

# Paradigm uniformity effects on French liaison\*

Benjamin Storme

June 30, 2023

## Abstract

French liaison is a type of external sandhi involving the use of a special consonant-final allomorph before vowel-initial words. Consonants occurring at the end of these allomorphs are challenging for phonological theory because of evidence that their prosodic and segmental realization is intermediate between the realizations of word-final and word-initial consonants. This puzzling behavior of French liaison has been used to motivate new phonological and lexical representations, including floating consonants, lexical constructions and gradient symbolic representations. This paper proposes an alternative analysis: the variable realization of liaison is derived as a paradigm uniformity effect, assuming traditional phonological and lexical representations. In a Word1-Word2 sequence, the liaison consonant at the boundary between the two words ends up acquiring properties of both word-final and word-initial consonants because of a pressure to make contextual variants of Word1 and Word2 similar to their citation forms. The proposal is implemented in a probabilistic constraint-based grammar including paradigm uniformity constraints and is shown to account for the intermediate behavior of liaison both in terms of prosodic attachment and segmental realization. The paper provides evidence for two key predictions of this analysis, using judgment data on the prosodic attachment of liaison consonants in European French and phonetic data on the interaction between liaison and affrication in Quebec French.

Keywords: French liaison; paradigm uniformity; phonetic detail; constraint-based grammar

---

\*This paper greatly benefited from the feedback of two anonymous reviewers and two NLLT editors, Arto Anttila and Michael Kenstowicz. Parts of this work were presented at the *2019 Annual Meeting on Phonology*, at *Going Romance XXIV* in 2020, at the *Séminaires de Recherches en Phonétique et Phonologie (SRPP)* in 2021, and at UCL in 2022. Among attendants at these talks, I am particularly grateful to Birgit Alber, Cécile Fougeron, and Rachid Ridouane for very helpful suggestions. Special thanks go to: Giorgio Magri (for helping me with CoGeTo), Marie-Hélène Côté (for giving me access to the corpus used in Study 2), Aris Xanthos (for writing the Python script to obtain lexical frequencies from the OpenSubtitles corpus in Study 1), Marjolein Hageman, Liselotte Masdupuy, and Stéphane Térosier (for their help with Study 1). All supplementary files accompanying the paper may be found in [Storme \(2022\)](#).

# 1 Introduction

French liaison is a type of external sandhi involving the use of a special consonant-final allomorph before vowel-initial words. The consonant at the end of this allomorph is called a liaison consonant. For instance, the adjective *grand* ‘great’ is generally realized as [gʁɑ̃], as shown in (1a) and (1b), but may appear under its liaison form [gʁɑ̃t] before vowel-initial words, as shown in (1c).

(1) French liaison

	Word1	Word2		Context
a.	<i>grand</i>	[gʁɑ̃]	(none)	‘great’ (citation form)
b.	<i>grand monsieur</i>	[gʁɑ̃]	[mɔ̃sjø]	‘great man’ (before C-initial words)
c.	<i>grand ami</i>	[gʁɑ̃t]	[ami]	‘great friend’ (before V-initial words)

Liaison is a complex phenomenon that is influenced by a range of language-internal and -external variables beyond the basic phonological conditioning described in (1). Due to this complexity, liaison has featured prominently in many theoretical debates over the last decades, including debates on the syntax-phonology interface, the nature of phonological and lexical representations, and the role of lexical frequency in phonology (see [Côté 2011](#) for an overview).

In recent years, there has been renewed interest in a particular challenge that French liaison raises for phonological theory. Liaison consonants vary in their pattern of realization. More specifically, their prosodic and segmental realization is intermediate between the realization of word-final and word-initial consonants. For instance, in the presence of an audible pause between a liaison word and the following word, liaison consonants can be attached either at the end of the liaison allomorph, like word-final consonants (= *liaison non-enchaînée*), or at the beginning of the following word, like word-initial consonants (= *liaison enchaînée*; [Encrevé 1988](#); [Durand & Lyche 2008](#)). This variability in prosodic attachment is illustrated in (2). Prosodic domains are indicated using parentheses.

(2) Variable realization of French liaison across a prosodic break

*un grand, ami* ‘a great, friend’

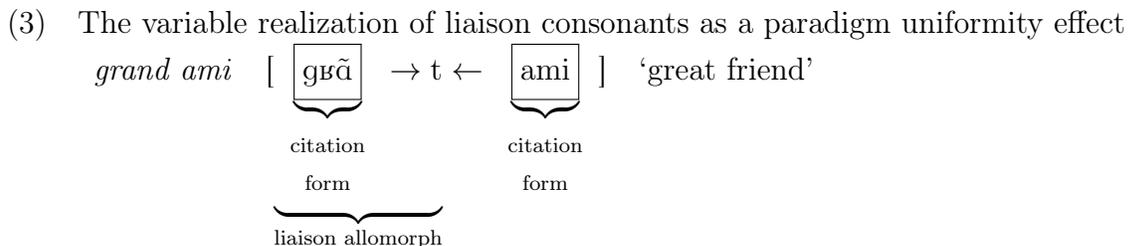
- a. (ẽgʁɑ̃t) (ami)
- b. (ẽgʁɑ̃) (tami)

This variability has led some researchers to propose special underlying phonological representations for liaison consonants, including floating segments ([Encrevé 1988](#); [Tranel 1990](#)) and gradient symbolic representations ([Smolensky & Goldrick 2016](#); [Smolensky et al. 2020](#); [Tessier & Jesney 2021](#)). It has also been used to argue against the traditional view according to which liaison consonants are lexically affiliated to Word1, either by positing that they belong to a lexical construction involving both Word1 and Word2 ([Bybee 2001](#)) or that they are independently affiliated to both Word1 and Word2 ([Smolensky & Goldrick 2016](#)).

This paper proposes another account: the variable behavior of liaison consonants is not captured through an enriched phoneme inventory or lexicon. Instead it emerges from the structure of the lexicon as a paradigm uniformity effect, similar to what has been proposed to account for incomplete devoicing in German ([Roettger et al. 2014](#)). More specifically, this paper builds on a hypothesis proposed by [Steriade \(1999\)](#) and according to which the liaison

allomorph of a word (e.g. [gβãt] in (1c)) is required to be faithful to the corresponding citation form, i.e. the word as pronounced in isolation (e.g. [gβã] in (1a)). Crucially, the liaison consonant is typically absent from the citation form. In a Word1-Word2 sequence, the requirement to be faithful to the citation form of Word1 will push the liaison consonant away from the end of this word onto the following word, therefore favoring a word-initial behavior at the beginning of Word2.

The present paper extends Steriade’s original analysis by hypothesizing that the realization of Word2 is also subject to paradigm uniformity effects, as proposed by Zuraw & Hayes (2017) in their analysis of French *h-aspiré*. The requirement to be faithful to the citation form of Word2 will push the liaison consonant away from the beginning of this word, therefore favoring a word-final behavior at the end of Word1. These two opposite uniformity effects, represented with arrows going in opposite directions in (3), are proposed to underlie the variable realization of liaison consonants. The forms enclosed in boxes in (3) correspond to the citation forms of the two words involved in the sequence *grand ami*. The hypothesis that the citation form of Word2 also plays a role will be crucial to explain why liaison consonants do not behave identically with word-initial consonants but also share properties with word-final consonants, both prosodically and segmentally.



Section 2 summarizes the evidence for the variable prosodic and segmental realization of French liaison. Section 3 implements the paradigm uniformity analysis schematically represented in (3) in a probabilistic constraint-based grammar with paradigm uniformity constraints and shows how this analysis can derive the *prosodic* variability of liaison. Section 4 presents the results of an experimental study that both confirms the prosodic variability of liaison and provides evidence for the role of citation forms in this sound pattern. The evidence comes from a comparison of two types of liaison allomorphs differing in their similarity with the corresponding citation forms.

Section 5 shows how the analysis can be extended to account for the *segmental* variability of French liaison. This analysis builds on Steriade (2000)’s proposal that phonetic detail such as consonant-vowel coarticulation matters in paradigm uniformity effects. Section 6 reports on a phonetic study looking at the interaction of affrication and liaison in Quebec French, using data from the *Phonologie du Français Contemporain* (PFC) project. The results of this study provide evidence for the phonetic mechanism that is proposed to underlie the variable segmental realization of liaison in Quebec French.

Section 7 compares the paradigm uniformity analysis with previous analyses and highlights some of the strengths and weaknesses of these different approaches. In a nutshell, the paradigm uniformity analysis presented in this paper is shown to provide the most comprehensive account of the *realization* of liaison, as it is currently the only analysis that derives both types of variabilities (prosodic and segmental). However it is insufficient to account for

a range of effects on the *rate* of liaison. This section sketches how this problem can be remedied by indexing paradigm uniformity constraints to properties that are relevant to speech production (e.g. lexical frequency, contextual predictability), according to the production planning theory of sandhi phenomena (Tanner et al. 2017; Kilbourn-Ceron 2017).

## 2 Background on the realization of French liaison

The liaison allomorph of a word is used when the following word starts with a vowel, as illustrated in (1c). However the allomorph without liaison is still available in this context. Most of the work on French liaison has been dedicated to understanding which factors shape the distribution of the two allomorphs prevocally (see Section 7 for more discussion on this topic).

This paper focuses on yet another important question in the literature on French liaison: how liaison consonants are realized when pronounced. As will be further reviewed in this section, liaison consonants pattern in a very puzzling way, as they share prosodic and segmental properties with stable word-final as well as word-initial consonants. Stable consonants differ from liaison consonants in being present regardless of the surrounding phonological context. For instance, word-initial [t] and word-final [t] in *trente* [tʁɛ̃t] ‘thirty’ are stable because they are present regardless of the nature of adjacent segments in neighboring words. Section 2.1 focuses on the prosodic variability of liaison and Section 2.2 on its segmental variability.

### 2.1 Prosodic variability

This section assumes the prosodic organization of French into three types of phrases proposed by Delais-Roussarie et al. (2015: 65-74): accentual phrase (AP), intermediate phrase (ip), and intonational phrase (IP). This proposal serves as a convenient descriptive framework to present the prosodic realization of liaison but is not key to the analysis.

The accentual phrase (AP) is the smallest prosodic constituent and includes minimally a content word along with all function words that depend on it syntactically, such as the AP *il est devenu* ‘he has become’ in (4). It can also include other content words that are construed as modifiers, such as the adjective *grand* ‘great’ in the AP *mon grand ami* ‘my great friend’ in (4). The intermediate phrase (ip) is a larger prosodic constituent that includes minimally one AP. Dislocations such as *le voisin* ‘the neighbour’ in (4) form their own ip. Finally, the intonational phrase (IP) is the largest prosodic constituent and is often followed by a pause. In (4), it corresponds to the whole sentence.

- (4) *Il est devenu mon grand ami, le voisin.*  
 $\{ \{ (Il\ est\ devenu)_{AP} (mon\ grand\ ami)_{AP} \}_{ip} \{ (le\ voisin)_{AP} \}_{ip} / IP$   
 ‘He has become my great friend, the neighbour.’

The prosodic behavior of liaison depends on the position of liaison words in these prosodic domains. Within an AP, liaison consonants are syllabified with the following vowel, as are stable word-final and word-initial consonants (Spinelli et al. 2002; Gaskell et al. 2002; Durand & Lyche 2008: section 3.3). For instance, in  $(un\ grand\ ami)_{AP}$  [ɛ̃ɡʁɑ̃tami] ‘a great

friend’, the liaison consonant is coarticulated with the vowel at the beginning of the following word. The same happens with stable word-final consonants, as in (*mes trente amis*)<sub>AP</sub> [metʁãtami] ‘my thirty friends’, and with stable word-initial consonants, as in (*un grand tamis*)<sub>AP</sub> [ẽgʁãtami] ‘a big sieve’. For word-final consonants and liaison consonants, syllabification with the following vowel is described as resyllabification or as *enchaînement* in French linguistic terminology (Encrevé 1988).

When a prosodic boundary intervenes between Word1 and Word2, liaison consonants pattern differently from stable consonants: although lexically dependent on the identity of Word1,<sup>1</sup> the liaison consonant will typically behave like a word-initial consonant prosodically and attach to Word2. In this case, it is described as a *liaison enchaînée*. But it may also attach to Word1, in which case it is described as a *liaison non-enchaînée*. By contrast, stable final and initial consonants remain prosodically affiliated to their lexical hosts (Word1 and Word2, respectively) and do not show prosodic variability.

This puzzling prosodic behavior of French liaison can be illustrated with Word1-Word2 sequences in right dislocations (Tranel 1990; Côté 2005). Right dislocated elements belong to a distinct prosodic unit from their nucleus sentence: their prosody copies the prosody of the nucleus but is characterized by decreased intensity, lower pitch and a flat intensity contour (De Cat 2007: 34-43). In Delais-Roussarie et al. (2015)’s prosodic analysis, right dislocation involves two distinct intermediate phrases (ip), as illustrated in (4). When Word2 is right dislocated, the liaison consonant is separated from the word it is lexically affiliated to (Word1) and attaches at the beginning of the following word (Word2) across the prosodic break. For instance, in (5a), liaison [t] is separated from its lexical host [gʁã] by a prosodic boundary between two intermediate phrases (ip) and attaches to [elefã]. By contrast, stable consonants remain prosodically attached to their lexical hosts, as shown in (5b) for word-final consonants (they attach to Word1) and in (5c) for word-initial consonants (they attach to Word2).

- (5) Liaison vs. stable consonants in right dislocations
- |    |                        |  |
|----|------------------------|--|
| a. | Liaison consonant      | <i>J’en ai un grand, éléphant.</i><br>(ʒãneẽgʁã) <sub>ip</sub> (tɛlefã) <sub>ip</sub><br>‘I have a big one, elephant.’     |
| b. | Word-final consonant   | <i>J’en ai trente, éléphants.</i><br>(ʒãnetʁãt) <sub>ip</sub> (elefã) <sub>ip</sub><br>‘I have thirty of them, elephants.’ |
| c. | Word-initial consonant | <i>J’en ai un beau, tabouret.</i><br>(ʒãneẽbo) <sub>ip</sub> (tabuʁɛ) <sub>ip</sub><br>‘I have a beautiful one, stool.’    |

Although Word2-attachment of liaison consonants is reported as the preferred option in the presence of a prosodic break, liaison consonants may still attach at the end of Word1. In other words, the prosodic realization of liaison is variable. When the liaison consonant

---

<sup>1</sup>The liaison consonant is lexically dependent on Word1 because the choice of a specific liaison consonant among the set of potential liaison consonants (i.e. /t z n ʁ/) covaries only with Word1, at least in adult speech (see Chevrot et al. 2009 on child speech). For instance, the liaison consonant is /t/ in the sequence *grand ami* ‘great friend’ but it becomes /n/ if *grand* is replaced by the indefinite determiner *un* in the sequence *un ami* ‘a friend’.

is attached at the end of Word1, the liaison consonant is said to be *non-enchaînée*. The availability of *liaison non-enchaînée* was famously described by [Encrevé \(1988\)](#), using a corpus of political speeches by prominent French politicians. Encrevé’s data are summarized in Table 1. The data show that the liaison allomorph is used in about half of the potential sites for liaison (49%). When present, the liaison consonant behaves either as a word-initial segment at the beginning of Word2 (= *liaison enchaînée*; this is the most common pronunciation) or as a final segment at the end of Word1 (= *liaison non-enchaînée*).

		Count	Frequency
Liaison present	<i>enchaînée</i>	2488	43%
	<i>non-enchaînée</i>	321	6%
Liaison absent		2978	51%

Table 1: Prosodic variability of French liaison in Encrevé’s corpus ([Encrevé 1988](#): 56)

An example of *liaison non-enchaînée* by the late French president Jacques Chirac is provided in (6) (cited from [Durand & Lyche 2008](#): 51). This example can be found on the YouTube channel of the *Institut national de l’audiovisuel* (INA) (<https://youtu.be/t72zuDsSHAw?t=289>; 4’49”). In this example, *font* ‘do.3pl’ and *honneur* ‘honor’ are not phrased together prosodically, as they normally would be, but belong to two distinct phrases. For concreteness, these two phrases are treated here as APs. The example in (6) provides a case of *liaison non-enchaînée*: liaison [t] in *font* is pronounced in the same prosodic phrase as Word1 and not at the beginning of Word2 in the following prosodic phrase. *Liaison enchaînée* would also be possible in this case, as reported by [Encrevé \(1988\)](#): the pronunciation would then be [(ilfɔ̃)<sub>AP</sub> (tɔ̃nɔ̃ɛʁ)<sub>AP</sub>].

- (6) Example of *liaison non-enchaînée* by Jacques Chirac ([Durand & Lyche 2008](#): 51)  
*(Ils font)<sub>AP</sub> (honneur)<sub>AP</sub> (à la République)<sub>AP</sub>*  
 (ilfɔ̃t)<sub>AP</sub> (ɔ̃nɔ̃ɛʁ)<sub>AP</sub> (alɑʁepyblik)<sub>AP</sub>  
 ‘They are a credit to the Republic.’

