following simple onsets) in the output.

As shown in these examples, this model could account for any variable high vocoid inital morpheme (cf. discussion for Martinet and Walter 1973 in section 1.3.3, Lyche 1979 in section 3.4 and Durand and Lyche 1999 section 4.6). In Table 110 below we provide a list of high vocoid initial morphemes drawn from Thiele (1987). Note that high vowel initial morphemes involve /I/ primarily[35].

---

[35] The vast majority of high vocoid initial suffixes involve /I/. The few instances of suffixes involving other high vocoids are unproductive as in the diminutive *–ule* (*globule, fécule*) or related to highly specialised fields as in *–um* (minerologie), -*ure* (chemistry).

Table 108: High vocoid initial morphemes (Thiele 1987)

Classe	Suffixes	Example
N → Adj.	*-ien m. /ienne f. "qui a rapport avec"*	*sartrien, hugolienne*
N → Adj.	*-ien m. /ienne origine géographique*	*Parisien, Vietnamienne*
N → N	*-(ic)ien m./(ic)ienne f. "nom d'agent"*	*acousticien, opticienne*
N → N	*-(i)er m. /(i)ère f. "nom d'agent"*	*routier, fermière*
N → N	*-ier "sert à denommer des arbres"*	*poirier, cerisier*
N → N	*-ier m./ ière f. "indique le contenant"*	*cendrier, cafetière*
N → N	*-ier m./ ière f. "designer le contenant, objet ou lieu"*	*poudrier, chaumière, encrier*
N → Adj.	*-(i)eux m./(i)euse f.*	*astucieux, mystèrieuse*
N → Adj.	*-(i)al m./ (i)ale f.*	*facial, familliale (iale = allomorphe*
Adj. Numéraux	*-ième (à former les ordinaux)*	*deuxième, dixième*

This model accounts for the variability observed in our data as lexical exceptions to a rule of glide formation thereby accounting for unexpected dieresis in our data. In this model, we can also account for exceptional syneresis as shown in (133a.) above. These exceptional lexical items (or morphemes) realised with dieresis when syneresis is expected by rule can be accounted for with a lexically preassociated rime in an individual's grammar that is employed for stylistic purpose (such as careful speech, or emphasis, related to "media style" cf. Bullock 2002: 23. See also the discussion of initial stress based on Lyche and Girard 1995 in section 4.3).

We turn now to an analysis of the variability found in derived /HV+V/ tokens.

7.2 Derived context data: analysis

The 459 tokens observed in derived contexts represent only 14.3% of our overall data, and show considerably more variability than is observed in our lexicalised data. Syneresis is most common in derived contexts representing (61.1%), a lower rate than lexicalised contexts (95.7%). Thus, there is more room for variability in derived contexts. Realisations with dieresis represent 28.1%, while high vocoid deletion represents 10.9% of overall derived tokens. For each of the three regions studied, the distribution of tokens across three high vocoids shows the same pattern wherein /I/ dominates overall (73.7%), then /Y/ (20.9%) ahead of /U/ (5.4%), a pattern which corresponds to the markedness

hierarchy for French high vocoids discussed above (cf. 4.4.3 (83)). The data show all three variants (deleted high vocoid, dieresis, and syneresis) with the exception of deleted high vocoid for /U/, which we attribute to a more restrained distribution of /U/ in these derived tokens.

Our data for derived glide contexts can be further divided based on the environment in which they occur: /HV+V/ in word medial and in cross-word-boundary environments.

7.2.1 Derived word medial environment

Tokens of derived word medial contexts number only 76 and represent 17.8% of all derived data. Glide distribution in these tokens follows from the general MH that consistently shows a re-ranking for rounded high vocoids: /I/ (64.5%) ≫ /Y/ (25%) ≫ /U/ (10.5%). Word medial tokens show no deleted high vocoids and overall realisations are relatively balanced across the three high vocoids. Tokens for /I/ represent 64.5% with 83.7% syneresis and 16.3% dieresis; tokens for /Y/ represent 25% of word medial derived tokens with 84.2% syneresis and 15.8% dieresis, and tokens for /U/ represent 10.5% of all word medial contexts with 87.5% syneresis and 12.5% dieresis.

In this context we find that all three high vocoids show comparable rates for dieresis. Though numbers are small, the distribution of tokens and their variable realisations appears to follow from the MH: /I/ 49 (41 syneresis, 8 dieresis); /Y/ 19 (16 syneresis, 3 dieresis; /U/ 8 (7 syneresis, 1 dieresis). Given the consistent proportion of dieresis across all three high vocoids (12.5%-16.3%), we can best account for instances of dieresis in word medial contexts as exceptions to the expected outcome (represented in a similar fashion as (125-129) i.e. via the preassociation of the high vocoid at the end of a verb stem which has the effect of blocking the glide formation rule).

In the following examples (133-136), we illustrate a model for lexical preassociation to account for the exceptional dieresis observed in our data. These examples are based on the model shown for *rien*. In (133a.-136a.) we model each token with glide epenthesis, as a possible outcome, for the forms *continuer, louer* and *varier* realised with dieresis. Each of these illustrations shows lexically preassociated stem final high vocoids precedeed by

a simple onset (C) involving each of the three high vocoids (*continuer* [y], *louer* [u], *varier* [i]) to account for realisations with exceptional dieresis.

(133) Preassociation of stem final high vowels (exceptional dieresis): *continuer*

a. /ny+e/ → [ny.ɥe]
|
x
|
R

No preassociation (syneresis): *continuer*

b. /ny+e/ → [nɥe]
|
x

(adapted from Guerssel 1986)

Forms like (133a.) show the high vowel /y/ lexically preassociated to a rime head and will not undergo GF. The form in (133b.), on the other hand, without preassociation will undergo the rule for GF resulting in syneresis.

