Paradigm uniformity effects on French liaison*

Benjamin Storme

December 16, 2022

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 they pattern ambiguously between word-final and word-initial consonants. This ambiguous behavior has led to proposals to expand the French phoneme inventory or lexicon with new representations (e.g. floating consonants, lexical constructions, gradient symbolic representations). This paper proposes an alternative analysis: the ambiguous patterning 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 (i.e. words as pronounced in isolation). The proposal is shown to account for the ambiguity of liaison both in terms of prosodic attachment and phonetic 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. The ambiguity of liaison is modeled using a probabilistic constraint-based grammar including paradigm uniformity constraints.

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

*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: Arto Anttila and 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] (with a liaison consonant [t]) before vowel-initial words, as shown in (1c).

(1) French liaison

		Word1	Word2		Context
a.	*grand*	[gʁɑ̃]	(none)	'great'	(citation form)
b.	*grand monsieur*	[gʁɑ̃]	[məsjø]	'great man'	(before C-initial words)
c.	*grand ami*	[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 pattern ambiguously between word-final and word-initial consonants, both prosodically and phonetically. For instance, in the presence of a prosodic break between a liaison word (e.g. *grand* 'great') and the following word, liaison consonants can be attached both at the end of the liaison allomorph, like word-final consonants (=*liaison non-enchaînée*), and at the beginning of the following word, like word-initial consonants (=*liaison enchaînée*; Encrevé 1988; Durand & Lyche 2008). This ambiguous behavior has led some researchers to propose specific 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 an alternative account: the ambiguity of liaison consonants is not captured through inherently ambiguous phonological representations or enriched lexical entries. Instead it emerges from the structure of the lexicon as a paradigm uniformity effect, similarly to what has been proposed to account for incomplete devoicing in German (Roettger et al. 2014). More specifically, this paper builds on a hypothesis put forth by Steriade (1999) and according to which the liaison allomorph of a word (e.g. [gʁɑ̃t] in (1c)) is attracted to the pronunciation of the corresponding citation form, i.e. the word as pronounced in isolation (e.g. [gʁɑ̃] in (1a)). Crucially, the liaison consonant is typically absent from the citation form. In a Word1-Word2 sequence, uniformity with the citation form of Word1 will push the liaison consonant away from the end of this word, therefore favoring a word-initial behavior.

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*. Uniformity with the citation form of Word2 will push the liaison consonant away from the beginning of Word2, therefore favoring a word-final behavior. These two opposite uniformity effects, represented with arrows going in opposite directions in (2), are proposed to underly the ambiguous realization of liaison consonants. The forms enclosed in boxes in (2) 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 just like word-initial consonants but also share properties with word-final consonants, both prosodically and phonetically.

(2) Ambiguity of liaison consonants as a paradigm uniformity effect

grand ami [$\underbrace{\boxed{\text{gʁɑ̃}}}_{\text{citation form}}$ → t ← $\underbrace{\boxed{\text{ami}}}_{\text{citation form}}$] 'great friend'

$\underbrace{\qquad\qquad\qquad\qquad}_{\text{liaison allomorph}}$

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

Section 5 shows how the analysis can be extended to account for the ambiguous phonetic realization of French liaison. This analysis builds on Steriade (2000)'s proposal that phonetic detail (e.g. 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 coarticulatory effect that is proposed to underly the phonetic ambiguity of liaison /t/ before /i/ in Quebec French.

Section 7 compares the paradigm uniformity analysis with previous analyses (floating consonants, lexical constructions, and gradient symbolic representations) and highlights some of the strengths and weaknesses of these different approaches. This section also sketches how a more comprehensive model of liaison might be achieved 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 prevocalically (see Section 7 for more discussion on this topic).

This paper yet focuses on another important research question in the literature on French liaison: how liaison consonants are realized when present. As will be further reviewed in this section, liaison consonants pattern in a very puzzling way, as they share prosodic and phonetic properties with both stable word-final and 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* [tʁɑ̃t] 'thirty' are stable because they are present regardless of the nature of adjacents segments in neighboring words. Section 2.1 focuses on the evidence for prosodic ambiguity and Section 2.2 on the evidence for phonetic ambiguity.