Although *liaison non-enchaînée* has sometimes been described as uniquely confined to high register and planned speech, [Durand & Lyche \(2008: 50-51\)](#) found examples occurring in natural daily interactions, in particular in the presence of prosodic breaks involving hesitations, as shown in (7a), or repetitions, as shown in (7b). [Durand & Lyche \(2008\)](#) write: ‘These examples seem to us extremely interesting: despite the clear predominance of *liaison enchaînée* in our corpus, they provide possible evidence against an analysis which simply treats a liaison consonant as an onset of Word2.’ This behavior is not reported for stable word-initial consonants at the beginning of Word2: these are not allowed to be attached at the end of Word1 across a prosodic break occurring in the middle of a Word1-Word2 sequence (see Section 4 for experimental evidence).

- (7) Examples of *liaison non-enchaînée* in conversational speech ([Durand & Lyche 2008](#): 50-51)

- a. *vingt euh* (hesitation)  
 (vɛ̃t)<sub>AP</sub> œ  
 ‘twenty uh’
- b. *un, un Aveyronnais* (repetition)  
 (œ̃n)<sub>AP</sub> (œ̃naverone)<sub>AP</sub>  
 ‘an, an Aveyronese man’

## 2.2 Segmental variability

The preceding section has provided evidence for the prosodic variability of liaison. This section shows that variability extends to the *segmental* realization of liaison. In the absence of a prosodic break between Word1 and Word2, the distinction between liaison consonants, stable word-final consonants and stable word-initial consonants is not completely neutralized. There remain segmental cues that distinguish the three types of consonants, as will be reviewed in this section. In particular, liaison consonants have a segmental realization that is intermediate between that of stable word-final and word-initial consonants.

Fougeron (2007) showed that word-final consonants before a vowel (VC#V) do not have the same acoustic realization as word-initial consonants (V#CV), even in contexts that are traditionally treated as involving *enchaînement*/resyllabification of the final consonant. In particular, she found that word-final consonants tend to be shorter than word-initial consonants (Fougeron 2007: 13).

Liaison consonants are also reported to behave distinctly from both stable word-final and word-initial consonants phonetically in this context. For instance, an early study by Durand (1936: 238) found that stable word-final consonants (e.g. the final [t] of *petite* [pətit] ‘small.FEM’ in *une petite orange* ‘a small orange’) differ from liaison consonants (e.g. liaison [t] at the end of *petit* [pətit] ‘small.MASC’ in *un petit orage* ‘a small storm’) in retaining some cues of their implosive/non-prevocalic nature. Liaison consonants have also been found to differ from word-initial consonants and in particular to be characterized by a shorter duration on average (Gaskell et al. 2002: 805, Spinelli et al. 2002: 93, Spinelli et al. 2003: 239). For instance, Spinelli et al. (2003: 239) report that liaison consonants are 17% shorter than initial consonants. Taken together, these studies suggest that, on average, liaison consonants have a phonetic realization that is intermediate between that of stable word-final and word-initial consonants.

More targeted studies found that this effect actually depends on the nature of the consonant, as summarized in Table 2. Word-initial /t/ was found to be longer than word-final /t/ and liaison /t/, but no significant durational difference was found between word-final /t/ and liaison /t/. For /z/, a study by Nguyen et al. (2007) found a shorter duration word-finally than word-initially or in a liaison context. But this result was not replicated by Bagou et al. (2009). Neither study found a significant durational difference between word-initial and liaison /z/. For /n/, phonetic realization does not seem to be affected by the lexical status of the consonant, according to available studies: the duration of the consonant does not significantly differ whether the consonant is word-final, liaison or word-initial. Overall, taken together, the liaison consonant appears to pattern intermediately between word-final and word-initial consonants in phonetic realization, with some differences depending on the specific liaison consonant.

	word-final		liaison		word-initial
/t/	dur(C)	= <sup>a</sup>	dur(C)	< <sup>a,d</sup>	dur(C)
/z/	dur(C)	< <sup>b</sup> /= <sup>a</sup>	dur(C)	= <sup>a,b</sup>	dur(C)
/n/	dur(C)	= <sup>b</sup>	dur(C)	= <sup>b,c</sup>	dur(C)

<sup>a</sup>Bagou et al. 2009: 454, <sup>b</sup>Nguyen et al. 2007: 12

<sup>c</sup>Wauquier-Gravelines 1996, <sup>d</sup>Dejean de la Bâtie & Bradley 1995: footnote 2

Table 2: Differences in duration for consonants /t z n/ as a function of lexical type (stable word-final consonant, liaison consonant, stable word-initial consonant).

The clearest evidence for the intermediate segmental realization of liaison consonants comes from data on affrication in Quebec French (Côté 2014). Quebec French has a process of affrication that turns /t d/ into [ts dz] before /i y j ɥ/. But this process affects liaison consonants, stable word-final consonants, and stable word-initial consonants differently, as shown in Table 3. More specifically, liaison /t/ exhibits a rate of affrication that is intermediate between stable word-final /t/ and word-initial /t/: liaison /t/ is more prone to affrication than stable word-final /t/ but less so than stable word-initial /t/. This result was obtained using data from the PFC project (Côté 2016). The presence or absence of affrication was determined perceptually for each token by at least two judges (Côté 2014: 36-37).

Consonant type	Consonant is absent	Consonant is present	
		Affrication	No affrication
Liaison consonant	21 (17.0%)	68 (55.0%)	35 (28.0%)
Word-final consonant		31 (36.5%)	54 (63.5%)
Word-initial consonant		715 (99.2%)	6 (0.8%)

Table 3: Affrication before /i y j ɥ/ in the PFC Trois-Rivières survey: count and frequency data (Côté 2014: 38)

It is important to note that only contexts involving resyllabification/*enchaînement* were included by Côté in the data reported in Table 3. For instance, among the 121 occurrences involving a stable word-final consonant before /i/ in the corpus, only 85 were included in the analysis (Côté 2014: 38). The remaining 36 occurrences did not involve resyllabification/*enchaînement*, i.e. there was a prosodic break (a pause or hesitation) between /t/ and the following /i/<sup>2</sup> and no affrication. This means that the differences in affrication observed in Table 3 cannot be explained away as a by-product of prosodic variability: all contexts involve consonants that would be normally treated as onsets.

## 2.3 Summary

Liaison consonants have a very puzzling behavior in Word1-Word2 sequences. Prosodically, they pattern intermediately between stable word-final and word-initial consonants. In the presence of a prosodic break between Word1 and Word2, liaison consonants may attach at

<sup>2</sup>The rate of *non-enchaînement* (30%) is similar to that reported by Fougeron & Delais-Roussarie (2004) in another corpus (40%).

the end of Word1 (their lexical host), as expected for word-final consonants, but they may also be separated from their lexical host and attach at the beginning of Word2. In contexts that do not involve a prosodic break between the two words and are traditionally treated as onset contexts (i.e. the consonant and the following vowel are articulated together), the segmental realization of liaison consonants has also been found to be intermediate between the realization of stable word-final and word-initial consonants.

### 3 Prosodic variability of French liaison as a paradigm uniformity effect

This section proposes an analysis of the variable behavior of French liaison as a paradigm uniformity effect, focusing first on prosody. More specifically, this variability is proposed to result from a pressure to make contextual variants of Word1 and Word2 in Word1-Word2 sequences similar to the corresponding citation forms. The relevance of the citation form of Word1 for French liaison was originally proposed by Steriade (1999) in her work on lexical conservatism. This section shows that generalizing the uniformity requirement to both Word1 and Word2 predicts that liaison consonants should be variable in their prosodic attachment whereas stable consonants (word-final and word-initial) should only attach to their lexical hosts (Word1 and Word2, respectively).

Section 3.1 describes and motivates the general grammatical architecture that is assumed in the analysis. Section 3.2 shows how the analysis derives citation forms for words with liaison allomorphs and for words that lack such allomorphs. Section 3.3 shows how the prosodic variability of liaison consonants in Word1-Word2 sequences can be derived as a result of paradigm uniformity with the corresponding citation forms, using Encrevé (1988)’s corpus as a case study. Section 3.4 further shows that the asymmetry between liaison and stable consonants is a necessary consequence (an implicational universal) of the proposed constraint set. The analysis can be found in Storme (2022) under the name `prosodic-ambiguity-final-2.txt`.

#### 3.1 Grammatical architecture

The analysis is implemented in a probabilistic constraint-based grammar including input-output (IO) and output-output (OO) faithfulness constraints evaluated in parallel, according to the model in Figure 1. The different components of this model are described and motivated below.

The key ingredient in the analysis is the assumption that there is a family of output-output faithfulness constraints evaluating the similarity among contextual variants of a word (see Benua 1997; Kenstowicz 1996 on output-output faithfulness in general; see Kawahara 2002 for an application of output-output faithfulness to the similarity among contextual variants of a word). Specifically, it is assumed that there are constraints that penalize dissimilarities between connected-speech variants of a word and the corresponding citation form (Steriade 1997: 55-58). Citation forms correspond to the form of a word as pronounced in isolation, with the beginning and end of the word matching the beginning and end of the utterance. Evidence for the role of citation forms in paradigm uniformity comes from data

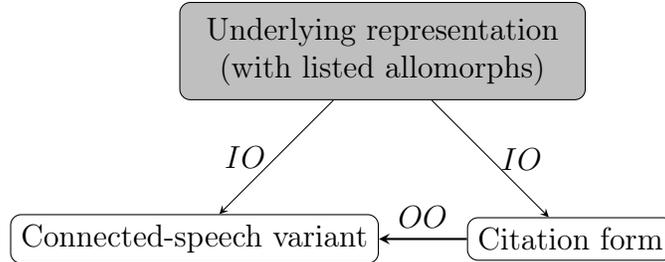


Figure 1: Grammatical model assumed in the analysis (IO = input-output correspondence, OO = output-output correspondence)

		Citation form	Prevocalic form
Liaison words	<i>grand</i> ‘great’	[gʁã]	[gʁãt]
Other words	<i>trente</i> ‘thirty’	[tʁãt]	[tʁãt]
	<i>tableau</i> ‘table’	[tablo]	[tablo]

Table 4: Liaison consonants are absent from citation forms whereas stable (final and initial) consonants are present in citation forms.

showing that phonological alternations that are phonetically motivated at utterance edges are extended utterance-medially (Steriade 1997: 55-58; Myers & Padgett 2014). For instance, word-final devoicing is motivated phonetically utterance-finally (by the lack of robust cues to the voicing contrast), but not utterance-medially. Yet languages overapply word-final devoicing utterance-medially.

In Delais-Roussarie et al. (2015: 65-74)’s prosodic framework for French (see Section 2.1), citation forms are intonational phrases that consist of a single word. Connected-speech variants are the forms of words as they occur in all other prosodic contexts, i.e. within accentual phrases, at the boundary between two accentual phrases or between two intermediate phrases.

What makes liaison consonants special vis-à-vis stable consonants is that they are (generally) absent from citation forms, as shown in Table 4. For instance, liaison [t] at the end of the liaison form [gʁãt] does not appear at the end of the corresponding citation form [gʁã]. However, stable [t] at the end of *trente* ‘thirty’ appears at the end of the corresponding citation form (i.e. [tʁãt]). Similarly, stable [t] at the beginning of *tableau* ‘table’ is also present at the beginning of the corresponding citation form. The requirement for connected-speech variants to be similar to the corresponding citation forms will be key to explain two properties of French liaison consonants: (i) why they do not always surface in connected speech and (ii) why they may be pushed away from both Word1 and Word2 in Word1-Word2 sequences.

The evaluation of output-output faithfulness constraints will assume base priority, following Benua (1997: 240): the phonology of the citation form is computed first and then the resulting output form is used in the evaluation of connected-speech variants, as shown by the unidirectional horizontal arrow in Figure 1. For French liaison, it means that the non-liaison allomorph that is used utterance-finally (e.g. [gʁã]) will serve as a base in the evaluation of connected-speech variants, including in liaison contexts (i.e. before vowel-initial words).

		Underlying representation
Liaison words	<i>grand</i> ‘great’	/gʁã, gʁãt/
Other words	<i>trente</i> ‘thirty’	/tʁãt/
	<i>tableau</i> ‘table’	/tablo/

Table 5: Assumptions about underlying representations.

The analysis further assumes that liaison words come with two underlyingly listed allomorphs: an allomorph without liaison and an allomorph with liaison (e.g. [Gaatone 1978](#); [Steriade 1999](#); [Bonami & Boyé 2005](#)). For instance, the masculine adjective *grand* ‘great’ has two suppletive allomorphs /gʁã/ and /gʁãt/. By contrast, non-liaison words have a single underlying representation, as shown in Table 5.

The assumption that liaison involves suppletion is not key to the analysis. The classic generative analysis where liaison words correspond to a single underlying representation (for instance [Dell 1985](#): 180–193) could also be adopted. In this analysis, *grand* ‘great’ corresponds to a unique underlying representation /gʁãd/. All surface realizations of the adjective, including the masculine non-liaison form [gʁã] and the masculine liaison form [gʁãt], are derived from /gʁãd/ by phonological rules (deletion and final devoicing, respectively). This approach is not adopted here because it requires fairly abstract underlying representations. For instance, this approach assumes that all stable word-final consonants are followed by schwa underlyingly (e.g. *trente* is analyzed as /tʁãtə/), even though these schwas are virtually never pronounced, at least in non-meridional varieties (see [Eychenne 2019](#) on final schwa). Moreover, the suppletion analysis adopted in this paper is further motivated by cases where it is not possible to derive liaison and non-liaison forms from a single underlying representation, as in the case of *beau/bel* [bo]/[bɛl] ‘beautiful.MASC’. These cases will be further discussed in Section 4.

The evaluation of words with several underlyingly listed allomorphs is based on the model proposed by [Mascaró \(2007\)](#) for phonologically optimizing allomorph selection (e.g. *h/u*-selection in Moroccan Arabic and exceptional post-nasal voicing in Basque; [Mascaró 2007](#): 711, 718–723). In this model, both allomorphs are treated as inputs in the evaluation, without any lexical ordering among them.<sup>3</sup> The allomorph that best satisfies the phonotactic constraints of the language is selected. For French, the distribution of the two allomorphs will be analyzed as a case of phonologically optimizing suppletive allomorphy (see [Inkelas 2014](#): 282–284 for crosslinguistic evidence for this kind of pattern). The non-liaison allomorph (e.g. /gʁã/) is preferred in general because it is phonotactically simpler (it does not have a final consonant) but the liaison allomorph (e.g. /gʁãt/) may be preferred before a vowel as a strategy to avoid a vowel hiatus. Phonologically optimizing suppletive allomorphy is independently motivated in French. For instance, some verbal stems have two alternating allomorphs in [ə] and [ɛ] that cannot be derived from a single underlying form and yet have a phonotactically optimizing distribution in the present tense, with the [ɛ]-allomorph occurring

<sup>3</sup>There are also cases of allomorph selection discussed by Mascaró that require an ordering among allomorphs as their distribution cannot be predicted from the phonotactics alone, e.g. overassimilation in Baix Empordà Catalan ([Mascaró 2007](#): 723–726). These cases require an additional constraint (Priority) that is not needed in the analysis of French liaison.

under stress (e.g. *jette* [ʒɛt] ‘throw.3SG’) and the [ə]-allomorph in unstressed syllables (e.g. *jetez* [ʒə'te] ‘throw.2PL’; Storme 2020a).

The final important ingredient in the analysis is the use of a probabilistic grammar. Input-output mappings must be probabilistic in order to account for the variation observed in the realization of French liaison. The analysis will be implemented in a Maximum Entropy grammar (MaxEnt; Goldwater & Johnson 2003; Hayes & Wilson 2008). In this framework, probabilities of input-output mappings are derived from their harmony using the exponential function. And the harmony of an input-output mapping is equal to the weighted sum of its constraint violations, as in Harmonic Grammar (Smolensky & Legendre 2006). The choice of MaxEnt as a framework for probabilistic grammars is motivated by earlier research showing that this framework is well adapted to deal with variable phonological patterns (Zuraw & Hayes 2017; Flemming 2021).

### 3.2 Deriving citation forms

According to the model described in Figure 1, the phonology of citation forms is derived first. At that stage, only input-output faithfulness and phonotactic markedness play a role.

For liaison words, the preference for the allomorph without liaison (e.g. [gʁã] ‘great’) can be attributed to the effect of a markedness constraint penalizing utterance-final consonants (\*C<sub>U</sub>). \*C<sub>U</sub> is a specific version of the constraint NO-CODA (Kager 1999: 94) that penalizes coda consonants for their lack of good perceptual cues (Ohala 1990; Wright 2004). In the case study at hand, this constraint penalizes the liaison allomorph with a final consonant (e.g. [gʁãt] ‘great’), as shown in Table 6a. In this table, indices are used to indicate correspondence relations between input allomorphs and output allomorphs in the candidate set, following Mascará (2007: 721). Row (a) shows the (faithful) mapping from input /gʁã/ to output [gʁã] and row (b) the (faithful) mapping from input /gʁãt/ to output [gʁãt].<sup>4</sup> As for faithfulness, the two allomorphs tie because they are both listed underlyingly. For instance, the vowel-final candidate does not violate the faithfulness constraint penalizing consonant deletion (MAX-IO) because the consonant is absent from the corresponding input allomorph.<sup>5</sup>

<sup>4</sup>As in Mascará (2007: 721), unfaithful mappings (e.g. /gʁã/ → [gʁãt] and /gʁãt/ → [gʁã]) are ignored here because they incur strictly more constraint violations than the corresponding faithful mappings. For instance, /gʁã/ → [gʁãt] violates the same markedness constraint \*C<sub>U</sub> as /gʁãt/ → [gʁãt] but violates in addition a faithfulness constraint that bans epenthesis (DEP-IO).