The examples below model the preassociation of the high vowel [u] and [i] for the words *louer* /U/ and *varier* /I/ (134a-135a.), while forms with no preassociation will undergo the rule of GF (134b.-135b.).

(134) Preassociation of stem final high vowels (exceptional dieresis): *louer*

a. /lu+e/ → [lu.we]
|
x
|
R

No preassociation (syneresis): *louer*

b. /lu+e/ → [lwe]
|
x

(135) Preassociation of stem-final high vowels (exceptional dieresis): *var<u>ier</u>*

a. /vaʀi+e/ → [va.ʀi.je]
        |
        x
        |
        R

No preassociation (syneresis): *var<u>ier</u>*

b. /vaʀi+e/ → [va.ʀje]
        |
        x

Under a lexical preassociation analysis, tokens realised with dieresis can be seen as exceptions to the glide formation rule that would produce the surface forms in (133b.-135b.).

Recall that among the categorical results observed from our judgement task (cf. section 5.4), we observed a single token for 81 *trouer* judged categorically by our panel of judges to be realised with syneresis. In any model, this token is more than exceptional; it is aberrent really. Its exceptionality suggests that it is more likely a performance error. In any case, it would require further study which is beyond the scope of this thesis. We prefer to call this particular token an outlier. A model of lexical preassociation to a rime head accounts for these unexpected realisations.

In the following section, we propose an alternative model to account for the more generalised variability in the OT framework.

7.2.2 Variability of derived word medial data in OT

Variability in glide data can also be modelled in an Optimality Theoretical (OT) account wherein the different pronunciations of glides (GF versus dieresis) can be seen as variable aspects of a grammar. In an OT account, we show that variable realisations of cross-word-boundary data may correspond to slightly different grammars i.e. different constraint rankings for derived glide contexts. In this model, lexical exceptions (like those above) are recast as the ideal candidate in a constraint ranking of markedness and faithfulness constraints. The lexical exceptions (dieresis preferred over syneresis) can be

modelled in a constraint ranking that places faithfulness to input morae over markedness: Max-μ ≫ ONSET. This ranking corresponds to the modified grammar that speakers employ for stylistic (affective) purposes. When syneresis is preferred over dieresis the markedness constraint ONSET must dominate the constraint ranking.

This account relies on constraints proposed by Hall (2006, cf. section 4.8). We accept the notion of morae as a convention used to model nuclear constituency for high vowels and the non nuclear constiuency of glides, while we acknowledge that glide formation in Modern French does not exhibit the kind of compensatory effects for the loss of a mora as is observed in other languages exhibiting glide phenomena (Rosenthall 1997, cf. section 4.5.1)

We begin with an account for *louer*. As in the various OT analyses surveyed (cf. section 4.6-4.8) glide formation is accounted for with a markedness dominant ranking. Note that as unordered constraints both ONSET and NOBRANCH show fatal constraint violations. Without reranking these constraints are equally violated. Since markedness dominates faithfulness in the constraint ranking, marked structures like hiatus or epenthesis (both realised with dieresis) are penalised at the cost of glide formation with syneresis (136b.).

(136) *louer*

/ly+e/	ONSET	NOBRANCH	MAX-μ
a. lu.e	*!		
b. ☞ lwe			*
c. lu.we		*!	

(Hall 2006 (36))

The two candidates showing dieresis in (136a. hiatus) and (136c. epenthesis) are excluded since each suffers a fatal constraint violation. While Hall acknowledges that glide formation may be optional GF is represented here as categorical.

Hall is also mindful of the potential for differing treatments across glides and offers this:

> In French phonology it is usually assumed that there is a palatal glide between the [i] and following vowel, e.g. [plije] and not [plie] (see, for example, Tranel 1987). For other authors (e.g. Kaye and Lowenstamm 1984) there are homorganic glides between [u]/[y] plus vowel sequences as well. I follow the latter approach in my analysis.
>
> (Hall 2006: 835, note 31)

In a slightly modified analysis in (137) below, we propose the unordered ranking of the constraints NoBranch and Max-μ which can account for two optimal candidates, glide formation (137b.) or glide epenthesis (137c.), with no need to revaluate candidates or add constraints (cf. Anttilla 1997, Nagy and Reynolds 1997).

(137) *lier*

/li+e/	Onset	NoBranch	Max-μ
a. li.e	*!		
b. ☞ lje			*
c. ☞ li.je		*	

(Based on Hall 2006 (36))

In (137) above, the unordered constraints NoBranch and Max-μ make the candidates in (137b.) and (137c.) equally optimal, while the candidate showing hiatus in (137a.) suffers a fatal violation of the Onset constraint. These two candidates represent the variants observed in our overall findings with the caveat that across three high vocoids, data for derived word medial tokens that privilege glide formation (syneresis as in 137b.) occur at a rate of roughly 80% while dieresis (with or without hiatus) averages 20%. However, across three high vocoids and for lexicalised or derived contexts, data are not equally distributed. It is beyond the scope of this thesis to propose a means to weight these competing optimal candidates in a way that reflects their distribution, their rate of realisations with syneresis versus with dieresis, and the environment studied (cf. Reynolds and Nagy 1997). The difference noted in the distribution of tokens across contexts has implications for the markedness hierarchy and its interactions with the margin fitness hierarchy, as proposed by Bullock (2002).