2.1 Prosodic ambiguity

In connected speech, stable consonants and liaison consonants generally have the same prosodic behavior when followed by a vowel: they tend to be syllabified with that vowel (Spinelli et al. 2002; Gaskell et al. 2002; Durand & Lyche 2008: section 3.3). This happens whether the vowel is in the same word or in another word. The vowel is in the same word when the consonant is a stable word-initial consonant (e.g. *grand tamis* [gʁɑ̃#tami] 'big sieve'). The vowel is in a different word when the consonant is a stable word-final consonant (e.g. *trente amis* [tʁɑ̃t#ami] 'thirty friends') or a liaison consonant (e.g. *grand ami* [gʁɑ̃t#ami] 'great friend'). 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).

In a corpus study, Fougeron & Delais-Roussarie (2004) found that resyllabification of stable word-final consonants actually happens in 60% of cases. In the 40% remaining cases, the consonant and the vowel are not coarticulated together and there is a prosodic break (a pause, a glottal stop, or a hesitation) that intervenes between the two segments. It's in this context that the behavior of liaison consonants and stable consonants actually comes apart. Although lexically dependent on the identity of Word1,[1] the liaison consonant will typically behave like a word-initial consonant prosodically and attach to Word2.

This can be illustrated with the behavior of 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 contour intensity (De Cat 2007: 34-43). When

[1]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). The liaison consonant is /t/ in the sequence *grand ami* 'great friend' but it becomes /n/ if *grand* is replaced by indefinite determiner *un* in the sequence *un ami* 'a friend'.

Word2 is right dislocated, the liaison consonant has a very puzzling behavior: it is separated from the word it is lexically affiliated to (Word1) and is attached at the beginning of the following word (Word2). In this case, the liaison consonant is described as *enchaînée*. For instance, in (3a), liaison [t] is separated from its lexical host [gʁɑ̃] by a prosodic boundary and attaches to [elefɑ̃]. The prosodic boundary is materialized by a whitespace in the phonetic transcription. By contrast, stable consonants remain prosodically attached to their lexical hosts, as shown in (3b) for word-final consonants (they attach to Word1) and in (3c) for word-initial consonants (they attach to Word2).

(3) Liaison vs. stable consonants in right dislocations

a.	Liaison consonant	*J'en ai un grand, éléphant.*	[gʁɑ̃ t̲elefɑ̃]
		'I have a big one, elephant.'	
b.	Word-final consonant	*J'en ai trente, éléphants.*	[tʁɑ̃**t** elefɑ̃]
		'I have thirty of them, elephants.'	
c.	Word-initial consonant	*J'en ai un beau, tabouret.*	[bo **t**abuʁɛ]
		'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 this case, the liaison consonant is said to be *non-enchaînée*. The availability of *liaison non-enchaînée* has been 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, it mostly behaves as a word-initial consonant at the beginning of Word2 (= *liaison enchaînée*) but it can also behave as a final consonant at the end of Word1 (= *liaison non-enchaînée*).

		Count	Frequency
Liaison present	enchaînée	2488	43%
	non-enchaînée	321	6%
Liaison absent		2978	51%

Table 1: 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 (4) (cited from Durand & Lyche 2008: 51).

(4) *Liaison non-enchaînée* in a speech by Jacques Chirac (November 14, 2005)
font honneur [fɔ̃t̲ ɔnœʁ] 'do.PRES.3PL honor'
(the whitespace between Word1 and Word2 represents a prosodic break)

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 contexts of prosodic breaks involving hesitations, as shown in (5a), or repetitions, as shown in (5b). 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.' Indeed, 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).

(5) Examples of *liaison non-enchaînée* in conversational speech (Durand & Lyche 2008: 50-51)

a.	*vingt euh*	[vɛ̃t̲ œ]	(hesitation)
b.	*un, un Aveyronnais*	[œ̃n̲ œ̃naverone]	(repetition)

2.2 Phonetic ambiguity

Even in case of *enchaînement*, the distinction between liaison consonants, stable word-final consonants and stable word-initial consonants is not completely neutralized. There remain phonetic cues that distinguish the three types of consonants, as will be reviewed in this section. In particular, liaison consonants have a phonetic 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] in *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; Spinelli et al. 2002, 2003).