<sup>5</sup>Conversely, the liaison allomorph does not violate the faithfulness constraint penalizing consonant epenthesis (DEP-IO). This constraint is not included in the analysis of citation forms because it does not play a role in distinguishing the candidates that are considered here. However it will play a role in the analysis of connected-speech variants.

UR: /gʁã <sub>1</sub> , gʁãt <sub>2</sub> /	MAX-IO $w = 50$	*C <sub>U</sub> ] $w = 5.51$	$H$	Predicted	Observed
a. [gʁã <sub>1</sub> ]			0.00	1.00	1.00
b. [gʁãt <sub>2</sub> ]		1	5.51	0.00	0.00

(a) Liaison word (*grand* ‘great’)

UR: /tʁãt/	MAX-IO $w = 50$	*C <sub>U</sub> ] $w = 5.51$	$H$	Predicted	Observed
a. [tʁãt]		1	5.51	1.00	1.00
b. [tʁã]	1		50.00	0.00	0.00

(b) Consonant-final word (*trente* ‘thirty’)

UR: /tablo/	MAX-IO $w = 50$	*C <sub>U</sub> ] $w = 5.51$	$H$	Predicted	Observed
a. [tablo]			0.00	1.00	1.00
b. [ablo]	1		50.00	0.00	0.00

(c) Consonant-initial word (*tableau* ‘table’)

Table 6: Citation forms (UR = underlying representation)

There is a categorical preference for the vowel-final allomorph in citation forms in French. In MaxEnt, this can be captured by setting a high weight to \*C<sub>U</sub>]. The specific weight shown in Table 6a for this constraint ( $w = 5.51$ ) was inferred jointly with the weights of all the other constraints presented in Section 3, using OT-Soft (Hayes et al. 2013).<sup>6</sup>

In order to block consonant deletion for words with final consonants but no vowel-final listed allomorph (e.g. *trente* [tʁãt] ‘thirty’), the constraint that penalizes consonant deletion (MAX-IO) must take precedence over the constraint penalizing utterance-final consonants (\*C<sub>U</sub>]), as shown in Table 6b. This condition ensures that a vowel-final allomorph is preferred in citation forms only in case it is listed underlyingly, as for liaison words in Table 6a. To put it differently, the preference for the vowel-final allomorph in liaison words can be analyzed as a case of emergence of the unmarked (see Mascaró 2007). Table 6b shows a concrete choice of weight for MAX-IO that derives the observed categorical preference for the candidate with a final consonant when there is no listed vowel-final allomorph ( $w = 50$ ). The weight for \*C<sub>U</sub>] is of course the same as in Table 6a because the same single grammar must derive all attested forms. Table 6c shows how the same grammar also blocks consonant deletion at the beginning of words that begin with a consonant underlyingly (e.g. *tableau* [tablo] ‘table’). The reader can check that the analysis in Table 6 derives the correct citation forms for liaison words and non-liaison words, as listed in Table 5.

<sup>6</sup>The number of iterations was set to 500, the minimum weight to 0, and the maximum weight to 50. Note that the predicted probability for the liaison allomorph appears to be equal to 0 in Table 6a but this is due to rounding. In MaxEnt, harmonically bounded candidates always have a small probability of occurrence. Here it is considered to be sufficiently small so that it can be ignored. An even higher weight on the phonotactic constraint would be required to make it even less likely.

### 3.3 Deriving connected-speech variants

In connected speech, markedness constraints that do not play a role utterance-finally become relevant and drive alternations. The analysis here focuses specifically on the context where liaison consonants and stable word-final/word-initial consonants have different prosodic behaviors, namely in Word1-Word2 sequences with a prosodic boundary between the two words and where Word2 starts with a vowel (see Section 2.1). In particular, the analysis focuses on the prosodic boundary between two accentual phrases (AP), as this is the context where the prosodic variability of liaison is the clearest, in particular in natural conversations (see Section 2.1).

The crucial markedness constraint that motivates the use of the liaison allomorph in this context is the anti-hiatus constraint \*VV (see Steriade 1999; Tranel 2000). It is assumed that this constraint remains relevant in the presence of a prosodic break between the two vowels, at least at the boundary between two accentual phrases or intermediate phrases (see Section 2.1).<sup>7</sup> This assumption is crucial to explain why liaison might occur across a prosodic break. Also, in addition to input-output faithfulness, paradigm uniformity with citation forms derived in Section 3.2 will also play a role, in line with the grammatical architecture described in Figure 1.

Table 7a shows the analysis of liaison words in this context. The prosodic boundary between the two APs is indicated by |. Candidate (a) corresponds to the allomorph without liaison. Whereas this candidate did not violate any constraint in citation forms, it now violates the anti-hiatus constraint \*VV because it is followed by a vowel-initial word (*ami* ‘friend’ in this case). This violation will leave room for the liaison allomorph to surface. Note that the constraint weights reported in Table 7 were inferred jointly for the analysis of citation forms and connected-speech variants.

---

<sup>7</sup>To the author’s knowledge, liaison has never been reported to occur across a very strong prosodic boundary such as the boundary between two intonational phrases.

URs: /gʁã <sub>1</sub> , gʁãt <sub>2</sub> /+/ami/ CFs: [gʁã]+[ami]	DEP-IO 50	R-ANCH-OO 4.78	L-ANCH-OO 2.80	*VV 2.63	<i>H</i>	Predicted	Observed
a. [gʁã <sub>1</sub>  ami]				1	2.63	0.51	0.51
b. [gʁãt <sub>2</sub>  ami]		1			4.78	0.06	0.06
c. [gʁã t <sub>2</sub> ami]			1		2.80	0.43	0.43
d. [gʁã <sub>1</sub> l ami]	1	1			54.78	0.00	0.00
e. [gʁã <sub>1</sub>  lami]	1		1		52.80	0.00	0.00

(a) Liaison Word1 + vowel-initial Word2 (*grand ami* ‘great friend’)

URs: /tʁãt/+/ami/ CFs: [tʁãt]+[ami]	DEP-IO 50	R-ANCH-OO 4.78	L-ANCH-OO 2.80	*VV 2.63	<i>H</i>	Predicted	Observed
a. [tʁãt ami]					0.00	1.00	1.00
b. [tʁã tami]		1	1		7.59	0.00	0.00

(b) Consonant-final Word1 + vowel-initial Word2 (*trente amis* ‘thirty friends’)

URs: /vʁɛ/+/tablo/ CFs: [vʁɛ]+[tablo]	DEP-IO 50	R-ANCH-OO 4.78	L-ANCH-OO 2.80	*VV 2.63	<i>H</i>	Predicted	Observed
a. [vʁɛt ablo]		1	1		7.59	0.00	0.00
b. [vʁɛ tablo]					0.00	1.00	1.00

(c) Vowel-final Word1 + consonant-initial Word2 (*vrai tableau* ‘true table’)

Table 7: Connected-speech variants (CF = citation form, UR = underlying representation). The vertical bar | between the two strings in each candidate represents a prosodic break.

Candidates (b) and (c) both feature the liaison consonant [t], with a prosodic attachment to Word1 (*liaison non-enchaînée*) and to Word2 (*liaison enchaînée*), respectively. These candidates fare better than candidate (a) phonotactically because they do not violate \*VV. Furthermore, contrary to candidates (d) and (e) (with liaison [l]), they do not violate the input-output faithfulness constraint penalizing consonant epenthesis (DEP-IO) because /t/ (but not /l/) is listed underlyingly in the liaison allomorph.

However these candidates are not completely optimal because they are not identical to the citation form [gʁã], contrary to candidate (a): they feature a consonant that is epenthetic in the output-output dimension. In Table 7a, this lack of paradigm uniformity is penalized by two output-output faithfulness constraints, based on Kager (1999: 251): RIGHT-ANCHOR-OO (abbreviated as R-ANCH-OO) and LEFT-ANCHOR-OO (abbreviated as L-ANCH-OO). These constraints play a crucial role in accounting for both the availability of the vowel-final allomorph and the prosodic variability of the liaison consonant in the liaison allomorph. They are defined in (8) and (9) and further explained below.

(8) RIGHT-ANCHOR-OO

Assign one violation mark if the segment at the right edge of an output form does not stand in correspondence with the segment at the right edge of the corresponding citation form.

(9) LEFT-ANCHOR-OO

Assign one violation mark if the segment at the left edge of an output form does

not stand in correspondence with the segment at the left edge of the corresponding citation form.

RIGHT-ANCHOR-OO has the same definition as the Lex A-Phrase constraint proposed by Steriade (1999: 267) to account for the word-initial attachment of liaison consonants in dislocations (see Section 2.1). This constraint requires that any element at the right edge of a citation form have a correspondent at the right edge of the corresponding contextual variant and vice versa. In a Word1-Word2 sequence, this constraint has the effect of penalizing the deletion/epenthesis of a segment at the end of Word1 (in the output-output dimension). It penalizes candidate (b) in Table 7a because this candidate features a consonant at the right edge of [gʁãt] that is absent from the corresponding citation form [gʁã]. This constraint does not penalize candidate (c) if one assumes that a segment has to be contiguous with the rest of an output form to qualify as its edge. In the presence of a prosodic break between [gʁã] and [t], [gʁã] and [t] are not interpreted as contiguous and therefore [t] does not qualify as a right edge for the output of the adjective *grand*. With this assumption in place, RIGHT-ANCHOR-OO favors a prosodic attachment of liaison [t] at the beginning of Word2 (candidate (c)).

The present proposal extends Steriade’s analysis by assuming that paradigm uniformity does not only apply to the right edge but also to the left edge of words in connected speech. In a Word1-Word2 sequence, these two requirements will conflict and drive the variable prosodic attachment of liaison consonants. It should be noted that the idea that connected-speech variants must be faithful to their citation forms along their left edge was actually proposed by Zuraw & Hayes (2017) in the context of a discussion of *h-aspiré* words in French. *H-aspiré* words block a range of sandhi phenomena, including liaison but also other processes such as vowel elision. Zuraw & Hayes (2017) proposed that this is due to *h-aspiré* words being subject to a greater pressure to be uniform with their citation forms along their left edge than other vowel-initial words. They used lexical indexation of constraints to derive this effect. However, the formal implementation of their analysis does not use paradigm uniformity constraints, as in this paper, but alignment constraints. Alignment constraints require that syllable boundaries and morpheme boundaries match. Alignment constraints could be used to derive the prosodic variability of liaison. However, they would fail to derive *segmental* variability: as mentioned in Section 2.2, liaison consonants are segmentally variable (e.g. in terms of their phonetic duration) even in contexts where they are traditionally treated as onsets. By contrast, as will be shown in Section 5, paradigm uniformity can explain the segmental variability of liaison.

The requirement to protect the left edge of a word is modeled using the paradigm uniformity constraint LEFT-ANCHOR-OO. LEFT-ANCHOR-OO requires that any element at the left edge of a citation form have a correspondent at the left edge of the corresponding contextual variant and vice versa. In a Word1-Word2 sequence, this constraint has the effect of penalizing segment deletion/epenthesis at the beginning of Word2 (candidate (c) in Table 7a) and therefore favors a prosodic attachment of liaison [t] at the end of Word1 (candidate (b)).<sup>8</sup>

---

<sup>8</sup>Note that [l]-epenthesis in candidates (d) and (e) is also penalized by RIGHT-ANCHOR-OO and LEFT-ANCHOR-OO, respectively: indeed [l] is absent not only from the underlying representations of *grand* and *ami* but also from their citation forms.

Table 7a shows a concrete choice of constraint weights that can derive the frequencies attested in [Encrevé \(1988\)](#)’s corpus. Constraint weights were inferred jointly for the analysis of citation forms and connected-speech variants.<sup>9</sup> Crucially, the analysis is not only able to model the variation between allomorphs with and without liaison (candidates (b)/(c) vs. candidate (a)) but also to capture the prosodic variability of liaison consonants: attachments to both Word1 and Word2 are derived (candidate (b) vs. candidate (c)). The analysis also does a very good job at matching the corpus frequencies of the different variants for liaison words.

The analysis predicts prosodic variability for liaison consonants but not for stable word-final and word-initial consonants. This is shown in Tables 7b and 7c, respectively.

In Table 7b, candidate (b) deletes word-final [t] in the citation form of *trente* and epenthesizes a [t] at the beginning of the citation form of *ami*, hence violating the two output-output faithfulness constraints RIGHT-ANCHOR-OO and LEFT-ANCHOR-OO. By contrast, candidate (a) does not violate any constraint and therefore harmonically bounds candidate (b). The specific choice of weights that makes the prosodic attachment of liaison consonants variable between Word1 and Word2 (Table 7a) predicts a categorical attachment to Word1 for stable word-final consonants (Table 7b).<sup>10</sup>

In Table 7c, candidate (a) epenthesizes a [t] at the end of the citation form of *vrai* and deletes word-initial [t] in the citation form of *tableau*, hence violating the two output-output faithfulness constraints LEFT-ANCHOR-OO and RIGHT-ANCHOR-OO. By contrast, candidate (b) does not violate any constraint and therefore harmonically bounds candidate (a). The specific choice of weights that makes the prosodic attachment of liaison consonants variable between Word1 and Word2 (Table 7a) predicts a categorical attachment to Word2 for stable word-initial consonants (Table 7c).

### 3.4 Implicational generalizations

The preceding section has shown that the paradigm uniformity analysis can capture the variable prosodic patterning of liaison consonants as attested in [Encrevé’s](#) corpus. This section further establishes that the asymmetry between liaison and stable consonants in (10) follows as a necessary consequence of the proposed constraint set, and this holds regardless of the particular framework chosen for probabilistic grammars: Stochastic Optimality Theory (OT; [Boersma & Hayes 2001](#)), Noisy Harmonic Grammar (HG; [Hayes 2017](#)), or Maximum Entropy grammars (MaxEnt; [Goldwater & Johnson 2003](#); [Hayes & Wilson 2008](#)).

- (10) Statistical implicational generalizations derived by the analysis
- a.  $P(\text{Word1-attachment}|\text{Liaison C}) \leq P(\text{Word1-attachment}|\text{Final C})$
  - b.  $P(\text{Word2-attachment}|\text{Liaison C}) \leq P(\text{Word2-attachment}|\text{Initial C})$

---

<sup>9</sup>Note that they could also be inferred independently since there happens to be no overlap between the constraints used in the two analyses.

<sup>10</sup>Word1-attachment is actually predicted to be *almost* categorical, because harmonically bounded candidates have a non-zero probability in MaxEnt (see Section 3.4). The prediction would be completely categorical in other frameworks such as Noisy Harmonic Grammar ([Hayes 2017](#)). The difference between ‘almost categorical’ and ‘categorical’ is assumed to be sufficiently small so that it can be ignored for the purposes of this paper.

In words: Liaison consonants are less likely to attach to Word1 than final consonants and less likely to attach to Word2 than initial consonants.

In Stochastic OT and Noisy HG, the demonstration makes use of harmonic bounding. For words that do not have liaison allomorphs (e.g. *trente* ‘thirty’ and *tableau* ‘table’), the candidates that involve displacing the consonant from its underlying position to the following or preceding word (candidate (b) in Table 7b and candidate (a) in Table 7c, respectively) violate a strict superset of the constraints violated by the candidates that involve no displacement (candidate (a) in Table 7b and candidate (b) in Table 7c, respectively). In other words, these candidates are harmonically bounded and there is no way they can win under any constraint ranking or weighting. Consequently, in the presence of a prosodic break, stable final consonants in Word1 are predicted to attach categorically to Word1 and stable initial consonants in Word2 to attach categorically to Word2.

However for liaison words, the candidates involving attachment of the liaison consonant to Word1 and to Word2 (candidates (b) and (c) in Table 7a, respectively) violate different constraints (RIGHT-ANCHOR-OO and LEFT-ANCHOR-OO, respectively). So it is not the case that one pronunciation is inherently better than the other. Varying rates of Word1-/Word2-attachment are expected to be observed depending on the ranking values (in Stochastic OT) or the weights (in Noisy HG) of RIGHT-ANCHOR-OO and LEFT-ANCHOR-OO. In other words, the analysis predicts that the rate of Word2-attachment for liaison consonants should always be located between the rate of Word2-attachment for final consonants (equal to 0) and the rate of Word2-attachment for initial consonants (equal to 1).

In MaxEnt, the situation is different because the fact that a candidate is harmonically bounded does not entail that it has a nonzero probability (e.g. Hayes 2017). The demonstration is more involved and the constraint-violation profiles of the different candidates must be carefully compared across the different types of consonants (final, liaison, and initial), as discussed in Anttila & Magri (2018). This demonstration is not provided here for reasons of space. In a nutshell, generalization (10a) holds because the candidate with a liaison consonant at the end of Word1 fares worse within the candidate set for the liaison word (Table 7a) than the candidate with a stable consonant at the end of Word1 does within the candidate set for the consonant-final word (Table 7b). And generalization (10b) holds because the candidate with a liaison consonant at the beginning of Word2 fares worse within the candidate set for the liaison word (Table 7a) than the candidate with a word-initial consonant at the beginning of Word2 does within the candidate set for the consonant-initial word (Table 7c). The reader is referred to the output of the CoGeTo analysis (Magri & Anttila 2019) in the supplementary files. CoGeTo is an online tool developed by Giorgio Magri and Arto Anttila to analyze the typological predictions of probabilistic constraint-based grammars. The CoGeTo analysis shows that indeed the implications hold in MaxEnt as well.