Crucial to an account of derived forms (Tableaux 136-137), the markedness constraint Onset must dominate faithfulness. Thus candidates showing dieresis are found to be suboptimal (136ac.) while the unordered constaints (NoBranch and Max-μ) proposed in

(137) select both candidates realised with syneresis and realised with dieresis GF in (137b.) and epenthesis in (137c.) as optimal output candidates. In this account an unordered constraint ranking in (137) permits us to represent the variability observed in our data as resulting from a grammatical preference for a syllable onset CGV/CV.GV versus no onset (136a.-137a.). Note that the token in (137) shows the high vocoid /I/ while in (136) we show the high vocoid /U/. As our data suggest, the distribution and variability of realisations appear strongly conditioned by the particular vocoid involved and the environment in which it is found. Following from a general MH our data show /I/ consistently dominates the two rounded glides. In derived contexts realisations can be seen to follow from the margin fitness hierarchy: *M[ɥ] ≫ *M[w] ≫ *M[j] proposed by Bullock (2002: 18). According to this hierarchy the front rounded [ɥ] is least optimal or most marked, therefore least likely to be parsed in a syllable margin. Following [ɥ], is the slightly less marked back rounded [w] and finally the front nonround [j] is least marked thus most likely to be parsed to a syllable margin. According to Bullock, while the implications for this MH are representative of glide distribution in French, “more important, it may simply reflect a more universal markedness of the front round glide: 10)”, Bullock (2002: 18). We agree with Bullock: our findings show that glide distribution in French corresponds closely to cross-linguistic glide typology observed in Maddieson (1984, cf. section 4.4.2). A detailed examination of the interactions of the MH with the margin fitness hierarchy would go beyond the current study.

We turn now to an analysis of data with HVV contexts occurring across word boundary (cf. section 6.3.2 above).

7.2.3 Cross-word-boundary context

We have chosen to examine data in cross-word-boundary contexts because it became evident to us that this environment is very often the target of GF. Discussion of cross-word-boundary glide formation is often absent in the literature or, if discussed, it is limited to a very specific environment (high vowel final prefixes, for example, of which we find only one token in this corpus, *pluriemploi,* realised with dieresis). Discussions of these data most often involve the blocking of glide formation (cf. Hannahs 1995, Tranel

1987). Our study shows that glide formation across-word-boundary occurs more systematically than is usually described (cf. section 6.3.2). In our spontaneous speech data, we find that cross word boundary contexts are much more prevalent than word medial derived contexts which dominate the descriptive literature. Fully 350 or 76.3% of all derived tokens are observed in cross-word-boundary environments a significant finding that must be investigated further. All three high vocoids are observed in word-final position followed by a vowel initial word (*fourni avec, venu à, loue un*, for example).

Overall findings for cross-word-boundary tokens show that syneresis is the favoured outcome in 60.3% of all tokens, dieresis is observed in 32.9% of tokens and deletion represents 6.9%. Tokens present distributions that vary according to the high vocoid involved. High vocoid /I/ privileges syneresis (66.6%) while tokens for /Y/ and /U/ favour dieresis (69.1% and 72.2% respectively). These data follow Bullock's markedness hierarchy for margin fitness. According to Bullock "[t]he distributional facts alone suggest that there is a fairly high ranked constraint in French that penalizes parsing glides, particularly the rounded glides, into syllable margins: *M/ɥ ≫ *M/w ≫ *M/j (2002: 18). Cross-word-boundary data in our study appear to support this position.

A majority of tokens (74.4%) involve a high vocoid contained in a monosyllabic word followed by a vowel initial form (*dit après, qui est, où on,* for example), whereas the remainder of cross-word-boundary tokens (24.5%) involve word-final high vocoids in polysyllabic words as in *revenu à*, for example. Distributions across high vocoids differ for monosyllabics bearing a high vocoid in word-final position compared to polysyllabics. The differences observed serve to reinforce a markedness hierarchy that shows the rounded vocoids are dominated by /I/ but are themselves are unranked /I/ ≫ /U/, /Y/. Note also that the high rate of tokens for *qui* + V (56.9%) has an undue influence on the imbalanced distribution observed across high vocoids. Nonetheless, if we elimate these tokens with *qui* + V, this hierarchy maintains (45.3% /I/, 5.4% /U/, 12% /Y/). Recall that due to the high frequency of high vocoid deletion, we have eliminated 28

tokens for the pronoun *tu*+V[36](cf. section 6.3.2).

In our cross-word-boundary data involving high vocoids in monosyllabics, we observe many instances of the same token which often involve prepositions or pronouns immediately following the monosyllabic with a high vocoid: *vu à*, *ou à*, *dit après*, *si on*. We might propose to account for glide formation in this environment as a post-lexical phonological rule motivated by a preference for syneresis in this context (cf. Kiparsky 1985, section 4.4 above). In this light, glide formation could be considered an allophonic variant of the high vocoid. However given the variability of realisations, we feel glide formation does not fit the model. In these cross-word-boundary environments, monosyllabic words with a high vocoid that undergoes glide formation when followed by a vowel initial word can be treated like proclitics. Monosyllabics bearing high vocoids appear to form a unit with the vowel-initial word that follows: *qui est, où est, ou à, si on*. Lyche 1979 discusses the implication of a weak syntactic boundary in this sort of context (cf. Chapter 3, 62c.). This unit emulates the structure observed in clitic groups or phonological words (PW, cf. section 4.4 above). The prevalence of glide formation in forms such as *qui est* [kjɛ] supports the analysis of monosyllables as similar to proclitics in PWs. Glide formation of course remains optional (/Y/ and /U/ prefer dieresis). Further research is required to determine if all monosyllables in this position can be proclitics, or only certain lexical sub-classes, and exactly what factors favour or disfavour glide formatino in this context.