More targeted studies found that the phonetic realization of the three types of consonants (liaison, word-final, and word-initial) might differ depending on the nature of the consonant, as summarized in Table 2. For /t/, the duration is longer for word-initial consonants than for both word-final and liaison consonants, but without clear durational difference between the latter two types. For /z/, a study by Nguyen et al. (2007) found a shorter duration for word-final consonants. 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 ambiguously 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)	$=^{a}$	dur(C)	$<^{a,d}$	dur(C)
/z/	dur(C)	$<^{b}/=^{a}$	dur(C)	$=^{a,b}$	dur(C)
/n/	dur(C)	$=^{b}$	dur(C)	$=^{b,c}$	dur(C)

[a]Bagou et al. 2009, [b]Nguyen et al. 2007: 12

[c]Wauquier-Gravelines 1996, [d]Dejean de la Bâtie & Bradley 1995: footnote 2

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

The clearest evidence for a phonetic ambiguity 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 differently liaison consonants, stable word-final consonants, and stable word-initial consonants, as shown in Table 3. More specifically, liaison /t/ has 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/. Rates of affrication were obtained by Côté (2014) on a perceptive basis, using data from the PFC project (Côté 2016).

	Consonant is absent	Consonant is present	
Consonant type		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/.[2] This means that the differences in affrication observed in Table 3 cannot be explained away as a by-product of prosodic ambiguity: 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 ambiguously between stable word-final and word-initial consonants. Indeed, in the presence of a prosodic break between Word1 and Word2, liaison consonants may attach at the end of Word1 (their lexical host), as expected for word-final consonants, but they may

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

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 coarticulated together), the phonetic realization of liaison consonants has also been found to be intermediate between the realization of stable word-final and word-initial consonants.

3 Prosodic ambiguity of French liaison as a paradigm uniformity effect

This section proposes an analysis of the ambiguous behavior of French liaison as a paradigm uniformity effect, focusing first on the attachment ambiguity across a prosodic break. More specifically, this ambiguity is proposed to result from a pressure to make contextual variants of Word1 and Word2 similar to the corresponding citation forms in Word1-Word2 sequences. The relevance of the citation form of Word1 for French liaison was originally put forth 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 host (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 ambiguity 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 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 first key ingredient in the analysis is the assumption that some underlying representations may consist of several listed allomorphs, as in Mascaró (2007). Liaison words are assumed to come with two underlyingly listed allomorphs: an allomorph without liaison and an allomorph with liaison (e.g. Gaatone 1978; Steriade 1999). By contrast, non-liaison words have a single underlying representation, as shown in Table 4. This underlying difference between liaison and non-liaison words will turn out to be crucial to account for the asymmetry between liaison and stable consonants. The key insight on which the analysis will build is that the liaison consonant is present in the underlying representation of a word with a liaison allomorph but (generally) absent from the corresponding citation form, as shown in Table 4 for *grand* 'great'. By contrast, stable word-final and word-initial consonants are

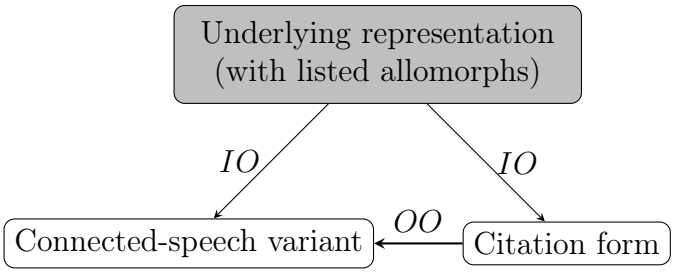


Figure 1: Grammatical model assumed in the analysis (IO = input-output correspondence, OO = output-output correspondence)

		Underlying representation	Citation form
Liaison words	*grand* 'great'	/gʁɑ̃, gʁɑ̃t/	[gʁɑ̃]
Other words	*trente* 'thirty'	/tʁɑ̃t/	[tʁɑ̃t]
	tableau 'table'	/tablo/	[tablo]

Table 4: Assumptions about underlying representations

systematically present in both representations, as shown in Table 4 for *trente* 'thirty' and *tableau* 'table'.

The evaluation of words with several underlyingly listed allomorphs (such as liaison words) 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 used as inputs in the evaluation, without any lexical ordering among them.[3] The allomorph that best satisfies the phonotactic constraints of the language is selected. In 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 patterns). The non-liaison allomorph (e.g. /gʁɑ̃/) is preferred in general because it does not have a final consonant but the liaison allomorph (e.g. /gʁɑ̃t/) 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 under stress (e.g. *jette* [ˈʒɛt] 'throw.3SG') and the [ə]-allomorph in unstressed syllables (e.g. *jetez* [ʒəˈte] 'throw.2PL'; Storme 2020a).