## 4 Study 1: epenthetic and suppletive liaison in European French

In the paradigm uniformity analysis of French liaison proposed in Section 3, the variable prosodic patterning of liaison consonants ultimately follows from their being *absent* from

		Underlying representation	Citation form
Epenthetic liaison	grand ‘great’	/gʁã, gʁãt/	[gʁã]
Suppletive liaison	beau ‘beautiful’	/bo, bɛl <sub>FEM</sub> /	[bo], [bɛl <sub>FEM</sub> ]

Table 8: Epenthetic liaison and suppletive liaison

the corresponding citation form. But no variability is predicted if the liaison consonant is *present* in the citation form. The goal of this section is to test this prediction experimentally by exploiting the distinction between two types of liaison consonants that differ in whether they are present in the corresponding citation form.

Section 4.1 introduces this distinction and explains how it provides a testing ground for the role of citation forms in the realization of liaison consonants. Section 4.2 presents the methods used to test this prediction. Section 4.3 presents the results. Section 4.4 concludes with a brief discussion. The data and code for Study 1 are available in [Storme \(2022\)](#) under the names `study1b-data.csv` and `study1b-code.R`, respectively.

## 4.1 Epenthetic and suppletive liaison

Epenthetic liaison describes cases where the liaison allomorph contains the morphologically corresponding citation form as a substring, with the liaison consonant being epenthesized after this substring (e.g.  $[gʁãt]_{\text{liaison allomorph}} = [gʁã]_{\text{citation form}} + [t]$ ). Suppletive liaison describes cases where the liaison allomorph does not contain the morphologically corresponding citation form as a substring but is based on a morphologically distinct form in the paradigm. For instance, the adjective *beau* [bo] ‘beautiful.MASC’ uses the form [bɛl] as a liaison variant (e.g. *bel ami* [bɛlami] ‘beautiful friend.MASC’). This form cannot be analyzed as the masculine citation form plus an epenthetic consonant. Rather it corresponds to the feminine form of the adjective (*belle* [bɛl] ‘beautiful.FEM’). The two types of liaisons are represented in Table 8. It is important to note that epenthesis is not used here in its usual sense to mean that the liaison consonant is epenthesized in the *input-output* dimension. Rather it is used to characterize the relationship between the liaison allomorph and the citation form in the *output-output* dimension. The liaison consonant is still assumed to be present underlyingly in the liaison allomorph for both types of liaison words, as shown in Table 8 (see Section 3.1).

The distinction between the two types of liaison is well-known and has been discussed by [Delattre \(1947: 150\)](#) and [Tranel \(1990, 2000\)](#) among others. Its relevance for the hypothesis of paradigm uniformity effects was first discussed by [Steriade \(1999\)](#). The paradigm uniformity analysis predicts that only epenthetic liaison should pattern variably between stable word-final and word-initial consonants. For suppletive liaison, the liaison consonant is present at the end of the corresponding citation form (e.g. the [l] in *bel* [bɛl] is present at the end of the feminine citation form *belle* [bɛl]) and this word-final attachment should be maintained in contextual realizations through paradigm uniformity.

Table 9 shows that the same grammar that derived prosodic variability for epenthetic liaison in Table 7a predicts a non-variable word-final behavior for suppletive liaison: candidate (e) (with liaison [l] attaching to Word2 across a prosodic break) is harmonically bounded by candidate (d) (with liaison [l] attaching to Word1). Epenthetic liaison (candidates (b)

and (c)) is also ruled out because /t/ is not present in any listed allomorph for the adjective *beau*. Note that no constraint penalizes the liaison allomorph [bɛl] in Table 9. In a comprehensive analysis, this candidate would violate a constraint banning mismatches between the morphosyntactic specification of an adjective (masculine singular here) and the corresponding morphological exponent (feminine singular here), as proposed by Steriade (1999: 256). Adding this constraint would not affect the prediction concerning the non-variable prosodic behavior of suppletive liaison: this constraint would be violated equally by the candidates involving Word1-attachment (candidate d) and Word2-attachment (candidate e).

URs: /bo, bɛl <sub>FEM</sub> /+/ami/ CFs: [bo, bɛl <sub>FEM</sub> ]+[ami]	DEP-IO 50	R-ANCH-OO 4.78	L-ANCH-OO 2.80	*VV 2.63	<i>H</i>	Predicted
a. [bo ami]				1	2.63	0.07
b. [bot ami]	1	1			54.78	0.00
c. [bo tami]	1		1		52.80	0.00
d. [bɛl ami]					0.00	0.93
e. [bɛ lami]		1	1		7.59	0.00

Table 9: Suppletive liaison consonants are predicted to behave categorically like word-final consonants and attach to Word1 (*bel ami* ‘beautiful friend’). The vertical bar | between the two strings in each candidate represents a prosodic break. CF = citation form, UR = underlying representation.

There is preliminary evidence for the prediction that epenthetic liaison and suppletive liaison differ in this way, as pointed out by Steriade (1999). In right-dislocation contexts, Tranel (1990) reports that epenthetic liaison consonants attach to Word2 whereas suppletive liaison consonants attach to Word1, as illustrated in (11).

- (11) Epenthetic liaison vs. suppletive liaison in right dislocations
- a. Epenthetic liaison *J'en ai un grand, éléphant.*  
 (ʒãneẽgʁã)<sub>ip</sub> (tɛlɛfã)<sub>ip</sub>  
 ‘I have a big one, elephant.’
- b. Suppletive liaison *J'en ai un bel, éléphant.*  
 (ʒãneẽbɛl)<sub>ip</sub> (ɛlɛfã)<sub>ip</sub>  
 ‘I have a beautiful one, elephant.’

However, right-dislocation contexts are probably not the most appropriate context to make a case for the variable patterning of epenthetic liaison, as they very strongly favor a prosodic attachment to Word2 (= *liaison enchaînée*). As noted by Durand & Lyche (2008: 50), the contexts where epenthetic liaison consonants are more readily found to attach to Word1 (= *liaison non-enchaînée*) involve a hesitation between Word1 and Word2 (see Section 2.1).<sup>11</sup> The goal of Study 1 is therefore to compare the behavior of liaison consonants (both

<sup>11</sup>In the paradigm uniformity analysis, the difference between right-dislocation contexts and hesitations could be captured by indexing paradigm uniformity constraints (e.g. RIGHT-ANCHOR-OO) to prosodic domains. In right dislocations such as (11), the adjective (Word1) is under focus. When there is a hesitation between Word1 and Word2, Word1 is not necessarily under focus. If the requirement to be similar to the citation form is more strongly enforced under focus, then this might explain why epenthetic liaison consonants are more likely to be pushed onto Word2 in right dislocations. See Section 7.4 for more discussion on the role of indexed paradigm uniformity constraints in French liaison.

epenthetic and suppletive) and stable consonants (both word-final and word-initial) across a prosodic break involving a hesitation.

## 4.2 Methods

Adjective-noun (Adj-N) sequences were chosen as Word1-Word2 sequences. This choice was motivated by the fact that both epenthetic liaison (e.g. *grand*) and suppletive liaison (e.g. *beau/bel*) can be found among adjectives. Each of the four experimental conditions (epenthetic liaison, suppletive liaison, stable word-final consonants, stable word-initial consonants) was represented by 12 Adj-N sequences, for a total of 48 Adj-N sequences. Six adjectives were used by condition, as shown in Table 10, and each adjective appeared in two Adj-N sequences varying by the strength of their collocation. For instance, *petit* appeared both in *petit ami* ‘boyfriend’ (more frequent) and in *petit anneau* ‘small ring’ (less frequent).<sup>12</sup> This manipulation was meant to control for potential effects of the following noun on the behavior of liaison consonants, as this variable has been shown to influence some aspects of French liaison in previous research (Fougeron et al. 2001; Kilbourn-Ceron 2017).

Condition	Adjectives
Adjectives with stable word-final consonants	<i>énorme, jeune, large, magnifique, meilleur, superbe</i>
Adjectives with epenthetic liaison	<i>faux, grand, gros, mauvais, parfait, petit</i>
Adjectives with suppletive liaison	<i>ancien, bel, bon, prochain, vieil, nouvel</i>
Adjectives pronounced with a final vowel	<i>affreux, charmant, gentil, joli, long, vrai</i>

Table 10: Adjectives used in Study 1

Two French native speakers (one female and one male) read each of the 48 Adj-N sequences twice, with a hesitation (*euh* [œ]) between the two words. The two pronunciations varied in the prosodic attachment of the intervocalic consonant between the two words. In one pronunciation, the consonant was pronounced at the end of Word1 before the hesitation. This corresponds to a case of *liaison non-enchaînée* for liaison conditions. In the other pronunciation, the consonant was pronounced at the beginning of Word2 after the hesitation. This corresponds to a case of *liaison enchaînée*. Examples are shown in Table 11 for each of the four experimental conditions. The prosodic attachment of the consonant was clearly indicated to the speakers by using bold characters (e.g. *un gran-[t] euh hommage/un gran euh [t]-hommage*). The speakers were unaware of the purpose of the study when they were recorded.<sup>13</sup> In order to make the loudness of the sound files comparable, the root-mean-square amplitude was equalized across the set of sound files and scaled to a max peak value of 1 using a Praat script written by Gabriël J.L. Beckers.<sup>14</sup>

<sup>12</sup>The strength of the collocation was measured as the conditional probability of N given Adj in the corpus of movie and TV subtitles OpenSubtitles (Lison & Tiedemann 2016).

<sup>13</sup>In a pilot study run with Swiss French speakers, the stimuli were read by the author. The data and code for this pilot study are also available in Storme (2022) (see `study1-data.csv` and `study1-code.R`). The same results were found in the pilot study as in the study presented in the paper.

<sup>14</sup>The Praat script is available at <http://wwwbio.leidenuniv.nl/~eew/G6/staff/beckers/beckers.html>.

	Pronunciation		
	Word1 attachment	Word2 attachment	
Stable word-final C	<i>magnifi[k] euh hôtel</i>	<i>magnifi euh [k]hôtel</i>	‘magnificent hotel’
Epenthetic liaison	<i>grand[t] euh hommage</i>	<i>grand euh [t]hommage</i>	‘great tribute’
Suppletive liaison	<i>be[l] euh appartement</i>	<i>be euh [l]appartement</i>	‘beautiful apartment’
Stable word-initial C	<i>joli[s] euh ourire</i>	<i>joli euh [s]ourire</i>	‘nice smile’

Table 11: Experimental items

Twenty-four French participants were recruited via the CNRS RISC platform (<https://expesciences.risc.cnrs.fr/>) to participate in a study run online. The 48 Adj-N sequences were presented to participants in random order and each participant heard each sequence only once (they heard either the sequence produced by the female speaker or the sequence produced by the male speaker, and this assignment was done randomly for each sequence). For each Adj-N sequence, the two pronunciations were presented one after the other, with the pronunciation involving prosodic attachment to Word1 always preceding the pronunciation involving prosodic attachment to Word2 (with a one-second interstimulus interval).

Participants were asked to indicate which of the two pronunciations sounded more natural to them. The target Adj-N sequence was not presented graphically to participants in order to avoid any explicit orthographic bias. Liaison consonants (both epenthetic and suppletive) appear at the end of Word1 in the spelling and this could directly bias participants towards a word-final attachment. Participants were invited to wear headphones while taking the study. The LimeSurvey platform (LimeSurvey 2012) was used to carry out the study. The participants provided their informed consent to participate in the research and agreed to make their data available online. No sensitive information about participants was collected.

A Bayesian hierarchical logistic regression was fit to the participants’ responses as a function of the dummy-coded factor Consonant (reference level ‘stable word-final consonant’). Consonant has four levels, corresponding to the four types of consonants (stable word-final consonant, suppletive liaison, epenthetic liaison, stable word-initial consonant). The random-effects structure included: (i) random intercepts for each speaker, each participant, and each Adj-N sequence and (ii) by-speaker and by-participant random slopes for the effect of Consonant. The logistic regression was fit using the `brms` package (Bürkner 2017) in R (R Core Team 2020). A Bayesian approach was chosen (rather than a frequentist approach) because it provides more intuitive and meaningful inferences and also virtually always converges to accurate values of the parameters (Kruschke & Liddell 2018).

For hypothesis testing, the difference  $\Delta$  in the posterior log-odds ratios of attachment to Word2 was computed for the different experimental conditions. Compelling evidence for a difference between two conditions was considered to be provided only in case zero was outside of the posterior 95% Credible Interval (CI) for  $\Delta$ . Credible Intervals were obtained using the ETI (Equal-tailed Interval) method and the package `bayestestR` (Makowski et al. 2019).

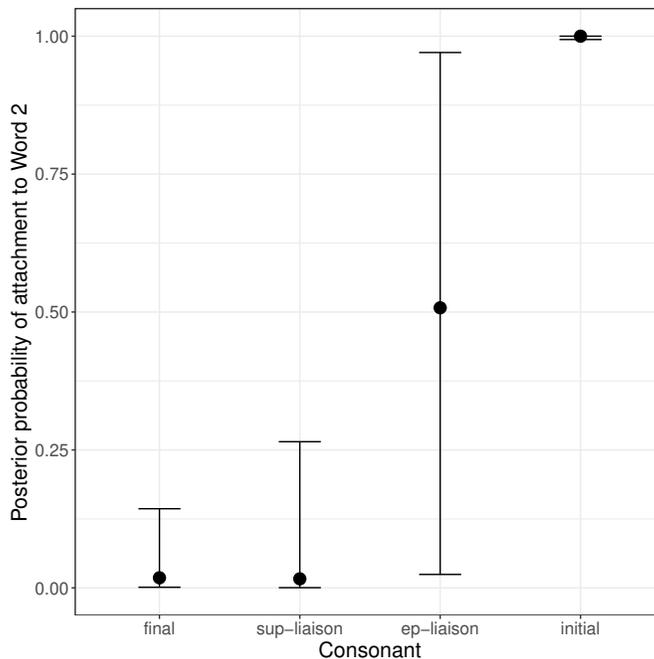


Figure 2: Posterior probability of attachment to Word2 as a function of consonant type (mean and 95% CI)

### 4.3 Results

Figure 2 shows the posterior probability that the consonant attaches prosodically to Word2 for each consonant type. As predicted by the paradigm uniformity analysis, suppletive liaison and epenthetic liaison were found to pattern differently, with suppletive liaison being less likely to attach to Word2 than epenthetic liaison ( $\Delta_{\text{sup liaison} - \text{ep liaison}} = -4.23$ ,  $CI = [-8.67, -0.04]$ ). Suppletive liaison was found to behave like stable word-final consonants ( $\Delta_{\text{sup liaison} - \text{final}} = -0.19$ ,  $CI = [-3.45, 2.66]$ ), favoring an attachment to Word1 almost categorically. Epenthetic liaison was found to have a rate of attachment to Word2 that is intermediate between stable word-final consonants ( $\Delta_{\text{ep liaison} - \text{final}} = 4.03$ ,  $CI = [0.96, 7.22]$ ) and word-initial consonants ( $\Delta_{\text{ep liaison} - \text{initial}} = -11.28$ ,  $CI = [-22.82, -4.41]$ ).

### 4.4 Discussion

An important prediction of the paradigm uniformity analysis was corroborated by the results of Study 1: liaison behaves variably as a word-final or word-initial consonant if and only if the liaison consonant can be analyzed as epenthesized at the end of the citation form. When the liaison allomorph is identical to the feminine form of the adjective and cannot be analyzed as a substring of the masculine form (= suppletive liaison), no prosodic variability arises. This study is to the author’s knowledge the first controlled study that establishes this difference between epenthetic liaison and suppletive liaison.

## 5 Segmental variability of French liaison as a paradigm uniformity effect

This section shows how the model proposed in Section 3 can be extended to deal with the segmental variability of French liaison by making paradigm uniformity constraints sensitive to phonetic detail, as proposed by Steriade (2000). This section focuses on the interaction of liaison and affrication in Quebec French as this phenomenon provides the clearest case of variable realization for French liaison: liaison /t/ has a rate of affrication that is intermediary between that of word-final /t/ and word-initial /t/ (see Section 2.2).<sup>15</sup> Also, this pattern can be modeled with the type of categorical phonological grammars used in Section 3 since affrication is usually described in a categorical manner (presence/absence of affrication) and therefore can be analyzed using discrete phonological representations. Modeling phonetic realization in a continuous space would require moving from phonological to phonetic grammars (Flemming 2001). This task is left for future research.

The key ingredient in the analysis will be the observation that coarticulation is bidirectional, namely it affects both the realization of C and V in a CV sequence. Bidirectionality of coarticulation predicts that a change in C (e.g. affrication in the case of Quebec French) correlates with a change in the following V. In combination with paradigm uniformity, this correlation will be crucial to explain why liaison consonants might pattern intermediately between word-final and word-initial consonants phonetically. In a nutshell, CV-coarticulation at word boundaries will potentially result in violations of paradigm uniformity for both words in a Word1-Word2 sequence. However this coarticulation will have different effects for the three types of consonants. Coarticulated *liaison* consonants will incur *less* output-output faithfulness violations than coarticulated *final* consonants (due to the liaison consonant being absent from the corresponding citation form). But coarticulated *liaison* consonants will imply *more* output-output faithfulness violations for the following vowel than coarticulated *initial* consonants will (because initial consonants are already articulated with the following vowel in the corresponding citation form whereas liaison consonants are not).