7.3 Conclusions

Our study of glide data in Continental French shows that glide phenomena are more variable than most modern descriptions report. We have shown that variation is conditioned by the context in which a /HV(+)V/ token is found. Lexicalised glide contexts overwhelmingly favour realisations with syneresis. Word medial derived glide

[36] As it turns out, inclusion of these 28 tokens changes little in the overall distribution of tokens across high vocoids but their recurring behaviour such as a prevalence for deleted /Y/ observed is consistent with our analysis of the tokens that remain.

data shows syneresis is also preferred, however, we observe systematic variable realisations for all three high vocoids. Where we do find variable realisations among lexicalised data, these are vanishingly few and involve, near categorically, the high vocoid /I/. We have shown that exceptional instances of dieresis invoving /I/ are best accounted for as lexically marked with the preassociation of the high vocoid to a rime node which prevents the application of a rule like glide formation (cf. section 4.2). Variability in word medial derived contexts can also be accounted for using the same mechanism of a preassociation of rime as lexical exceptions or, as shown in an alternate analysis above, variability can be modeled as variable grammars through constraint ranking in an OT account (cf. sections 4.7-4.9). Given the nature of the guided interviews (cf. Chapter 5), this sample can be considered to exhibit a semi-formal style with periods of informal and affective speech. Some level of lexical exceptions can be attributed to a speaker's stylistic practice (formal or careful speech, affective pronunciations etc.). However, further analysis is required to determine whether variation in constraint ranking can actually be correlated with style.

Despite descriptions that ignore or downplay cross-word-boundary environments, our study shows that there is, in fact, considerable evidence of cross-word-boundary glide formation. In particular, we find that the nature of this environment involves monosyllabic words bearing high vocoids followed by a vowel-initial word wherein the prevalence of glide formation can be seen as a stategy in the cliticisation of these formally separate words (cf. section 6.3.2). These data also show systematicity in their realisation. Syneresis is privileged in contexts involving /I/ though tokens for the rounded glides show that dieresis is preferred, which suggests that the markedness hierarchy is at least partially active: I ≫ U, Y. Cross-word-boundary glide formation occurs frequently. These data follow many gramatical constraints described for other contexts, i.e. realisation of glide formation versus non-realisation is not random. These tokens can be seen to follow from the same grammar and constraints observed in word medial contexts and can be modelled with the same constraints in an OT analysis as shown for word medial data above.

We have shown that across all three regions studied, the distribution of high vocoids in lexicalised and derived glide contexts is surprisingly consistent, which we take as an indication of the inherent composition of spontaneous speech. The distribution of glide contexts across all high vocoids appears to be conditioned first by the distribution of high vocoids in the lexicon and secondly by the contexts in which they occur: lexicalised versus derived. Distribution of high vocoids in lexicalised contexts consistently corresponds to a general markedness hierarchy: /I/ ≫ /U/ ≫ /Y/ (cf. Greenberg 1966, Maddieson 1985, Bullock 2002). Overall data from derived contexts show that GF is favoured by yod and dispreferred by the two rounded glides, which corresponds to the general markedness hierarchy for glides: *ɥ ≫ *w ≫ *j, as presented by Bullock (2002:18, cf. section 4.7) and supported also by increased rates for deleted /Y/. These data present a slightly altered markedness hierarchy wherein the more marked rounded high vocoids /U/ and /Y/ are uncontroversially dominated by /I/ but the rounded high vocoids remain variably ranked: /I/ ≫ /U/, /Y/. Evidence from lexicalised contexts supports the ranking /U/ ≫ /Y/ while derived contexts support the ranking /Y/ ≫ /U/ (cf. Chapter 4), a partial inversion of the MH for which we as yet have no explanation. Our data confirm that glide data follow a markedness hierarchy in evidence in the grammar of French but also cross linguistically. Across all contexts tokens involving the front high vocoid /I/ are most prevalent and show the greatest variability.

Our study of the particular contexts in which glide data arise provides a more nuanced understanding of the distribution as well as systematic variability of glide data in French. Data patterns in lexicalised contexts are not identical to derived glide contexts. In particular, derived glide contexts show more systematic variability that involves all three high vocoids and they can be distinguished by the environment in which they occur. Tokens for cross-word-boundary glide formation occur in frequently and their realisations follow many of the same grammatical constraints as word medial glide formation.

Future studies of French glides would be enhanced by examining the various aspects of the glide phenomena in the specific environments in which a glide can occur. Glide formation is not a categorical process as is presented in descriptive studies that are based

on data sets that present a single glide formation environment. By identifying mechanisms and constraints involved, we can better understand what varies and what is constant in these different environments.

This thesis has shown that glide phenomena in a spontaneous speech sample correspond very closely to a general markedness hierarchy that privileges the front non round high vocoid over the back rounded and finally front rounded high vocoid (Bullock 2002 cf. section 4.8 and section 6.3.2). Future studies could focus more closely on the implications of markedness in glide realisations since our data have shown that all high vocoids are not alike and in fact the environment in which they occur may condition their behaviour in /HVV/ environments (cf. section 6.3).

Results from the Judgement Task showed that variable realisations for a single high vocoid may be conditioned by the preceding environment. While our spontaneous speech data provided too few tokens to examine the effect of a preceding environment in detail, the effects of the preceding environment in glide realisations could be studied in a laboratory environment wherein a controlled sample would provide a larger number of tokens that involve a broad spectrum of preceding segmental material (including complex onsets) in order to examine more closely how preceding segments or classes of segments might condition realisations in /HVV/ environments across all three high vocoids.