The second important 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

[3]There are also cases of allomorph selection discussed by Mascaro 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.

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. The requirement for connected-speech variants to be similar to the corresponding citation forms will be key to explain two properties of 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.

Evidence for the role of citation forms in paradigm uniformity comes from data 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. For instance, in Lithuanian, underlying /daug/ is realized as [dauk] with devoicing before a vowel-initial word, despite the fact that the following vowel would have permitted to provide sufficient cues to the voicing contrast. Steriade (1997) proposes that devoicing overapplies in this case to increase similarity with the citation form. Although French does not have word-final devoicing, other types of processes provide similar evidence for paradigm uniformity with citation forms in this language. For instance, closed-syllable laxing in French (e.g. *fête* [fɛt] 'party') is motivated utterance-finally by a lack of sufficient consonant-place cues for the final consonant and is extended in utterance-medial contexts regardless of the nature of the following segment. For instance, it applies before a vowel-initial word, as in *la fête* [fɛt] *est finie* 'the party is over', even though this vowel should provide sufficient cues to place of articulation (Storme 2017, 2019).

The evaluation of output-ouput 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 allomorph without liaison that is used utterance-finally (e.g. [gʁɑ̃]) will be used in the evaluation of connected-speech variants, including in liaison contexts (i.e. before vowel-initial words).

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. [gʁɑ̃] 'great') can be attributed to the effect of a markedness constraint penalizing utterance-final consonants (*C_U]). *C_U] 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. [gʁɑ̃t] 'great'), as shown in Table 5a. In this table, indices are used to indicate correspondence relations between input allomorphs and output allomorphs in the candidate set, following Mascaró (2007: 721). Row (a) shows the (faithful) mapping from input /gʁɑ̃/ to output [gʁɑ̃] and row (b) the (faithful) mapping from input /gʁɑ̃t/ to output [gʁɑ̃t].[4] 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.[5]

UR: /gʁɑ̃$_1$, gʁɑ̃t$_2$/	Max-IO $w = 50$	*C_U] $w = 5.51$	H	Predicted	Observed
a. [gʁɑ̃$_1$]			0.00	1.00	1.00
b. [gʁɑ̃t$_2$]		1	5.51	0.00	0.00

(a) Liaison word (*grand* 'great')

UR: /tʁɑ̃t/	Max-IO $w = 50$	*C_U] $w = 5.51$	H	Predicted	Observed
a. [tʁɑ̃t]		1	5.51	1.00	1.00
b. [tʁɑ̃]	1		50.00	0.00	0.00

(b) Consonant-final word (*trente* 'thirty')

UR: /tablo/	Max-IO $w = 50$	*C_U] $w = 5.51$	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 5: 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_U]. The specific weight shown in Table 5a for this constraint ($w = 5.51$) was inferred jointly with the weights of all the

[4] As in Mascaró (2007: 721), unfaithful mappings (e.g. /gʁɑ̃/ → [gʁɑ̃t] and /gʁɑ̃t/ → [gʁɑ̃]) are ignored here because they incur strictly more constraint violations than the corresponding faithful mappings. For instance, /gʁɑ̃/ → [gʁɑ̃t] violates the same markedness constraint *C_U] as /gʁɑ̃t/ → [gʁɑ̃t] but violates in addition a faithfulness constraint that bans epenthesis (Dep-IO).

[5] 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.

other constraints presented in Section 3, using OT-Soft (Hayes et al. 2013).[6]

In order to block consonant deletion for words with final consonants but no vowel-final listed allomorph (e.g. *trente* [tʁɑ̃t] 'thirty'), the constraint that penalizes consonant deletion (MAX-IO) must take precedence over the constraint penalizing utterance-final consonants (*C_U]), as shown in Table 5b. 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 5a. 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 5b 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_U] is of course the same as in Table 5a because the same single grammar must derive all attested forms. Table 5c 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 5 derives the correct citation forms for liaison words and non-liaison words, as listed in Table 4.

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 break between the two words and where Word2 starts with a vowel (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 even in the presence of a prosodic break between the two vowels. 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 6a shows the analysis of liaison words in this context. 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.

[6]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 5a 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 to be ignored. An even higher weight on the phonotactic constraint would be required to make it even less likely.