Section 5.1 provides some background on the bidirectionality of coarticulation, and shows how it applies in the case of Quebec French affrication. Building on these results, Section 5.2 shows how the analysis derives citation forms for words with liaison allomorphs and for words that lack such allomorphs. Section 5.3 shows how the segmental variability of liaison consonants as documented by Côté (2014)’s corpus can be derived as a paradigm uniformity effect, assuming bidirectionality of coarticulation. Section 5.4 further shows that the asymmetry between liaison consonants and stable consonants is a necessary consequence (an implicational universal) of the proposed constraint set. The analysis can be found in Storme (2022) under the name `phonetic-ambiguity-final-2.txt`.

### 5.1 Bidirectionality of coarticulation in CV sequences

Probably the best studied case of coarticulation is the assimilation in second formant (F2) frequency between consonants and vowels (for example Flemming 2001: 16-23). A large number of studies have shown that C assimilates to V in CV, in particular F2 at consonant

---

<sup>15</sup>The analysis presented in this section supersedes the analysis presented in Storme (2020b).

release can be described as an increasing linear function of F2 in the middle of the vowel: as the F2 in the middle of the vowel increases, the F2 at consonant release also increases (Lindblom 1963; Sussman et al. 1991). In turn, V has also been found to assimilate to C in CV, with F2 in the middle of the vowel being higher when F2 at consonant release is higher (Lindblom 1963; Broad & Clermont 1987). These results suggest that coarticulation is bidirectional in CV sequences: both C and V are affected when the two sounds are combined in a CV sequence, and any change affecting one of the two sounds should also affect the other one.

Bidirectionality of coarticulation extends beyond this well studied case and applies in particular to Quebec French affrication. In Quebec French, affrication of /t d/ before high front vowels and glides applies almost categorically morpheme-internally (Côté 2014). Phonetically, affrication involves a change in consonant manner: the stop burst is followed by a frication noise (Stevens 1998: 412). But affrication before high vowels does not only affect the realization of the consonant. It also correlates with changes in the following vowel (Cedergren & Simoneau 1985; Dow 2019). In particular, Cedergren & Simoneau (1985: 72-80) report that high vowels tend to be reduced/deleted in the vicinity of fricatives, including after affricates. This effect is stronger with voiceless fricatives/affricates. Because /t/ maps to a voiceless affricate [ts] after affrication, high-vowel reduction is expected to be particularly common after this sound. In the remainder of this paper, [i̥] will be used to note this reduced/deleted high vowel. In other words, an underlying sequence /ti/ tends to be realized as [tsi̥] on the surface in Quebec French, with both affrication and high-vowel reduction.

This coarticulatory pattern involving fricatives/affricates and high vowels is found in other languages such as Japanese (Beckman & Shoji 1984; Whang 2018). For instance, Whang (2018: 1166) found a positive correlation between lengthening of [tʃ] (from an underlying /t/) and high-vowel devoicing in [tʃi̥] sequences in Japanese. This result suggests that, as /t/ gets more affricated (the frication noise gets longer), the following high vowel gets more reduced, in line with what has been found in Quebec French. Based on such parallels, Cedergren & Simoneau (1985: 189) propose that this interaction stems from a universal phonetic constraint, but without providing more details. One possible mechanism relating the two changes is compensatory lengthening/shortening: there is a trading relationship between the duration of C and V such that if the frication noise of C lengthens then V shortens and conversely (see Whang 2018: 1160 and literature therein).

The present paper will remain agnostic as to what the precise coarticulatory mechanism underlying this pattern is. In what follows, the constraint that drives the interaction between affrication and high-vowel reduction will be noted descriptively as \*tsi. Assuming that affrication of /t/ to [ts] is independently motivated before [i] in a language (by a markedness constraint \*ti), the constraint \*tsi will favor a candidate [tsi̥] involving a concomitant change in vowel quality over a candidate [tsi] involving affrication but no high-vowel reduction.

## 5.2 Deriving citation forms

The same grammatical architecture is assumed in the analysis of segmental variability as in the analysis of prosodic variability (see Figure 1). According to this model, the phonology of citation forms is derived first. At that stage, only input-output faithfulness and phonotactic markedness play a role.

The preference for realizing underlying /ti/ as [tʃi] (candidate (c) in Table 12a) can be attributed to the effect of the two markedness constraints \*ti and \*tʃi described in the previous section. If these constraints have higher weights than the input-output faithfulness constraints protecting against changes in consonant continuancy (IDENT-IO(CONT)) and in vowel voicing (IDENT-IO(VOI)), then candidate (c) is predicted to be favored over the faithful candidate (candidate (a)) and over the candidate that only changes consonant continuancy (candidate (b)). Table 12a shows a specific choice of constraint weights in MaxEnt that derives the near-categorical rate of affrication of morpheme-internal /ti/ documented by Côté (2014). The weights were inferred using OT-Soft, as in Section 3.<sup>16</sup> \*C<sub>U</sub>] penalizes utterance-final consonants, as in Section 3.

UR: /tiḃã/	*tʃi	*ti	*C <sub>U</sub> ]	IDENT-IO(CONT)	IDENT-IO(VOI)	<i>H</i>	Predicted	Observed
	50	15.81	7.82	6.16	6.16			
a. [tiḃã]		1				15.81	0.03	0.01
b. [tʃiḃã]	1			1		56.16	0	0
c. [tʃiḃã]				1	1	12.33	0.97	0.99

(a) Word with initial /ti/ (*tyran* ‘tyrant’)

UR: /tḃãt/	*tʃi	*ti	*C <sub>U</sub> ]	IDENT-IO(CONT)	IDENT-IO(VOI)	<i>H</i>	Predicted	Observed
	50	15.81	7.82	6.16	6.16			
a. [tḃãt]			1			7.82	1.00	1.00
b. [tʃãts]			1	1		13.99	0	0

(b) Word with final /t/ (*trente* ‘thirty’)

UR: /inosã/	*tʃi	*ti	*C <sub>U</sub> ]	IDENT-IO(CONT)	IDENT-IO(VOI)	<i>H</i>	Predicted	Observed
	50	15.81	7.82	6.16	6.16			
a. [inosã]						0	1.00	1.00
b. [iḃnosã]					1	6.16	0	0

(c) Word with initial /i/ (*innocents* ‘innocent’)

UR: /gḃãt <sub>1</sub> , gḃãt <sub>2</sub> /	*tʃi	*ti	*C <sub>U</sub> ]	IDENT-IO(CONT)	IDENT-IO(VOI)	<i>H</i>	Predicted	Observed
	50	15.81	7.82	6.16	6.16			
a. [gḃãt <sub>1</sub> ]						0	1.00	1.00
b. [gḃãt <sub>2</sub> ]			1			7.82	0	0

(d) Word with liaison /t/ (*grand* ‘great’)

Table 12: Citation forms

Tables 12b and 12c show that the same grammar predicts that affrication and high-vowel reduction should not apply in the citation forms of words with stable word-final /t/ and words with word-initial /i/, respectively. In these cases, the relevant markedness constraints are not violated and therefore nothing motivates any change in consonant continuancy or vowel quality on the surface.

<sup>16</sup>The weights of all constraints used in the analysis in Sections 5.2 and 5.3 were inferred jointly. The number of iterations was set to 10,000, the minimum weight to 0, and the maximum weight to 50.

The analysis above accounts for the affrication of morpheme-internal /ti/ in Quebec French. The citation forms for liaison words must also be derived. In Quebec French, the allomorph without liaison (e.g. *grand* [gʁɑ̃]) is categorically favored in citation forms, as in the variety from France analyzed in Section 3. This preference is due to a phonotactic markedness constraint penalizing utterance-final consonants, as shown in Table 12d (see also Section 3.2).

### 5.3 Deriving connected-speech variants

At the boundary between two words, affrication will potentially result in changes in both Word1 and Word2, due to the bidirectional nature of coarticulation. But this will have different implications in terms of paradigm uniformity with the corresponding citation forms depending on the type of consonant.

Dissimilarities with respect to citation forms are penalized by two new output-output faithfulness constraints: IDENT-OO(CONT) and IDENT-OO(VOI). These constraints correspond to IDENT-IO(CONT) and IDENT-IO(VOI) used in the analysis of morpheme-internal affrication, but are active in the output-output dimension: they penalize dissimilarities between contextual variants and the corresponding citation forms in terms of consonant continuancy and vowel quality, respectively. Table 13 shows how affrication implies different patterns of violations of these constraints depending on whether the consonant is a stable word-initial consonant, a liaison consonant, or a stable word-final consonant.

	Citation forms	Affricated candidate	IDENT-OO (CONT)	IDENT-OO (VOI)	Rate of affrication (Côté 2014)
Word-final C	[tʁɑ̃t], [inosɑ̃]	[tʁɑ̃tʃinosɑ̃]	* <sub>t→ts</sub>	* <sub>i→i̥</sub>	36.5%
Liaison C	[gʁɑ̃], [inosɑ̃]	[gʁɑ̃tʃinosɑ̃]		* <sub>i→i̥</sub>	66.0%
Word-initial C	[vʁɛ], [tʃiʁɑ̃]	[vʁɛtʃiʁɑ̃]			99.2%

Table 13: How affrication affects the similarity with citation forms depending on the type of consonant

For liaison consonants, affrication implies a number of feature changes that is intermediate between the number of feature changes for stable word-final and word-initial consonants. Only the feature change affecting vowel quality at the beginning of Word2 (i → i̥) is penalized by paradigm uniformity. The change in consonant continuancy at the end of Word1 (t → ts) is not penalized because the liaison consonant is missing from the corresponding citation form. For stable word-final consonants, affrication implies two feature changes relative to the corresponding citation form (one on the consonant at the end of Word1 and another one on the vowel at the beginning of Word2). For stable word-initial consonants, affrication does not imply any feature change relative to the corresponding citation form, since affrication already applies (almost) categorically in this form.

As shown in the last column of Table 13, the rate of affrication is inversely correlated with the number of feature changes implied by affrication across the three types of consonants. This can be understood as a paradigm uniformity effect: the grammar militates for uniformity between contextual and citation forms, resulting in less affrication for forms in which affrication would imply more feature changes.

Note that the paradigm uniformity constraints used in the analysis of prosodic variability (RIGHT-ANCHOR-OO in (8) and LEFT-ANCHOR-OO in (9)) will not play a role here. The reason is that all output candidates that do not feature a prosodic break between Word1 and Word2 violate Anchor constraints equally regardless of how consonants are realized segmentally at the junction of the two words. Anchor constraints assess faithfulness for the segment as a whole regardless of its internal feature specifications. For instance, in the absence of a prosodic break between Word1 and Word2, a liaison consonant is in contact with both the right edge of Word1 and the left edge of Word2. As a consequence, it is penalized equally by the two Anchor constraints regardless of whether it is affricated or not. Similarly, candidates involving a word-final consonant followed by a vowel-initial word (as in *trente innocents*) incur the same violations of RIGHT-ANCHOR-OO and LEFT-ANCHOR-OO, regardless of whether the consonant is affricated or not. And the same goes with candidates involving a vowel-final word followed by a consonant-initial word (as in *vrai tyran*).

Table 14 shows how an analysis including IDENT-OO(CONT) and IDENT-OO(VOI) can derive the rates of affrication attested in Côté (2014)'s study for the different types of consonants: liaison consonants (Table 14a), stable final consonants (Table 14b), and stable initial consonants (Table 14c).<sup>17</sup> As in the analysis of prosodic variability, the use of the liaison allomorph is motivated by a markedness constraint penalizing vowel hiatuses (\*VV). The choice of a specific realization for the consonants occurring at the boundary between Word1 and Word2 is driven by the interaction between phonotactic constraints (\*ti, \*tsi), input-output faithfulness constraints (IDENT-IO(CONT) and IDENT-IO(VOI)), and output-output faithfulness constraints (IDENT-OO(CONT) and IDENT-OO(VOI)). IDENT is abbreviated as ID in Table 14.

---

<sup>17</sup>Côté (2014) does not include a phonetic analysis, therefore it is not possible to determine whether vowel reduction always accompanies affrication (candidate (d) in Table 14a, candidate (c) in Table 14b and Table 14b) or not (candidate (c) in Table 14a, candidate (d) in Table 14b and Table 14b) in her corpus. It will here be assumed that affrication is always accompanied by vowel reduction. See Section 6 for evidence in favor of a correlation between these two processes in Quebec French.

URs: /gʁã <sub>1</sub> , gʁã <sub>2</sub> / +/inosã/ CFs: [gʁã]+[inosã]	*tsi	*VV	*ti	ID-IO (cont)	ID-IO (voi)	ID-OO (voi)	ID-OO (cont)	<i>H</i>	Predicted	Observed
a. [gʁã <sub>1</sub> inosã]		1						16.43	0.15	0.15
b. [gʁã <sub>2</sub> inosã]			1					15.81	0.29	0.29
c. [gʁãts <sub>2</sub> inosã]	1			1				56.16	0	0
d. [gʁãts <sub>2</sub> inosã]				1	1	1		15.14	0.56	0.56

(a) Word1 with liaison /t/ + Word2 with initial /i/ (*grand innocent* ‘great innocent’)

URs: /tʁãt/ +/inosã/ CFs: [tʁãt]+[inosã]	*tsi	*VV	*ti	ID-IO (cont)	ID-IO (voi)	ID-OO (voi)	ID-OO (cont)	<i>H</i>	Predicted	Observed
a. [tʁãtinosã]			1					15.81	0.63	0.64
b. [tʁãtsinosã]	1			1			1	57.37	0	0
c. [tʁãtsinosã]				1	1	1	1	16.35	0.37	0.36

(b) Word1 with final /t/ + Word2 with initial /i/ (*trente innocents* ‘thirty innocent (persons)’)

URs: /vʁɛ/ +/tiʁã/ CFs: [vʁɛ]+[tiʁã]	*tsi	*VV	*ti	ID-IO (cont)	ID-IO (voi)	ID-OO (voi)	ID-OO (cont)	<i>H</i>	Predicted	Observed
a. [vʁɛtiʁã]			1			1	1	19.83	0	0.01
b. [vʁɛtsiʁã]	1			1		1		58.97	0	0
c. [vʁɛtsiʁã]				1	1			12.33	1.00	0.99

(c) Vowel-final Word1 + Word2 with initial /ti/ (*vrai tyran* ‘true tyrant’)

Table 14: Connected-speech variants

The analysis derives all three realizations attested in Côté (2014) for liaison words, as shown in Table 14a: absence of liaison consonant (candidate (a)), liaison without affrication (candidate (b)), and liaison with affrication (candidate (d)). It also derives the specific frequencies attested for each of these realizations.

The analysis also does a good job at matching the frequencies of affricated realizations and non-affricated realizations for stable word-final /t/ before /i/ (candidate (c) in Table 14b vs. candidate a in Table 14b). The fact that affrication is less likely for word-final /t/ than for liaison /t/ follows from differences in the way affrication is penalized by output-output faithfulness in these two cases: the candidate with affrication violates IDENT-OO(CONT) in the consonant-final word (candidate (c) in Table 14b) but not in the liaison word (candidate (d) in Table 14a), due to the liaison consonant being absent from the citation form.

The analysis also does a good job at deriving the quasi-categorical affricated realization of word-initial /t/ before /i/ (candidate (c) in Table 14c). The fact that affrication is more likely for word-initial consonants than for liaison consonants follows from differences in the way affrication is penalized by output-output faithfulness in these two cases. Affrication is not penalized at all in the consonant-initial word (candidate (c) in Table 14c) because it has already applied in the corresponding citation form. Affrication is penalized by IDENT-OO(VOI) in the liaison word (candidate (d) in Table 14a): affrication of liaison /t/ correlates with a reduction of the vowel at the beginning of Word2, in violation of paradigm uniformity

([i] is unreduced in the citation form).

## 5.4 Implicational generalizations

The preceding section has shown that the paradigm uniformity analysis can capture the intermediate rate of affrication of liaison consonants as attested in [Côté \(2014\)](#)'s corpus. But the analysis actually derives the intermediate realization of French liaison in (12) as a necessary consequence of the proposed constraint set, and this result obtains regardless of the framework chosen for probabilistic grammars (Stochastic OT, Noisy HG, or MaxEnt).

- (12) Statistical implicational generalizations derived by the analysis
- a.  $P(\text{No affrication}|\text{Liaison C}) \leq P(\text{No affrication}|\text{Final C})$
  - b.  $P(\text{Affrication}|\text{Liaison C}) \leq P(\text{Affrication}|\text{Initial C})$

In words: Liaison consonants are less likely to resist affrication ( $/\text{ti}/ \rightarrow [\text{tsi}]$ ) than final consonants and less likely to undergo affrication than initial consonants.