Variability in lexicalised contexts and perhaps some derived glide contexts might be better accounted for in correlation with an individual speaker's use of style. The effects of a speaker's affect, their use of style and by implication their speech rate on the realisation of /HVV/ environments in derived versus lexicalised contexts and for all three high vocoids also merit further study.

Finally, the phenomena of glide formation and glide epenthesis would benefit from further acoustic analysis. In a laboratory-controlled study, the participants could be balanced for social factors. Lexicalised and derived contexts could also be targeted while controlling for linguistically internal factors such as preceding and following environment. A controlled study would also allow the use of a variety of tasks to elicit formal and less formal style. With such controls in place we would better be able

to examine the presence or absence of epenthetic glides, to study the influence of age and gender on interspeaker variability, and to determine the role if any of the immediate internal phonological environment in the realisations with syneresis or with dieresis.

References

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Anttila, Arto. 1997. Deriving variation from grammar. *Amsterdam studies in the theory and history of linguistic science series* 4 (1997): 35-68.

Baehrens, W. A., 1922, *Sprachlicher Kommentar zur vulgärlateinischen Appendix Probi*. Hlle (Saale). Verlag von Max Neirmeyer

Bourciez, Edouard. 1967. *Phonétique française–étude historique*. Paris: Klincksieck.

Bullock, Barbara E. 2002. On constraining the vagaries of glide distribution in varieties of French. In Caroline R. Wiltshire and Joaquim Camps (eds.). *Current Issues in Linguistic Theory*, 11-25. Amsterdam/Philadelphia: John Benjamins.

Bybee, Joan. 2003. *Phonology and language use*. Cambridge: Cambridge University Press.

Carreira, Maria. 1989. The representations of rising diphthongs in Spanish. In *Theoretical analyses in Romance linguistics*: selected papers from the ninetheenth Linguistic Symposium on Romance Languages (LSRL XIX), C. Laefer and Terrell A. Morgan (eds.), 19-36, Ohio State University.

Carton, Fernand. 1974. *Introduction à la phonétique du français*. Montréal: Bordas.

Cedergren, Henrietta and Louise Simoneau.1985. La chute des voyelles hautes en français de Montreal: "As tu entendu la belle syncope?", in M. Lemieux and H.J. Cedergren (eds.). *Les tendences dynamiques du français parlé(s) à Montréal*. 57-144.

Centre National de Resources Textuelles et Lexicales (CNRTL online): http://www.cnrtl.fr/

Chomsky, Noam and Morris Halle. 1968. *The sound patterns of English*. MIT Press: Cambridge, Massachusetts.

Chow, Ivan and François Poiré. 2007a. Rhotic variation and schwa epenthesis in Windsor French. Interspeech 2007, 27-31 August, Antwerp, Belgium, 1022-1025.

Chow, Ivan and François Poiré. 2007b. Consonant-Rhotic Clusters in South-Western Ontario: A Study of Rhotic Variation and Schwa Epenthesis. Proceedings of the 33rd. Conference of the Linguistic Association of Canada and the United States (LACUS), Toronto, 93-103.

Colantoni, Laura and Jeffrey Steele. 2007. Voicing-dependent cluster simplification asymmetries in Spanish and French. In P. Prieto, M. J. Solé & J. Mascaró (eds.), *Segmental and prosodic issues in Romance linguistics* (109-129). Amsterdam/Philadelphia: John Benjamins.

Dell, François. 1985. *Les règles et les sons*, 2nd ed. Paris: Hermann.

Durand, Jacques. 1987. On the phonological status of glides: the evidence from Malay. In *Explorations in Dependency Phonology* (John Anderson and Jacques Durand, eds.), Dordrecht, Holland: Foris Publications. 79-107.

Durand, Jacques and Chantal Lyche. 1999. Regard sur les glissantes en français: français standard, français du Midi. In *Cahiers de grammaire* 24, Phonologie: théorie et variation. Ed. Nicole Serna. ERSS-Université de Toulouse-Le Mirail,Toulouse.

Durand, Jacques, Bernard Laks and Chantal Lyche. 2002. La phonologie du français contemporain: usages, variétés et structure. In: C. Pusch & W. Raible (eds.) *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen: Gunter Narr Verlag, pp. 93-106.

Durand, Jacques, Bernard Laks and Chantal Lyche (éds) (2002) - Protocole, Conventions et directions d'analyse. Bulletin PFC 1. http://www.projet-pfc.net/bulletins-et-colloques/cat_view/918-bulletins-pfc/919-bulletin-pfc-nd1.html

Durand, Jacques and Chantal Lyche. 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, Jacques. 2006. Mapping French Pronunciation: the PFC Project. In Montreuil, Jean-Pierre Y. (ed.), *New Perspectives on Romance Linguistics*: Vol. II: Phonetics, Phonology and Dialectology. pp. 65–82

Durand, Jacques, Bernard Laks and Chantal Lyche. 2009. Le projet PFC: une source de données primaires structurées. In J. Durand, B. Laks et C. Lyche (eds)(2009) *Phonologie, variation et accents du français*. Paris: Hermès. pp. 19-61.

Encrevé, Pierre. 1988. *La liaison avec et sans enchaînement*: *Phonologie tridimensionnelle et usages du français*. Paris: Seuil.

Fouché, Pierre. 1952. *Phonétique Historique du Français*. Paris: Klincksieck.

Fouché, Pierre. 1959. *Traité de prononciation française*. Paris: Klincksieck.