URs: /gʁɑ̃$_1$, gʁɑ̃t$_2$/+/ami/ CFs: [gʁɑ̃]+[ami]	Dep-IO 50	*VV 2.63	L-Anch-OO 2.80	R-Anch-OO 4.78	*H*	Predic- ted	Obser- ved
a. [gʁɑ̃$_1$ ami]		1			2.63	0.51	0.51
b. [gʁɑ̃t$_2$ ami]				1	4.78	0.06	0.06
c. [gʁɑ̃ t$_2$ami]			1		2.80	0.43	0.43
d. [gʁɑ̃$_1$l ami]	1			1	54.78	0.00	0.00
e. [gʁɑ̃$_1$ lami]	1		1		52.80	0.00	0.00

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

URs: /tʁɑ̃t/+/ami/ CFs: [tʁɑ̃t]+[ami]	Dep-IO 50	*VV 2.63	L-Anch-OO 2.80	R-Anch-OO 4.78	*H*	Predic- ted	Obser- ved
a. [tʁɑ̃t ami]					0.00	1.00	1.00
b. [tʁɑ̃ 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	*VV 2.63	L-Anch-OO 2.80	R-Anch-OO 4.78	*H*	Predic- ted	Obser- ved
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 6: Connected-speech variants (CF = citation form, UR = underlying representation). The white space 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 either because they are not identical to the citation form [gʁɑ̃], contrary to candidate (a): they feature a consonant that is epenthetic in the output-output dimension. In Table 6a, 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 ambiguity of the liaison consonant in the liaison allomorph. They are defined in the following paragraphs.

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 has 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 6a because this candidate features a consonant at the end of [gʁɑ̃t] that is absent from the corresponding citation form [gʁɑ̃]. This constraint therefore

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.[7] In a Word1-Word2 sequence, these two requirements will conflict and drive the ambiguous prosodic attachment of liaison consonants.

The requirement to protect the left edge of a word is modeled using LEFT-ANCHOR-OO. LEFT-ANCHOR-OO requires that any element at the left edge of a citation form has 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 6a) and therefore favors a prosodic attachment of liaison [t] at the end of Word1 (candidate (b)).[8]

Table 6a shows a concrete choice of constraint weights that can derive the frequencies attested in Encrevé (1988)'s corpus. As already mentioned in Section 3.2, constraint weights were inferred jointly for the analysis of citation forms and connected-speech variants.[9] 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 ambiguity of liaison consonants: both attachments to 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 ambiguity for liaison consonants but not for stable word-final and word-initial consonants. This is shown in Tables 6b and 6c, respectively.

In Table 6b, 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 ambiguous between Word1 and Word2 (Table 6a) predicts a categorical attachment to Word1 for stable word-final consonants (Table 6b).[10]

In Table 6c, 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 ambiguous between Word1 and Word2 (Table 6a) predicts a categorical attachment to Word2

[7] A similar idea was proposed by Zuraw & Hayes (2017) to account for the behavior of *h-aspiré* words in French and will be further discussed at the end of this section.

[8] 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.

[9] Note that they could also be inferred independently since there happens to be no overlap between the constraints used in the two analyses.

[10] 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 to be ignored for the purpose of this paper.

for stable word-initial consonants (Table 6c).

As mentioned in Footnote 7, the hypothesis that connected-speech variants are required to be similar to the corresponding citation forms along their left edge was already proposed by Zuraw & Hayes (2017) to account for the behavior of French *h-aspiré* words. *H-aspiré* words are a subset of vowel-initial words that tend to block external sandhi phenomena (elision, liaison). For instance, the *h-aspiré* word *le hibou* 'the owl' is typically realized as [lə#ibu], with blocking of schwa-elision in the determiner. For other vowel-initial words, schwa-elision usually applies, as in *l'hiver* [l#ivɛʁ] 'the winter'. Zuraw & Hayes (2017) also use uniformity with citation forms to account for this difference: they assume that *h-aspiré* words are subject to a greater pressure to be uniform with their citation forms than other vowel-initial words. The present paper shows that the same kind of paradigm uniformity pressure can also account for the prosodic ambiguity of French liaison.