The demonstration is more involved than in the case of prosodic variability (Section 3.4) because here no candidate is harmonically bounded. The constraint-violation profiles of the different candidates must be carefully compared across the different types of consonants (final, liaison, and initial). In a nutshell, generalization (12a) holds because the candidate with an unaffricated liaison consonant (candidate (b) in Table 14a) fares worse within the candidate set for liaison words than the candidate with an unaffricated final consonant (candidate (a) in Table 14b) does within its own candidate set. And, similarly, generalization (12b) holds because the candidate with an affricated liaison consonant (candidate (d) in Table 14a) fares worse within its candidate set than the candidate with an affricated initial consonant (candidate (c) Table 14c) does within its own candidate set. For more details, the reader is referred to the output of the CoGeTo analysis ([Magri & Anttila 2019](#)) in the supplementary materials. This analysis shows that indeed the implications in (12) hold in Stochastic OT, Noisy HG, and MaxEnt.

## 6 Study 2 : liaison and affrication in Quebec French

One key ingredient in the analysis of Quebec French liaison proposed in Section 5 is the idea that coarticulation affects both C and V in CV, and more specifically that affrication of  $/\text{t}/$  correlates with a reduction of  $/\text{i}/$  in  $/\text{ti}/$  sequences. This hypothesis was crucial to explain why affrication is less likely for liaison consonants than for word-initial consonants.

The goal of this section is to test whether affrication does indeed correlate with vowel reduction across a word boundary, for both stable word-final consonants and liaison consonants. Section 6.1 presents the methods used to test the hypothesis. Section 6.2 presents the results. Section 6.3 concludes with a brief discussion. The data and code for Study 2 are available in [Storme \(2022\)](#) under the names `study2-data.csv` and `study2-code.R`, respectively.

## 6.1 Methods

Data from the Quebec PFC project (Côté 2016) were used to investigate this question. The analysis focuses on one minimal pair from the PFC word lists that feature an underlying sequence /ti/ at the boundary between two words: *grand innocent* ‘great innocent’ (with liaison /t/) and *trente innocents* ‘thirty innocent (people)’ (with stable word-final /t/). These data are particularly interesting because they make it possible to test both whether affrication and high-vowel reduction are correlated and how this correlation might differ for liaison consonants and stable word-final consonants. The analysis does not focus on word-initial /ti/ sequences as the PFC word lists do not include minimal pairs allowing for a controlled comparison with liaison and word-final consonants.

The data from all locations available in the corpus in 2021 were selected, corresponding to a total of 394 participants.<sup>18</sup> Annotations were done manually in Praat (Boersma & Weenink 2021). /t/ duration was used as the acoustic correlate for affrication: an underlying /t/ that is affricated on the surface should be longer than an underlying /t/ that is not affricated. The choice to use duration as the dependent variable was also motivated by the fact that earlier phonetic studies on French liaison also focused on duration (see Section 2.2). The duration of /t/ included the burst and/or frication noise, following Whang (2018: 1163). Vowel reduction was also annotated, using the presence of formant structure as a criterion. In the absence of clear formant structure, no vowel /i/ was included on the corresponding tier. This does not mean that the vowel is completely absent phonetically as phonetic reflexes of /i/ could be present in the burst or frication noise of /t/. Pauses and schwas that sometimes occurred between /t/ and /i/ were also annotated, as well as cases of non-conventional consonant realizations (for instance, some participants pronounced a [z] between *trente* and *innocents*) and cases where the liaison consonant was absent (in these cases, no consonant was annotated on the corresponding tier). Segment durations were extracted automatically using a Praat script.

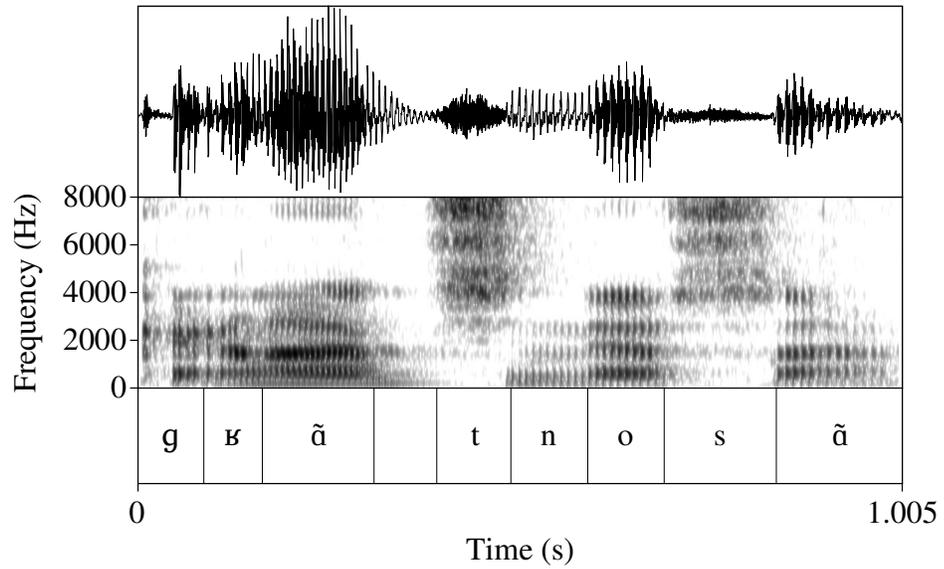
Figure 3 shows the annotation for two occurrences of /ti/ in the corpus. Figure 3a shows an occurrence of liaison /t/ and Figure 3b an occurrence of stable word-final /t/. In Figure 3a, liaison /t/ appears to be characterized by a long frication noise and there is no visible vowel after it, as it is directly followed by [n]. This realization corresponds to the transcription of /ti/ as [ts̥i̯] in Section 5. In Figure 3b, the burst of stable word-final /t/ is much shorter and /i/ is clearly visible in the spectrogram after it. This realization corresponds to the transcription of /ti/ as [ti] in Section 5. These two occurrences thus illustrate the correlation that should be found in the corpus if affrication correlates with vowel reduction in /ti/ sequences.

Only sequences that involve a [t] on the surface (affricated or not) and no pause between the consonant and the vowel were included in the final analyses, corresponding to a total of 322 participants and 494 occurrences of consonants (243 liaison consonants and 251 stable word-final consonants).

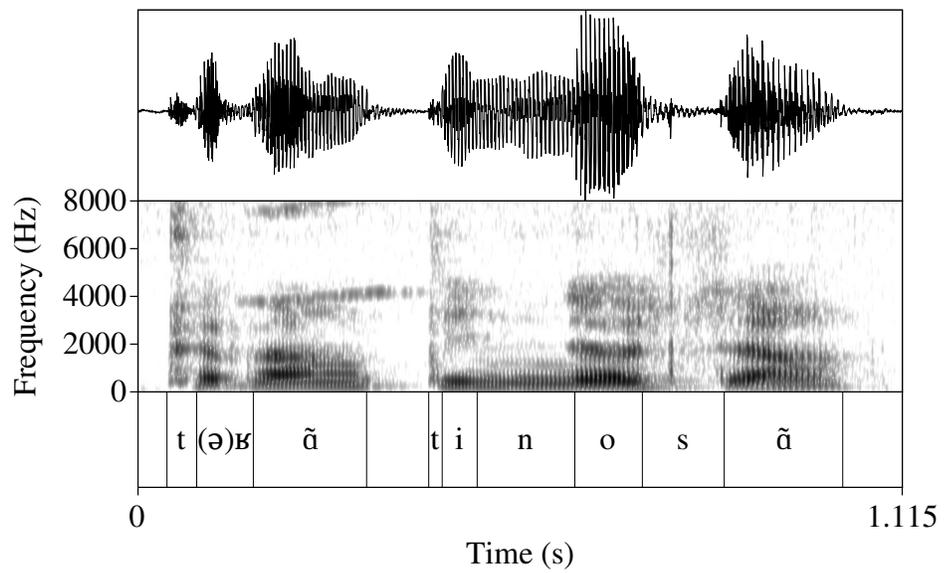
To assess the reliability of the coding of consonant duration and vowel presence, a subset of the stimuli were annotated again by the author two years after the first annotation was carried out. 50 occurrences of the /ti/ sequences included in the final analysis were randomly sampled from the original set and reannotated in Praat. Reliability was assessed using the

---

<sup>18</sup>I am grateful to Marie-Hélène for making the data available to me.



(a) *grand innocent* 'great innocent' (Speaker crbdp1mgx)



(b) *trente innocents* 'thirty innocents' (Speaker cqots1mgx)

Figure 3: Annotation in Praat

intraclass coefficient correlation (ICC) in the `irr` package in R. A good reliability was found for the coding of both consonant duration ( $ICC(A, 1) = 0.869, p < .001$ ) and vowel presence ( $ICC(A, 1) = 0.842, p < .001$ ). The datafile for the reliability study can be found in [Storme \(2022\)](#) under the name `check-study2-data.csv` and the corresponding R script is available under the name `study2-reliability-study.R`.

Two statistical analyses were conducted on the original dataset. A Bayesian logistic regression was fit to the data using the `brms` package in R, with Vowel (present, absent) as the dependent variable and Consonant (liaison, final), Consonant duration and their interaction as independent variables. The goal of this first analysis was to test whether vowel deletion correlates with lengthening of /t/, as expected under the hypothesis that affrication results in high-vowel deletion/reduction.

A Bayesian linear regression was also fit to the data, with Consonant duration as dependent variable and Consonant (liaison, final) as independent variable. The goal of this second analysis was to test whether liaison /t/ is phonetically longer than stable word-final /t/. A greater duration for liaison /t/ is expected if liaison /t/ is more affricated than stable word-final /t/, as reported by [Côté \(2014\)](#) on a perceptive basis, and if liaison consonants are generally longer than stable word-final consonants (see Section 2.2). The analyses did not include random effects because there was at most one occurrence of each type of consonant (liaison, final) per speaker.

## 6.2 Results

The results of the logistic regression confirm the hypothesis that /t/-lengthening correlates with a higher likelihood of /i/-deletion, as shown in Figure 4. An increase of 1 ms in /t/ duration corresponds to a decrease of 0.08 unit ( $CI = [0.06, 0.11]$ ) in the posterior log-odds ratio of /i/-presence. This result was found to hold for both liaison and stable word-final consonants, as the interaction term between duration and consonant type was not significantly different from zero ( $\beta = 0.02, CI = [-0.01, 0.05]$ ). Moreover, liaison /t/ was found to favor /i/-deletion more than than word-final /t/ ( $\beta = -2.25, CI = [-4.31, -0.27]$ ), independently from the effect of duration.

The results of the linear regression show that liaison /t/ is longer on average than word-final /t/ ( $\beta = 11.79, CI = [7.66, 16.01]$ ), as shown in Figure 5. This lengthening corresponds to an increase of 19% in duration. This is compatible with the observation in [Côté \(2014\)](#) that liaison /t/ is more affricated than stable word-final /t/ on average. This is also compatible with earlier observations about the relative duration of liaison consonants and stable word-final consonants more generally (see Section 2).

## 6.3 Discussion

The results of Study 2 support a key hypothesis of the paradigm uniformity account of the segmental variability of Quebec French liaison: affrication at a word boundary results in reduction/deletion of the following vowel. This hypothesis was crucial to explain why the rate of affrication is smaller for liaison /t/ than for word-initial /t/. Reduction of the initial vowel of Word2 after affricated liaison /t/ makes the connected-speech variant of Word2 less similar to its citation form (where reduction does not apply) and therefore is undesirable for

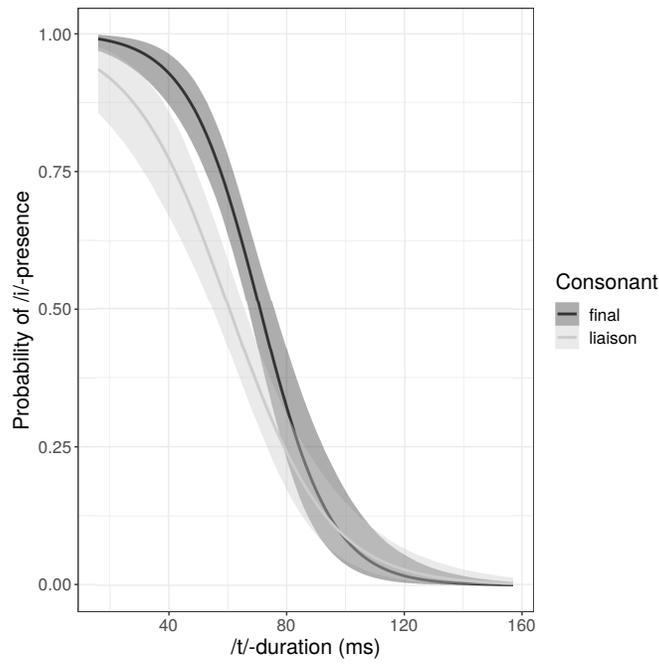


Figure 4: Posterior probability of /i/-deletion as a function of /t/-duration and consonant type (liaison, final)

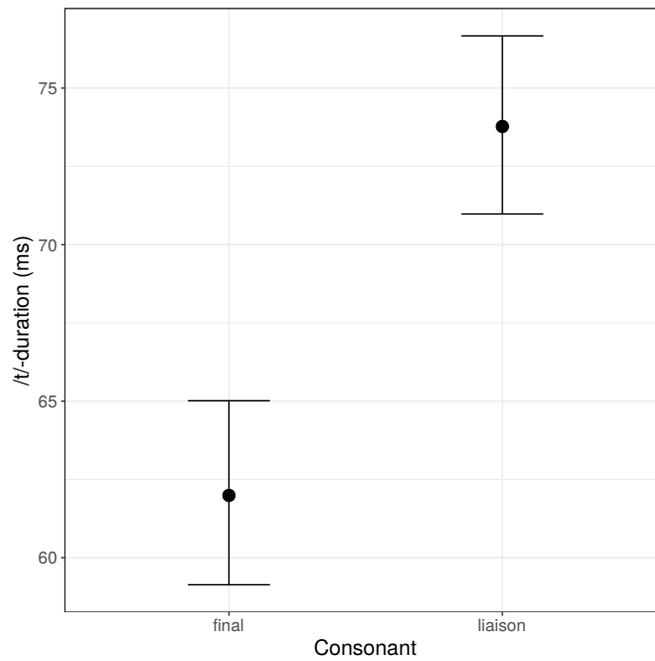


Figure 5: Posterior distribution of consonant duration (ms) for stable word-final /t/ and liaison /t/ (mean and 95% CI)

paradigm uniformity. By contrast, word-initial /ti/ sequences already undergo affrication in their citation forms and therefore there is no reason to block affrication in the corresponding connected-speech variants.

Moreover, the results also support the hypothesis that liaison /t/ is more prone to affricate than stable word-final /t/. In the paradigm uniformity analysis, this follows from the effect of the corresponding citation form. Stable word-final /t/ is influenced by the corresponding unaffricated [t] in the citation form. Liaison /t/ does not correspond to any [t] in the citation form and therefore is less likely to resist affrication.

## 7 Comparison with other analyses

This section proposes to take a step back and look at how the paradigm uniformity analysis of liaison proposed in this paper compares with earlier analyses (see [Côté 2011](#) for an overview). Before doing so, it is important to remind the reader that the analysis presented in this paper addresses only one of the two main research questions on French liaison, namely how liaison consonants are realized when pronounced (=Q1). The other main research question is about which factors affect whether the liaison consonant in a liaison word is pronounced or not (=Q2). This second research question is motivated by the observation that the rate of liaison in Word1-Word2 sequences depends not only on the initial segment of Word2 (vowel/consonant) but also on other phonological and non-phonological properties of Word1 and Word2 as well as language-external factors, as shown in Table 15.

Variables		Source
Morphosyntactic	PoS of Word1 and Word2	<a href="#">Fougeron et al. (2001)</a>
Phonological	Length of Word1	<a href="#">Fougeron et al. (2001)</a> ; <a href="#">Kilbourn-Ceron (2017)</a>
Lexical	Freq of Word1, Word2	<a href="#">Bybee (2001)</a> ; <a href="#">Kilbourn-Ceron (2017)</a> ;
	and Word1-Word2	<a href="#">Kilbourn-Ceron (2017)</a>
	Identity of Word1	<a href="#">Côté (2011)</a>
Sociolinguistic	Identity of Word2	<a href="#">Zuraw &amp; Hayes (2017: section 3)</a>
	Speech style	<a href="#">Fougeron et al. (2001)</a>
	French variety	<a href="#">Côté (2017)</a>
	Speaker identity	<a href="#">Encrevé (1988)</a> ; <a href="#">Meinschaefer et al. (2015)</a>

Table 15: A non-exhaustive list of variables reported to condition the rate of liaison along with a non-exhaustive list of sources (PoS = part of speech, Freq = frequency).

Analyses of French liaison differ in their scope: some of them, like the present analysis, only focus on Q1, others only on Q2 and finally some theories deal with both Q1 and Q2. Table 16 presents a sample of analyses of French liaison, with the first column indicating which research question they focus on. Analyses may also differ in how they answer the research question. Analyses of French liaison can be conveniently divided into two broad categories, depending on whether the explanatory burden relies primarily on representations (a richer phoneme inventory and/or a richer lexicon) or on computation (a richer constraint set), as indicated in the three rightmost columns of Table 16. Although the production planning hypothesis proposed by [Kilbourn-Ceron \(2017\)](#) is not directly formulated in terms of constraints, it can be implemented this way, as will be shown in Section 7.4.