Fischer-Jørgensen, Eli. 1975. *Trends in Phonological Theory*. Akademisk Forlag: Copenhagen.

Gilliéron, Jules and Edmond Edmont 1902-1912. *Atlas linguistique de la France*. Paris: Champion.

Gleason, H. A. 1955. *An Introduction to Descriptive Linguistics*. New York: Holt, Rinehart and Winston.

Goad, Heather. 2006. Are children's grammars rogue grammars? Glide substitution in branching onsets. *Recherches linguistiques de Vincennes* 35: 103-132.

Gougenheim, George. 1935. *Eléments de phonlogie française*. Paris : Les Belles Lettres.

Grammont, M., 1933. *Traite de phonetique*. Paris: Delagrave.

Greenberg, Joseph H. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In Greenberg, Joseph H. (ed.), Universals of Human Language, 73-113. Cambridge, Mass: MIT Press.

Greenberg, Joseph H. 1966a. Some Universals of Grammar with special reference to feature hierarchies. In Greenberg, Joseph H. (ed.), *Universals of Human Language*, 73-113. Cambridge, Mass: MIT Press.

Greenberg Joseph H. .1966b. Synchronic and diachronic universals in phonology. Language 42 (2): 508–17.

Guerssel, M., 1986. Glides in Berber and Syllabicity. *Linguistic Inquiry* 17, 1-12.

Hall, T. A., 2006. Derived environment blocking effects, In *Optimality Theory. Natural Language and Linguistic Theory* 24.3: 803-856.

Hannahs, S.J., 1995. Glide formation, prefixation, and the phonological word in French. In Jon Amaste, Grant Goodall, Mario Montalbetti, Marianne Phinney (eds.). Contemporary research in Romance linguistics, 13-22. Amsterdam/Philadelphia: John Benjamins.

Heap, David, Terry Nadasdi and Jeff Tennant. 1992. Élision et semi-voyelles en français : représentations sous-jacentes et variation. In Ph. Martin, ed., *Mélanges Léon : Hommages à Pierre Léon*. Éditions Mélodie : Toronto, 165-187.

Inkelas, Sharon and Yvan, Rose. 2008. Positional neutralization: A case study from child language. Language 83(4): 707-736.

International Phonetics Association (IPA): http://www.langsci.ucl.ac.uk/ipa/index.html

Jakobson, Roman, C. Gunnar M. Fant, and Morris Halle. 1952. *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge: MIT

Jakobson, Roman and Halle, Morris. 1956. Fundamentals of language. The Hague: Mouton.

Jakobson, Roman. 1963. Language Universals and Linguistics. In Universals of language. MIT Press, ed. Joseph H. Greenberg : Cambridge, Massachusetts.

Jones, Daniel. 1932. *An Outline of English phonetics*. Cambridge: Heffer & Sons.

Juilland, Alphonse. 1965. *Dictionnaire Inverse de la langue Française*. Mouton and Co.: Paris.

Kaye, Jonathan and Jean Lowenstamm. 1984. De la syllabicité, in F. Dell, D.Hirst and J.-R. Vergnaud (eds.), *Forme sonore du langage, Structure des représentations en phonologie*, Hermann, Paris, pp. 123–159.

Klein, Marc, 1991. Vers une approche substantielle et dynamique de la constituance syllabique. Le cas des semi-voyelles et des voyelles hautes dans les usages parisiens. Thèse de doctorat, Saint Denis, Université de Paris VIII.

Klein, Marc, 1993. La syllabe comme interface de la production et de la réception phoniques, in B. Laks & M. Plénat (éds), *De Natura Sonorum*. Essais de phonologie, Saint Denis, Presses Universitaires de Vincennes, pp. 101-141.

Ladefoged, Peter. 1982. *A Course in Phonetics*. New York: Harcourt.

Lanly, André. 1971. *Fiches de philologie française*. Paris : Bordas.

Léon, Pierre. R., 1978. *Prononciation du français standard*. Paris : Didier.

Levi, Susan. 2004. The representation of underlying glides: a cross-linguistic study. Doctoral dissertation. University of Washington.

Lyche, Chantal. 1979. Glides in French: Questions for natural generative phonology. In *Lingua* 49 (4), 315-330.

Lyche, Chantal, and Francine Girard. 1995. Le mot retrouvé. *Lingua* 95: 205–221

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge University Press: London.

Martinet, André. 1945. *La prononciation du français contemporain*. Droz: Paris.

Martinet, André. 1949. *Phonology as Functional Phonetics*. Oxford University Press: London.

Martinet, André. 1960. *Elements of General Linguistiques*. Faber and Faber: London.

Martinet, André and Henriette Walter. 1973. *Dictionaire de la prononciation française dans son usage réel.* France-Expansion: Paris.

Morin, Yves-Charles. 1976. Phonological Tensions in French. In *Current studies in Romance linguistics*. Marta Lujan and Fritz Hensey Editors. Washington: Georgetown University Press. pp. 37-49.

Nespor Marina and Irene Vogel.1986. *Prosodic Phonology*. Dordrecht: Foris.

Noske, Roland. 1993. A Theory of Syllabification and Segmental Alternation, With Studies on the Phonology of French, German, Tonkawa and Yawelmani, Niemeyer, Tübingen.

Nyrop, K. 1899-1930: Grammaire historique de la langue française. I-VI. Copenhague/Leipzig/Paris: Nordisk forlag (Gyldendal)/Otto Harrassowitz/Alphonse Picard & Fils.

Passy, Paul. 1913. *Les sons du français*. Paris: Didier.