However the present analysis differs from Zuraw & Hayes (2017) in the implementation details. The paradigm uniformity constraints in Zuraw & Hayes (2017) are formulated as alignment constraints (they penalize misalignments between morpheme and syllable boundaries) and therefore depend on assumptions about how segments are syllabified. By contrast, the paradigm uniformity constraints proposed in this section are formulated as ANCHOR constraints and do not refer to syllabification. These constraints only require segments to have correspondents in connected-speech variants and citation forms. As shown by the analysis proposed in this section, syllabification is not necessary to capture the prosodic ambiguity of liaison consonants across a prosodic break. Moreover, as previewed in the background section on the phonetic ambiguity of liaison (Section 2.2), syllabification is not sufficient either. Indeed liaison consonants and word-initial consonants differ in their phonetic realizations in contexts where they would both be expected to behave as onsets, according to traditional syllabification principles. An analysis using only ALIGN constraints would therefore not work to account for the phonetic ambiguity of liaison (see Section 5 for an alternative proposal).

3.4 Implicational generalizations

The preceding section has shown that the paradigm uniformity analysis can capture the ambiguous prosodic patterning of liaison consonants as attested in Encrevé's corpus. This section further establishes that the asymmetry between liaison and stable consonants in (6) follows as a necessary consequence of the proposed constraint set, and this regardless of the framework for probabilistic grammar: Stochastic Optimality Theory (OT; Boersma & Hayes 2001), Noisy Harmonic Grammar (HG; Hayes 2017), or Maximum Entropy grammars (MaxEnt; Goldwater & Johnson 2003; Hayes & Wilson 2008).

(6) Statistical implicational generalizations derived by the analysis

a. $P(\text{Word1-attachment}|\text{Liaison C}) \leq P(\text{Word1-attachment}|\text{Final C})$

b. $P(\text{Word2-attachment}|\text{Liaison C}) \leq P(\text{Word2-attachment}|\text{Initial C})$

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 6b and candidate (a) in Table 6c, respectively) violate a strict superset of the constraints violated by the candidates that involve no displacement (candidate (a) in Table 6b and candidate (b) in Table 6c, 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 6a, 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 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 comprised 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 (6a) holds because the candidate with a liaison consonant at the end of Word1 fares worse within the candidate set for the liaison word (Table 6a) than the candidate with a stable consonant at the end of Word1 does within the candidate set for the consonant-final word (Table 6b). And generalization (6b) holds because the candidate with a liaison consonant at the beginning of Word2 fares worse within the candidate set for the liaison word (Table 6a) than the candidate with a word-initial consonant at the beginning of Word2 does within the candidate set for the consonant-initial word (Table 6c). The reader is referred to the output of the CoGeTo analysis (Magri & Anttila 2019) in the supplementary files. This 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 ambiguous prosodic patterning of liaison consonants follows from their being *absent* from the corresponding citation form. But no ambiguity 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 (epenthetic liaison and suppletive liaison).

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

		Underlying representation	Citation form
Epenthetic liaison	grand 'great'	/gʁɑ̃, gʁɑ̃t/	[gʁɑ̃]
Suppletive liaison	beau 'beautiful'	/bo, bɛl$_{\text{FEM}}$/	[bo], [bɛl$_{\text{FEM}}$]

Table 7: Epenthetic liaison and suppletive liaison

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. [gʁɑ̃t]$_{\text{liaison allomorph}}$ = [gʁɑ̃]$_{\text{citation form}}$ + [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* [bɛl#ami] '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 7.[11]

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 has been first discussed by Steriade (1999). The paradigm uniformity analysis predicts that only epenthetic liaison should pattern ambiguously between stable word-final and word-initial consonants. Indeed, 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 reinforced in contextual realizations by paradigm uniformity.

Table 8 shows that the same grammar that derived prosodic ambiguity for epenthetic liaison in Table 6a indeed predicts an unambiguously 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*.

[11] 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 (see Section 3.1).

URs: /bo, bɛl$_{\text{FEM}}$/+/ami/ CFs: [bo, bɛl$_{\text{FEM}}$]+[ami]	Dep-IO 50	*VV 2.63	L-Anch-OO 2.80	R-Anch-OO 4.78	*H*	Predic-ted
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 8: Suppletive liaison consonants are predicted to behave unambiguously like word-final consonants and attach to Word1 (*bel ami* 'beautiful friend').

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 (7).

(7) Epenthetic liaison vs. suppletive liaison in right dislocations

a. Epenthetic liaison *J'en ai un grand, éléphant.* [gʁɑ̃ t̲elefɑ̃]
'I have a big one, elephant.'

b. Suppletive liaison *J'en ai un bel, éléphant.* [bɛl̲ elefɑ̃]
'I have a beautiful one, elephant.'