	Research question	Enriched phoneme inventory	Enriched lexicon	Enriched constraint set
Floating consonants <sup>a</sup>	Q1	yes	no	no
Lexical constructions <sup>b</sup>	Q1, Q2	no	yes	no
Gradient symbolic representations <sup>c</sup>	Q1, Q2	yes	yes	no
Paradigm uniformity <sup>d</sup>	Q1	no	no	yes
Production planning <sup>e</sup>	Q2	no	no	yes

<sup>a</sup>Encrevé 1988; Tranel 1990, 2000; <sup>b</sup>Bybee 2001; Chevrot et al. 2009, <sup>c</sup>Smolensky & Goldrick 2016; Smolensky et al. 2020; Tessier & Jesney 2021; <sup>d</sup>Steriade 1999; this paper; <sup>e</sup>Kilbourn-Ceron 2017

Table 16: Analyses of French liaison: a typology. Q1: How are liaison consonants realized when present? Q2: Which factors condition the rate of liaison before vowel-initial words?

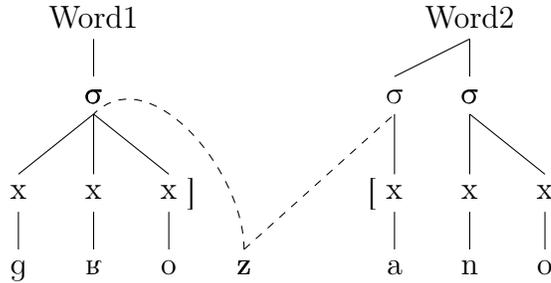
Sections 7.1-7.3 review the first three analyses listed in Table 16 and highlight how they compare to the paradigm uniformity analysis in accounting for Q1. Section 7.4 concludes by sketching how insights from Kilbourn-Ceron’s production planning hypothesis can be combined with the paradigm uniformity analysis to get a more comprehensive model of French liaison accounting for both Q1 and Q2.

## 7.1 Floating consonants

Liaison consonants have been analyzed as floating segments by several researchers, including by Encrevé (1988: 169-173) and Tranel (1990: 183-184) in the framework of autosegmental phonology and by Tranel (2000: 49-52) in the framework of Optimality Theory. In these approaches, liaison consonants are lexically affiliated to the first word but differ from stable word-final consonants in not being attached to the word’s early prosodic structure. This property allows them to be associated at a later stage of prosodic-structure building to either Word1 or Word2 in a Word1-Word2 sequence (Encrevé 1988: 182).

The specific proposal advanced by Tranel (1990) is represented in (13). In this analysis, liaison /z/ does not project a skeletal slot and is therefore ‘floating’ at the end of Word1 in the early prosodic structure. When the two words are combined together, the liaison consonant has to be attached somewhere prosodically. Rightward syllabification attaches it at the beginning of Word2, making it a *liaison enchaînée*. Leftward syllabification attaches it at the end of Word1, making it a *liaison non-enchaînée*. Liaison words that use the feminine form as liaison variant (e.g. *bel* [bɛl] ‘beautiful’) are analyzed differently: in these words, the final consonant is attached to the word’s early prosodic structure (Tranel 1990: 183) and therefore behaves like other stable word-final consonants (see Section 4 on the distinction between epenthetic and suppletive liaisons).

- (13) Liaison consonants as floating segments: *gros anneau* ‘big ring’ (based on Tranel 1990: 184)



Analyses using floating segments are the closest in scope to the paradigm uniformity analysis: their focus is also on deriving the variable realization of liaison consonants before vowel-initial words. However they differ from the paradigm uniformity analysis because they build the variable behavior of liaison consonants directly into their phonological representations. These approaches indeed require an enrichment of the French phoneme inventory with a new set of phonemes. For instance, [Tranel \(2000: 51-52\)](#) introduces a phonological feature to distinguish stable consonants (noted as C) from liaison consonants (noted as L).

By contrast, the paradigm uniformity analysis adopts the exact same phonological representation for liaison and non-liaison consonants. The difference between liaison consonants and other consonants ultimately emerges from differences in *lexical* representations: liaison words have several listed allomorphs varying by the presence/absence of a final consonant whereas other words don't (see Section 3). And the fact that epenthetic liaison and suppletive liaison behave differently emerges from differences in the similarity between the liaison allomorph and the corresponding citation form (see Section 4).

One empirical advantage of the paradigm uniformity analysis is that it accounts for both types of variability (prosodic and segmental). By contrast, the approach using floating segments currently only accounts for prosodic variability. If liaison consonants become identical to onset consonants after rightward resyllabification, then it is unclear why they should pattern differently from stable word-initial consonants segmentally in this context, as documented in Section 2.2. The same problem would arise for approaches using alignment constraints (e.g. [Zuraw & Hayes 2017](#); see the discussion in Section 3.3).

As pointed out by a reviewer, one possibility would be to exploit the fact that there remains a difference in the underlying representations of liaison and non-liaison consonants in Tranel's analysis: one projects a skeletal slot whereas the other does not, as shown in (13). For Quebec French, a rule of affrication that is sensitive to this underlying difference could then be entertained. However it is unclear why this affrication rule would necessarily have a probability of application that is intermediate between the rules applying to stable word-final consonants and to stable word-initial consonants. In the paradigm uniformity analysis, this asymmetry follows as an implicational generalization from independently motivated principles of coarticulation and paradigm uniformity with citation forms (see Section 5).

## 7.2 Lexical constructions

In the approach using lexical constructions, the liaison consonant belongs neither to Word1 nor to Word2 but to a construction involving the two words ([Bybee 2001](#)). For instance, there is a lexical construction of the form  $/g\beta\tilde{a}_{Adj} t X_{V-initial} N/$  'great N', where  $X_{V-initial}$  is

a vowel-initial noun and /t/ a consonant occurring between the Adj and the N. Nouns that are more frequently associated with the adjective *grand* ‘great’ are more likely to be stored under this frame, explaining for instance why the likelihood of the liaison consonant increases with the frequency of the Word1-Word2 sequence (see [Fougeron et al. 2001](#); [Kilbourn-Ceron 2017](#)).

Although lexical constructions are primarily motivated by the type of frequency effects reported in Table 15 (e.g. the rate of liaison is higher for liaison words with higher lexical frequency), Bybee mentions in passing that they can also account for the prosodic variability of French liaison. She argues that a prosodic break may intervene in the middle of a lexical construction in the same way as it may intervene in the middle of a word. For instance, it is possible to say *un élé phant* [ɛ̃nele|fɑ̃] ‘an ele (prosodic break) fant’ with a prosodic break in the middle of the word *éléphant*. *Liaison non-enchaînée* and *liaison enchaînée* would then correspond to situations where the prosodic break within a lexical construction intervenes after and before the liaison consonant, respectively.

This approach suffers from the same limitation as the approach using floating consonants in that it does not account for segmental variability. Moreover, it potentially presents another problem. Stable word-final consonants and word-initial consonants do not seem to be separable from their lexical host prosodically, even in high-frequency two-word sequences. For instance, a prosodic break seems much more natural after the stable final consonant of Word1 than before it in the compound *porte-avion* ‘aircraft carrier’ (*porte euh avion* [pɔʁt|œ|avjɔ̃]/\**por euh tavion* [pɔʁ|œ|tavjɔ̃]).

[Bybee \(2001\)](#) sketches an explanation for why this does not happen: ‘Since the words of a construction are usually associated with other instances of the same word, their identity as words is known, and the point between two words is a possible place to pause.’ In other words, a pause is more likely to occur between words than within words inside a multiple-word construction because the word forms inside this construction stand in correspondence with their base forms (which are independently stored outside of any construction). In other words, stable word-final [t] in *porte-avion* cannot be resyllabified across a prosodic break because there is a pressure from the base form *porte* to maintain the [t] at the end of Word1. Liaison consonants are not subject to the same pressure because they are absent from the base forms of Word1 and Word2. When fleshed out, this explanation must refer to principles of paradigm uniformity among morphologically related forms. This means that an analysis using lexical constructions alone is not sufficient to derive the variable realization of liaison and must be supplemented with a mechanism allowing for paradigm uniformity.

However the lexical-construction approach has an advantage over the paradigm uniformity analysis presented in this paper in that it can account for lexical effects on the rate of liaison. This difference derives from the fact that multiple-word constructions are stored in the lexical-construction analysis whereas they are not in the paradigm uniformity analysis presented in Sections 3 and 5. Although paradigm uniformity could in principle be combined with lexical constructions to get a more comprehensive model of French liaison, it is not immediately clear how to reconcile the differences in lexical representations between the two analyses. Liaison consonants do belong to Word1 underlyingly in the paradigm uniformity analysis (see Table 5 in Section 3) whereas they do not in Bybee’s analysis. Section 7.4 will show that the production planning hypothesis proposed by [Kilbourn-Ceron \(2017\)](#) is more directly compatible with the assumption that liaison consonants belong to Word1 and

therefore provides a better fit to the paradigm uniformity analysis proposed in this paper.

### 7.3 Gradient symbolic representations

In the approach using gradient symbolic representations, phonological representations of phonemes are enriched with an activity degree ranging from zero to one (Smolensky & Goldrick 2016; Smolensky et al. 2020). In the evaluation of input-output mappings, faithfulness violations are multiplied by the activity degree of segments in the input. This approach predicts that phonemes with a lower activity degree will be less protected by faithfulness and therefore more likely to undergo a phonological process. This approach derives the higher likelihood of deletion for liaison consonants as compared to stable consonants by assuming a lower activity degree for liaison segments underlyingly: liaison consonants have an activity degree strictly smaller than 1 whereas stable consonants have an activity degree equal to 1.

Two further assumptions are made in order to derive the prosodic variability of liaison consonants (Q1) and lexical effects on the rate of liaison (Q2). Liaison consonants are assumed to be stored both at the end of liaison words and at the beginning of all vowel-initial words. This explains why they might be realized both at the end of Word1 (*liaison non-enchaînée*) and at the beginning of Word2 (*liaison enchaînée*). Furthermore, liaison consonants are assumed to have word-specific activity degrees. Together with the assumption that liaison consonants are present both in Word1 and Word2 underlyingly, this allows for the rate of liaison to depend on the identity of Word1 and Word2.<sup>19</sup>

The representational assumptions of this analysis are illustrated in (14), using the sequence *grand innocent* as an example. The degree of activity of liaison consonants is indicated as a subscript (the specific activity degrees provided in (14) are only indicative). For stable consonants, the degree of activity is always equal to 1 and therefore not explicitly indicated. A word like *innocent* [inosã] ‘innocent’ is stored with all the possible liaison consonants that can be attached to it as first segment. When the words *grand* and *innocent* are combined, the activity level of liaison /t/ increases, due to the presence of /t/ at the end of Word1 and at the beginning of Word2. This increase in the activity degree of /t/ allows it to surface with some probability. Because /t/ is underlyingly present in both words, it can surface either at the end of Word1 or at the beginning of Word2.

- (14) Liaison consonants as gradient phonemes affiliated to both Word1 and Word2 (Smolensky & Goldrick 2016)
- $$/g\mathfrak{B}\tilde{a}t_{0.48}/ + / \{t_{0.09}, z_{0.09}, \mathfrak{n}_{0.09}, \mathfrak{B}_{0.09}, \mathfrak{p}_{0.09}\}inos\tilde{a}/$$

This approach derives some lexical effects on the rate of liaison (but not all; see Footnote 19) and accounts for the prosodic instability of liaison. However, as articulated to this point in time, it fails to derive the intermediate behavior of liaison in terms of segmental realization. The core intuition behind the account using gradient symbolic representations is the idea that liaison consonants are less protected by faithfulness than stable consonants. As a consequence, liaison consonants should always be more likely to undergo a sound process

<sup>19</sup>As noted by Smolensky et al. (2020), these assumptions are not sufficient to derive all lexical effects on the rate of liaison. In particular, they do not allow for interactive effects of Word1 and Word2 on the rate of liaison. Yet such effects have been reported in the literature (see Table 15).

than stable consonants. This prediction happens to be correct regarding deletion (liaison consonants are more likely to delete than stable consonants) but not regarding affrication (liaison /t/ is less likely to affricate than stable initial /t/). This problematic prediction is made because, due to its lower activity degree, a liaison /t/ that is affricated on the surface incurs strictly fewer faithfulness violations than a stable initial /t/ that is also affricated on the surface. For instance, affricating liaison /t/ in *grand innocent* in (14) would violate IDENT-IO(CONT)  $0.48 + 0.09 = 0.57$  times whereas affricating initial /t/ in *timide* would violate it one time (the degree of activity of a stable consonant is equal to 1). As a consequence, liaison /t/ should be more likely to affricate than initial /t/. But this is not the case (see Côté 2014 and Section 2.2). The mechanism proposed to account for the rate of liaison therefore turns out to be problematic for an account of its intermediate segmental realization.

## 7.4 Towards a comprehensive model of French liaison

The paradigm uniformity analysis provides the most comprehensive account of the *realization* of liaison consonants (Q1). Among the four approaches reviewed in this paper, it is the only one that can derive the variability of liaison in both prosodic and segmental terms. The approach using lexical constructions does not account for the variability of liaison without an accompanying mechanism allowing for associations among different realizations of the same lexical item. The approach using floating segments accounts for the prosodic variability of liaison but not for its segmental variability. The approach using gradient symbolic representations accounts for prosodic variability but makes incorrect predictions about the segmental realization of liaison consonants vis-à-vis stable consonants.

However the paradigm uniformity analysis presented in this paper does not account for the factors that affect the *rate* of liaison (Q2), contrary to the lexical-construction approach and the approach using gradient symbolic representations. One could attempt to combine these different approaches to get a more comprehensive model of French liaison. However they are based on very different assumptions about underlying phonological/lexical representations, making it difficult to find a common ground.

The analysis most compatible with the representational assumptions of the paradigm uniformity analysis offered in this paper is the production planning hypothesis (Wagner 2012; Tanner et al. 2017) as applied to French liaison by Kilbourn-Ceron (2017). Like the paradigm uniformity analysis, this analysis does not need to depart from traditional phonological/lexical representations. Moreover, it shares the same constraint-based orientation: the rate of liaison is determined by linguistic and cognitive constraints on speech-production planning.

According to the production planning hypothesis, an external sandhi process such as French liaison arises at the junction between two words when those two words are planned in the same planning window. And two words are more likely to be planned together if they are easy to retrieve from memory. The fact that the rate of liaison is higher in Word1-Word2 sequences in which Word1 and Word2 are frequent, contextually predictable, and short is compatible with the production planning hypothesis because these properties all facilitate word-form retrieval (Kilbourn-Ceron 2017: chapter 4).

Insights from the production-planning hypothesis can be incorporated into the paradigm

URs: /gʁã₁, gʁãt₂/+/ano/ Bases: [gʁã]+[ano]	$H$	Predicted	URs: /gʁã₁, gʁãt₂/+/ami/ Bases: [gʁã]+[ami]	$H$	Predicted
a. [gʁã₁ ano]	2.63	0.73	a. [gʁã₁ ami]	2.63	0.51
b. [gʁãt₂ ano]	4.78	0.08	b. [gʁãt₂ ami]	4.78	0.06
c. [gʁã t₂ano]	4.00	0.19	c. [gʁã t₂ami]	2.80	0.43
d. [gʁã₁l ano]	54.78	0.00	d. [gʁã₁l ami]	54.78	0.00
e. [gʁã₁ lano]	54.00	0.00	e. [gʁã₁ lami]	52.80	0.00

(a) *grand anneau* ‘large ring’

$$w_{\text{LEFT-ANCHOR-OO}_{\text{low predictability}}} = 4.00$$

(b) *grand ami* ‘great friend’

$$w_{\text{LEFT-ANCHOR-OO}_{\text{high predictability}}} = 2.80$$

Table 17: How an analysis including indexed output-output faithfulness constraints accounts jointly for the rate and realization of liaison. The vertical bar | between the two strings in each candidate represents a prosodic break

uniformity analysis by indexing output-output faithfulness constraints to properties that are relevant for production planning (see Pater 2007 on constraint indexation). A word that is harder to retrieve from memory will tend to be planned in its own planning window, and therefore to be less connected to adjacent words. In other words, it will tend to be more similar to the corresponding citation form. In the constraint-based model proposed in this paper, the parameters that regulate the similarity with citation forms are the weights of output-output faithfulness constraints. The effects that influence production planning can thus be captured by weighting output-output faithfulness constraints higher for words that are harder to retrieve. For instance, words with low lexical frequency would be evaluated by a set of output-output faithfulness constraints that have higher weights than the output-output faithfulness constraints evaluating words with high lexical frequency. As mentioned in Section 3.3, lexical indexation of paradigm uniformity constraints has already been proposed by Zuraw & Hayes (2017) to account for the greater resistance of French *h-aspiré* words to external sandhi processes. Generalizing this approach to other types of properties, and in particular to properties that are relevant to speech production, looks like a promising avenue to model the rate of French liaison.

Table 17 illustrates this approach using a toy example based on the analysis of prosodic variability from Section 3.3. The analysis focuses on one lexical property that has been shown to affect the rate of liaison, namely the conditional probability (or predictability) of Word2 given Word1. The production planning hypothesis predicts that, in Word1-Word2 sequences, the liaison allomorph of Word1 should become more likely as the contextual probability of Word2 given Word1 increases. This prediction is made because Word2 should be easier to retrieve and therefore more likely to be planned together with Word1 if Word2 is more predictable contextually. In line with this prediction, Kilbourn-Ceron (2017: 146) found that a higher conditional probability of Word2 given Word1 correlates with a higher rate of liaison.