Pike, Kenneth L. 1943. *Phonetics: a critical analysis of phonetic theory and a technic for the practical description of sounds*. Ann Arbor: The University of Michigan Press.

Pike, Kenneth L., 1947. *Phonemics: a technique for reducing languages to writing*. University of Michigan publications. Linguistics, 3. Ann Arbor: University of Michigan Press.

Pike, Kenneth L., and Pike, Evelyn V. 1947. Immediate Constituents of Mazateco Syllables. IJAL 13.78-91 (1947).

Poiré, François, Claire Gurski and Stephanie Kelly. 2007. L'analyse phonétique à partir de petit corpus de parole spontanée : le cas des glissantes en français de Windsor, Canada. *Bulletin PFC* no 7, CLLE-ERSS, UMR 5263, CNRS et Université Toulouse Le Mirail. pp. 341-358.

Prince, Alan and Paul, Smolensky. 1993. *Optimality Theory*, Rutgers University and University of Colorado.

Rey-Debove, Josette and Alain Rey. 1967/1993. Le nouveau Petit Robert. Dictionnaires le Robert: Paris.

Reynolds and Nagy, Naomi. (1997). Optimality Theory and variable word-final deletion in Faetar. In Language Variation and Change 9: 37-55 Cambridge University Press.

Rialland, Annie. 1994. The Phonology and Phonetics of Extrasyllabicity in French. In P.A. Keating (ed.), *Phonological Structure and Phonetic Form*: Papers in Laboratory Phonology III, Cambridge University Press, Cambridge, pp. 136–159.

Rice, Keren. 2007.Markedness. In The Cambridge handbook of phonology. Ed. Paul de Lacy. Cambridge University Press. New York.

Rosenthall, S., 1997. Vowel/Glide Alternation in a Theory of Constraint Interaction, New York, Garland.

de Saussure, Ferdinand. 1982. Cours de linguistique générale. Paris: Payot.

Schane, Sandford. A., 1968. *French Phonology and Morphology*. Cambridge (Mass.): MIT Press.

Selkirk, Elizabeth. 1982. The phrase phonology of English and French. Doctoral Dissertation. MIT. IULC.

Straka, Georges.1964. A propos de la question des semi-voyelles. *Zeitschrift fuer Phonetik Sprachwissenschaft und Kommunikationsforschung* 17. 301-23.

Tranel, Bernard. 1987. *The Sounds of French*. New York: Cambridge University Press.

Troubetzkoy, Nicolai Sergeevitch. 1964. *Principes de phonologie*. Paris: Klincksieck,

Walker, Douglas. 2001. *French sound structure*. Calgary: University of Calgary Press.

Warnant, Léon. 1987. *Dictionnaire de la prononciation française dans sa norme actuelle*. Paris: Éditions Ducoulot.

Appendices

Appendix A: PFC Reading Task with glide glide contexts studied in (Kelly 2005).

Le Premier Ministre ira-t-il à Beaulieu?

Le village de Beaulieu est en grand émoi. Le Premier Ministre a en effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises en soie, un champion local de course à $_{1}$pied (Louis Garret), $_{2}$quatrième aux jeux olympiques de Berlin en 1936, et plus récemment, son usine de pâtes $_{3}$italiennes. Qu'est-ce $_{4}$qui a donc $_{5}$valu à Beaulieu ce grand honneur? Le hasard, tout bêtement, car le Premier Ministre, lassé des circuits $_{6}$habituels qui tournaient toujours autour des mêmes villes, veut découvrir ce qu'il appelle "la campagne profonde".

Le maire de Beaulieu - Marc Blanc - est en revanche très $_{7}$inquiet. La cote du Premier Ministre ne cesse de baisser depuis les élections. Comment, en plus, éviter les manifestations $_{8}$qui ont eu tendance à $_{9}$se multiplier lors des visites $_{10}$officielles ? La côte escarpée du Mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les opposants de tous les bords manifestent leur colère. D'un autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout est fait pour le protéger. Or, un gros détachement de police, comme on en a $_{11}$vu à Jonquière, et des vérifications d'identité risquent de provoquer une explosion. Un jeune membre de l'opposition aurait déclaré: "Dans le coin, on est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous avons le $_{12}$soutien du village $_{13}$entier." De plus, quelques articles parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des activistes des communes voisines préparent une journée chaude au Premier Ministre. Quelques fanatiques auraient même entamé un jeûne prolongé dans l'église de Saint Martinville.

Le sympathique maire de Beaulieu ne sait plus à quel saint $_{14}$se vouer. Il a le sentiment de se trouver dans une impasse stupide. Il s'est, en désespoir de cause, décidé à écrire au Premier Ministre pour $_{15}$vérifier si son village était vraiment une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver au centre d'une bataille politique dont, par la télévision, seraient témoins des millions d'électeurs.

Texte PFC (© Projet PFC)

Appendix B: Instructions to judges for Judgement task.

À lire (et même imprimer) avant de commencer.

Merci de participer dans cette tâche de jugement.

Pour commencer vous êtes prié.e d'installer la version la plus récente de Praat (5.1.41) que vous trouverez en format Mac et Windows sur le site officiel au lien suivant : http://www.fon.hum.uva.nl/praat/

Pour faire dérouler cette tâche il va falloir une heure et demie environ. Vu le temps qu'il exige, la tâche a été séparée en trois sessions qui durent 30 minutes environ chacune (ce sont les trois documents .txt titrés « MFC5 votre nom1-2-3 »).

Vous pouvez faire les sessions dans l'ordre de votre choix, mais svp lisez les instructions suivant avant de commencer.