The analysis in Table 17 compares the pronunciation of *grand anneau* ‘large ring’ and *grand ami* ‘great friend’. The two sequences differ by the conditional probability of Word2 given Word1: after *grand*, *anneau* is less frequent than *ami* and therefore more likely to be planned in a separate planning window from the preceding word, according to the production-planning hypothesis. In the constraint-based model of paradigm uniformity, this

translates into the following conditions. The contextual pronunciation of *anneau* is evaluated by a constraint that penalizes dissimilarities between connected-speech variants and the corresponding citation forms for words with low contextual predictability (LEFT-ANCHOR-OO<sub>low predictability</sub>). Furthermore, this constraint has a higher weight than the constraint that evaluates the pronunciation of words with high contextual predictability such as *ami* (LEFT-ANCHOR-OO<sub>high predictability</sub>). Table 17 shows that this approach correctly predicts that the rate of liaison should be smaller before the less predictable word (27% before *anneau* vs. 49% before *ami*)<sup>20</sup> while still accounting for the prosodic variability of French liaison (the liaison consonant can attach at the end of Word1 or at the beginning of Word2 in both *grand ami* and *grand anneau*). All other constraints and constraint weights are the same as in the analysis presented in Section 3.3.

The toy analysis in Table 17 also shows that speech-production planning is predicted to affect not only the rate of liaison but also its prosodic realization: Word1-attachment of liaison consonants should be proportionally more likely when Word2 is contextually less probable (because, in this case, the output-output faithfulness constraint evaluating Word2 has a high weight and will therefore strongly repel the liaison consonant). And the same is expected to hold for the *segmental* realization of liaison: for instance, liaison /t/ should have a lower rate of affrication when Word2 is contextually less probable. These predictions should be tested in future work.

## 8 Conclusion

French liaison consonants have been shown in previous research to pattern in an intermediate way between stable word-final and word-initial consonants. The present paper has shown that it is not necessary to attribute this behavior to differences in the underlying phonological status of liaison consonants. Rather it can be derived from independently motivated principles of uniformity among paradigmatically related forms (contextual variants of a word and the corresponding citation form). Also, the variable behavior of liaison consonants can be derived without positing lexical constructions or massive allomorphy in the lexicon. It is sufficient to assume that liaison words have just two listed allomorphs. An explicit implementation of the analysis in a probabilistic constraint-based grammar was proposed and shown to be able to derive the variable behavior of French liaison in terms of both prosodic and segmental properties. Crucially, the analysis assumed standard lexical and phonological representations as inputs. In the end, liaison is only one among the many types of phonologically optimizing suppletion found in French, such as [ɛ]-[ə] alternations (Storpe 2020a). Its puzzling behavior comes from the way suppletion interacts with paradigm uniformity at word edges.

Quantitative evidence was provided for two important hypotheses of the paradigm uniformity analysis. Study 1 showed that liaison consonants are variable in their prosodic attachment only if they are absent from the corresponding citation form, thus making a clear argument for the role of paradigm uniformity with citation forms. Study 2 provided evidence for the phonetic mechanism that underlies the paradigm uniformity analysis of the

---

<sup>20</sup>These numbers are only indicative and do not represent the actual rates of liaison for these specific word sequences.

segmental variability of liaison in Quebec French. Affrication of /t/ was found to correlate with a higher likelihood of high-vowel reduction in Quebec French, for both liaison and stable word-final consonants. This result is in line with the hypothesis that affrication at word boundaries has consequences for uniformity with the citation forms of both Word1 (through a change affecting word-final /t/) and Word2 (through a change affecting word-initial /i/). This hypothesis was key to explain why liaison /t/ is less likely to affricate than word-initial /t/ before /i/.

There are two ways in which the paradigm uniformity analysis of French liaison could be further developed and evaluated. First, it should be extended to account for patterns of segmental realization that involve continuous phonetic representations. This will require moving away from *symbolic* constraint-based grammars and adopting *phonetic* constraint-based grammars (Flemming 2001) instead. Second, the analysis presented in this paper mainly focused on deriving the variable realization of French liaison. But a comprehensive model of French liaison should also include the factors that have been reported to affect the rate of liaison. Section 7.4 has sketched how some of these effects could be captured by indexing output-output faithfulness constraints to properties that are relevant in speech production, according to the production planning hypothesis of external sandhi (Kilbourn-Ceron 2017). Future work should test further the predictions of this approach.

## References

- Anttila, Arto & Giorgio Magri. 2018. Does MaxEnt overgenerate? Implicational universals in Maximum Entropy Grammar. In *AMP 2017: Proceedings of the 2017 Annual Meeting on Phonology*, ed. Gillian Gallagher, Maria Gouskova, & Yin Sora. Washington, DC: Linguistic Society of America.
- Bagou, Odile, Violaine Michel & Marina Laganaro. 2009. On the production of sandhi phenomena in French: Psycholinguistic and acoustic data. In *Tenth Annual Conference of the International Speech Communication Association*, 452–455.
- Beckman, Mary & Atsuko Shoji. 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41:61–71.
- Benua, Laura. 1997. Transderivational identity: phonological relations between words. Doctoral Dissertation, University of Massachusetts, Amherst.
- Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:45–86.
- Boersma, Paul & David Weenink. 2021. Praat: doing phonetics by computer [computer program]. version 6.1.41. URL <http://www.praat.org/>, retrieved 25 March 2021.
- Bonami, Olivier & Gilles Boyé. 2005. Construire le paradigme d’un adjectif. *Recherches linguistiques de Vincennes* 77–98.
- Broad, David J & Frantz Clermont. 1987. A methodology for modeling vowel formant contours in CVC context. *Journal of the Acoustical Society of America* 81:155–165.

- Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80:1–28.
- Bybee, Joan L. 2001. Frequency effects on French liaison. In *Frequency and the emergence of linguistic structure*, ed. Joan L. Bybee & Paul J. Hopper, 337–359. Amsterdam: John Benjamins Publishing Company.
- Cedergren, Henrietta J. & Louise Simoneau. 1985. La chute des voyelles hautes en français de Montréal: "As-tu entendu la belle syncope?". In *Les tendances dynamiques du français parlé à Montréal*, ed. Monique Lemieux & Henrietta J. Cedergren, volume 1, 57–145. Quebec: Office de la langue française.
- Chevrot, Jean-Pierre, Céline Dugua & Michel Fayol. 2009. Liaison acquisition, word segmentation and construction in French: A usage-based account. *Journal of Child Language* 36:557–596.
- Côté, Marie-Hélène. 2005. Le statut lexical des consonnes de liaison. *Langages* 158:66–78.
- Côté, Marie-Hélène. 2011. French liaison. In *The Blackwell Companion to Phonology*, ed. Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume, & Keren Rice, 1–26. Chichester, UK: John Wiley and Sons Ltd.
- Côté, Marie-Hélène. 2014. Liaison et assibilation en français laurentien. In *La liaison: approches contemporaines*, ed. Christiane Soum-Favaro, Annelise Coquillon, & Jean-Pierre Chevrot, 9–32. Berne: Peter Lang.
- Côté, Marie-Hélène. 2016. Variation in Canada: Trois-Rivières in Quebec. In *Varieties of spoken French*, ed. Sylvain Detey, Jacques Durand, Bernard Laks, & Chantal Lyche, 449–462. Oxford: Oxford University Press.
- Côté, Marie-Hélène. 2017. La liaison en diatopie: esquisse d’une typologie. *Journal of French Language Studies* 27:13–25.
- De Cat, Cécile. 2007. *French dislocation. interpretation, syntax, acquisition*. Oxford: Oxford University Press.
- Dejean de la Bâtie, Bernadette & Dianne C Bradley. 1995. Resolving word boundaries in spoken French: Native and non-native strategies. *Applied Psycholinguistics* 16:59–81.
- Delais-Roussarie, Elisabeth, Brechtje Post, Mathieu Avanzi, Carolin Buthke, Albert Di Cristo, Ingo Feldhausen, Sun-Ah Jun, Philippe Martin, Trudel Meisenburg, Annie Rialland, Rafêu Sichel-Bazin & Hi-Yon Yoo. 2015. Intonational phonology of French: Developing a ToBI system for French. In *Intonation in Romance*, ed. Sónia Frota & Pilar Prieto. Oxford: Oxford University Press.
- Delattre, Pierre C. 1947. La liaison en français, tendances et classification. *The French Review* 21:148–157.
- Dell, François. 1985. *Les règles et les sons*. Paris: Hermann, 2nd edition.

- Dow, Michael. 2019. Mind your /ti/'s and q's: A subsegmental approach to affrication in Québec French. *Toronto Working Papers in Linguistics* 41. URL <https://twpl.library.utoronto.ca/index.php/twpl/article/view/32770>.
- Durand, Jacques & Chantal Lyche. 2008. French liaison in the light of corpus data. *Journal of French Language Studies* 18:33–66.
- Durand, Marguerite. 1936. *Le genre grammatical en français parlé Paris et en région parisienne*. Paris: Bibliothèque du Français Moderne.
- Encrevé, Pierre. 1988. *La liaison avec et sans enchaînement: phonologie tridimensionnelle et usages du français*. Paris: Seuil.
- Eychenne, Julien. 2019. On the deletion of word-final schwa in Southern French. *Phonology* 36:355–389.
- Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18:7–44.
- Flemming, Edward. 2021. Comparing MaxEnt and Noisy Harmonic Grammar. *Glossa: a journal of general linguistics* 6. 141.
- Fougeron, Cécile. 2007. Word boundaries and contrast neutralization in the case of enchaînement in French. In *Papers in Laboratory Phonology 9: Change in Phonology*, ed. Jennifer Cole & José Ignacio Hualde, 609–642. Berlin: Mouton de Gruyter.
- Fougeron, Cécile & Elisabeth Delais-Roussarie. 2004. Liaisons et enchaînements «fais\_en à fez\_en parlant». *Actes des Journées d'Etudes sur la Parole* 221–224.
- Fougeron, Cécile, Jean-Philippe Goldman, Alicia Dart, Laurence Guélat & Clémentine Jaeger. 2001. Influence de facteurs stylistiques, syntaxiques et lexicaux sur la réalisation de la liaison en français. In *Actes de la 8ème conférence sur le Traitement Automatique des Langues Naturelles. Articles longs*, 172–181. Tours, France: ATALA.
- Gaatone, David. 1978. Forme sous-jacente unique ou liste d'allomorphes? (à propos des consonnes de liaison en français). *Linguistics* 214:33–54.
- Gaskell, M. Gareth, Elsa Spinelli & Fanny Meunier. 2002. Perception of resyllabification in French. *Memory & Cognition* 30:798–810.
- Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, & Östen Dahl, 111–120. Stockholm: Stockholm University, Department of Linguistics.
- Hayes, Bruce. 2017. Varieties of Noisy Harmonic Grammar. In *Proceedings of the 2016 Annual Meeting on Phonology*, ed. Karen Jesney, Charlie O'Hara, Caitlin Smith, & Rachel Walker. Washington, DC: Linguistic Society of America.

- Hayes, Bruce, Bruce Tesar & Kie Zuraw. 2013. OTSoft 2.5. URL <http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.
- Inkelas, Sharon. 2014. *The interplay of morphology and phonology*. Oxford: Oxford University Press.
- Kager, René. 1999. *Optimality theory*. Cambridge: Cambridge University Press.
- Kawahara, Shigeto. 2002. Similarity among variants: Output-variant correspondence. Bachelor's thesis, International Christian University (Tokyo, Japan).
- Kenstowicz, Michael. 1996. Base-identity and uniform exponence: Alternatives to cyclicity. In *Current trends in phonology: Models and methods*, ed. Jacques Durand & Bernard Laks, 363–393. University of Salford: European Studies Research Institute.
- Kilbourn-Ceron, Oriana. 2017. Speech production planning affects variation in external sandhi. Doctoral Dissertation, McGill University.
- Kruschke, John K. & Torrin M. Liddell. 2018. Bayesian data analysis for newcomers. *Psychonomic bulletin & review* 25:155–177.
- LimeSurvey. 2012. LimeSurvey: An open source survey tool. URL <http://www.limesurvey.org>.
- Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35:1773–1781.
- Lison, Pierre & Jörg Tiedemann. 2016. OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*. European Language Resources Association.
- Magri, Giorgio & Arto Anttila. 2019. CoGeTo: Convex geometry tools for typological analysis in categorical and probabilistic constraint-based phonology (version 1.0). URL <https://cogeto.stanford.edu/about>.
- Makowski, Dominique, Mattan S. Ben-Shachar & Daniel Lüdecke. 2019. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software* 4:1541.
- Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry* 38:715–735.
- Meinschaefter, Judith, Sven Bonifer & Christine Frisch. 2015. Variable and invariable liaison in a corpus of spoken French. *Journal of French Language Studies* 25:367–396.
- Myers, Scott & Jaye Padgett. 2014. Domain generalisation in artificial language learning. *Phonology* 31:399–433.

- Nguyen, Noël, Sophie Wauquier-Gravelines, Leonardo Lancia & Betty Tuller. 2007. Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications. In *Segmental and prosodic issues in Romance phonology*, ed. Pilar Prieto, Joan Mascaró, & Marie-Josep Solé, 3–23. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. *Papers in laboratory phonology* 1:258–275.
- Pater, Joe. 2007. The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In *University of massachusetts occasional papers in linguistics 32: Papers in optimality theory iii*, ed. Leah Bateman, Michael O’Keefe, Ehren Reilly, & Adam Werle. Amherst, MA: GLSA.
- R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Roettger, Timo B., Bodo Winter, Sven Grawunder, James Kirby & Martine Grice. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43:11–25.
- Smolensky, Paul & Matthew Goldrick. 2016. Gradient symbolic representations in grammar: The case of French liaison. *Ms. Johns Hopkins University and Northwestern University. Available as ROA 1286*.
- Smolensky, Paul & Géraldine Legendre. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.
- Smolensky, Paul, Eric Rosen & Matthew Goldrick. 2020. Learning a gradient grammar of French liaison. In *Proceedings of the 2019 Annual Meeting on Phonology*, ed. Hyunah Baek, Chikako Takahashi, & Alex Hong-Lun Yeung. Washington, DC: Linguistic Society of America.
- Spinelli, Elsa, Anne Cutler & James M. McQueen. 2002. Resolution of liaison for lexical access in French. *Revue française de linguistique appliquée* 7:83–96.
- Spinelli, Elsa, James M. McQueen & Anne Cutler. 2003. Processing resyllabified words in French. *Journal of Memory and Language* 48:233–254.
- Steriade, Donca. 1997. Phonetics in phonology: the case of laryngeal neutralization. Manuscript.
- Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison. In *Formal perspectives in Romance linguistics. Selected papers from the 28th linguistic symposium on Romance languages*, ed. Jean-Marc Authier, Barbara E. Bullock, & Lisa A. Reed, 243–270. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, ed. Michael B. Broe & Janet B. Pierrehumbert. Cambridge: Cambridge University Press.
- Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA/London: The MIT Press.
- Storme, Benjamin. 2020a. Evidence for suffix cohesion in French. *Journal of French Language Studies* 1–29.
- Storme, Benjamin. 2020b. Gradient behavior without gradient underlying representations: the case of French liaison. In *Proceedings of the 2019 Annual Meeting on Phonology*, ed. Hyunah Baek, Chikako Takahashi, & Alex Hong-Lun Yeung. Washington, DC: Linguistic Society of America.
- Storme, Benjamin. 2022. Paradigm uniformity effects on French liaison. [Data set]. URL <https://doi.org/10.5281/zenodo.7971975>.
- Sussman, Harvey M., Helen A. McCaffrey & Sandra A. Matthews. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America* 90:1309–1325.
- Tanner, James, Morgan Sonderegger & Michael Wagner. 2017. Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology* 8.
- Tessier, Anne-Michelle & Karen Jesney. 2021. Learning French liaison with gradient symbolic representations: Errors, predictions, consequences. In *Proceedings of the 2020 Annual Meeting on Phonology*, ed. Ryan Bennett, Richard Bibbs, Mykel Loren Brinkerhoff, Max J. Kaplan, Stephanie Rich, Amanda Rysling, Nicholas Van Handel, & Maya Wax Cavallaro.
- Tranel, Bernard. 1990. On suppletion and French liaison. *Probus* 2:87–104.
- Tranel, Bernard. 2000. Aspects de la phonologie du français et la théorie de l’optimalité. *Langue française* 126:39–72.
- Wagner, Michael. 2012. Locality in phonology and production planning. *McGill Working Papers in Linguistics* 22:1–18.
- Wauquier-Gravelines, Sophie. 1996. Organisation phonologique et traitement de la parole continue. Doctoral Dissertation, Université Paris 7.
- Whang, James. 2018. Recoverability-driven coarticulation: Acoustic evidence from Japanese high vowel devoicing. *Journal of the Acoustical Society of America* 143:1159–1172.
- Wright, Richard. 2004. A review of perceptual cues and cue robustness. In *Phonetically based phonology*, ed. Bruce Hayes, Robert Kirchner, & Donca Steriade, 34–57. Cambridge: Cambridge University Press.
- Zuraw, Kie & Bruce Hayes. 2017. Intersecting constraint families: an argument for Harmonic Grammar. *Language* 93:497–548.