- Ouvrir le logiciel Praat.
- Cliquez sur « Read » et sélectionnez le document MFC5 votrenom.txt 1, 2 ou 3. Le document apparaîtra sur « Praat Objects ». Cliquez sur « Run » pour faire rouler la tâche. À cette étape, ça prendra un peu du temps à lire les données de chaque document (260 environ à chaque reprise). Lorsque la tâche est prête, vous allez voir une fenêtre à plein écran. Continuez en suivant les instructions fournies.
- Lorsque vous avez fini la tâche, vous pouvez maintenant fermer l'écran de la tâche, mais SVP ne fermez pas le logiciel Praat avant de **sauvegarder les résultats de votre travail pour chaque session**!
- Sur le menu à droite de « Praat Objects » cliquez sur « Extract results ». Vous verrez ensuite un nouvel objet à gauche sur « Praat Objects », c'est le fichier « ResultsMFC MFCvotrenom1, 2 ou 3 ». Cliquez sur le bouton « Write » et choisissez « Write to text file ». Vous pouvez prendre le temps qu'il faut avant de continuer avec une autre session.
- Lorsque vous êtes prêt.e, continuez avec les deux autres fichiers mais à chaque fois ne fermez pas le logiciel Praat avant de **sauvegarder les résultats de votre travail** !
- Lorsque vous avez fini une session vous êtes prié.e de m'envoyer le document « ResultsMFC votrenom1-2 ou 3 » par courriel.
- MERCI :D

Appendix C: PFC - Fiche signalétique

NOM, prénom(s) : ....

Date de naissance : ....

Lieu de naissance : ....

Domiciles successifs (en nombre d'années) :

....

Domicile actuel :..

Professions successives:..

Profession actuelle : ....

Situation familiale (marié, célibataire, etc.) :..

enfants, âge, scolarisation : ....

Études (préciser jusqu'à quel âge et quel type d'études) :

....

Langues parlées :..

Père de l'informateur, année de naissance :..

lieu d'origine : ....

profession : ....

études : ....

langues parlées (étrangères ou régionales) : ....

Mère de l'informateur, année de naissance :..

lieu d'origine : ....

profession : ....

études : ....

langues parlées (étrangères ou régionales) : ....

Époux, épouse, autre : ....

lieu d'origine : ....

profession : ....

études : ....

langues : ....

Personnes ayant joué un rôle important au moment de l'apprentissage du français par l’enquêté (grands-parents, nourrice...) :

....

Type de logement de l’enquêté (maison, appartement...) :...

Intégration dans le quartier, relations de voisinage :...

Activités culturelles, loisirs,voyages:..

Autres informations :

....

....

Informations sur l’enquête

Nom de l’enquêteur (entretien guidé) :..

Nom de l’enquêteur (entretien libre) :..

Date de l'enregistrement, durée :..

Lieu de l'enregistrement :..

Localité :..

Lien enquêteur/enquêté :..

Professionnel :...

Amical :..

Familial : ....

Autre : ....

Observations sur l’enregistrement :...

Informel : qui étaient les locuteurs présents ?

....

Ordre des situations dans l’enregistrement (ex. guidé, liste de mots, texte, libre) :

....

Principaux thèmes abordés :

....

Qualité de l’enregistrement : ....

Autres observations: (interventions d’autre personnes, interruptions téléphoniques notables...) ....

....

Enregistrement de groupe

Établir une fiche signalétique pour chaque participant et préciser les relations à l'intérieur du groupe.

Curriculum Vitae

Name:	Stephanie Kelly
Post-secondary Education and Degrees:	The University of Western Ontario London, Ontario, Canada 1998-2002 B.A.
	The University of Western Ontario London, Ontario, Canada 2002-2003 M.A.
	The University of Western Ontario London, Ontario, Canada 2004-2014 Ph.D.
Honours and Awards:	Social Science and Humanities Research Council Canadian Graduate Scholarship 2005-2008
Related Work Experience	Teaching Assistant The University of Western Ontario 2004-2008

Publications:

Kelly, Stephanie. 2004. Theoretical implications of French nuclear diphthongisation. Linguistica Atlantica. Thematic Issue: Explanatory Adequacy. Sous la direction de Nick Bibis, Anna Moro et Alexandre Sévigny, 59 à 76.

Poiré, François, Stephanie Kelly et Darcie Williams. 2007. La réalisation des voyelles nasales en français de Windsor. REVUE PArole, Belgique, 259 à 282.

Kelly, Stephanie, Darcie Williams and François Poiré. 2007. Nasal Vowels in Windsor French, 33rd Conference of the Linguistic Association of Canada and the United States, submitted to the Proceedings of the 33rd Conference of the Linguistic Association of Canada and the United States Toronto, August 2006, 317 à 327.

Kelly, Stephanie (2010). Conversation à Québec ville (Québec, Canada): la religion pour un jeune Québecois. In S. Detey, J. Durand, B. Laks & C. Lyche (éds) Le français parlé contemporain dans ses variétés: ressources pour l'étude du français, Paris: Ophrys.

Poiré, François, Claire Gurski & Stephanie Kelly. 2007. L'analyse phonétique à partir de petit corpus de parole spontanée : le cas des glissantes en français de Windsor, Canada. Bulletin PFC no 7, CLLE-ERSS, UMR 5263, CNRS et Université Toulouse Le Mirail.

pp. 341-358.

Kelly, Stephanie. 2007. Prononciation et autres aspects du français au Québec. Bulletin PFC no 7, CLLE-ERSS, UMR 5263, CNRS et Université Toulouse Le Mirail. pp. 205-216.