However, right-dislocation contexts are probably not the most appropriate context to make a case for the ambiguous patterning of epenthetic liaison, as they seem to 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).[12] The goal of Study 1 is therefore to compare the behavior of liaison consonants (both 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 9, and each adjective appeared in two Adj-N sequences varying by the strength of their collocation. For instance, *petit* appeared

[12]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 (7), 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.

both in *petit ami* 'boyfriend' (more frequent) and in *petit anneau* 'small ring' (less frequent).[13] 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	*énorme, jeune, large, magnifique, meilleur, superbe*
Adjectives with epenthetic liaison	*faux, grand, gros, mauvais, parfait, petit*
Adjectives with suppletive liaison	*ancien, bel, bon, prochain, vieil, nouvel*
Adjectives followed by consonant-initial nouns	*affreux, charmant, gentil, joli, long, vrai*

Table 9: Adjectives used in Study 1

Two French native speakers (female and 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 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 10 for each of the four experimental conditions. The prosodic attachment of the consonant was clearly indicated to the speakers 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.[14] 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.[15]

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

Table 10: 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

[13]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).

[14]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.

[15]The Praat script is available at http://wwwbio.leidenuniv.nl/~eew/G6/staff/beckers/beckers.html.

produced by the male speaker, and the 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 Δ 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 Δ. Credible Intervals were obtained using the ETI (Equal-tailed Interval) method and the package `bayestestR` (Makowski et al. 2019).

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 the former being less likely to attach to Word1 than the the latter ($\Delta_{\text{sup liaison - ep liaison}} = -4.23$, $CI = [-8.67, -0.04]$). Suppletive liaison was found to behave like stable word-final consonants ($\Delta_{\text{sup liaison - final}} = -0.19$, $CI = [-3.45, 2.66]$), 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 ($\Delta_{\text{ep liaison - final}} = 4.03$, $CI = [0.96, 7.22]$) and word-initial consonants ($\Delta_{\text{ep liaison - initial}} = -11.28$, $CI = [-22.82, -4.41]$).

4.4 Discussion

An important prediction of the paradigm uniformity was corroborated by the results of Study 1: liaison behaves ambiguously if and only if the liaison consonant can be analyzed as

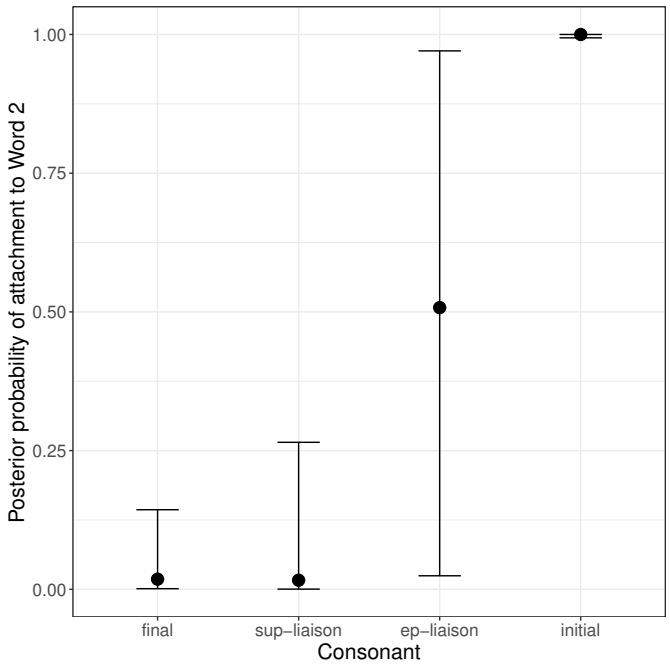


Figure 2: Posterior probability of attachment to Word2 as a function of consonant type (mean and 95% CI)

epenthesized at the end of the citation form. When the liaison allomorph is identical to the feminine form and cannot be analyzed as a substring of the masculine form (= suppletive liaison), no ambiguity arises. This study is to the author's knowledge the first controlled study that establishes this difference between epenthetic liaison and suppletive liaison.

5 Phonetic ambiguity 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 phonetic ambiguity 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 case study provides the clearest case of ambiguous realization for French liaison.[16] Also, this case study can be modeled with the type of symbolic 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 symbolic to phonetic grammars (Flemming 2001). This task is left for further 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 on C (e.g. affrication in the case of Quebec French)

[16]The analysis presented in this section supersedes the analysis presented in Storme (2020b).

correlates with a change on the following V. In combination with paradigm uniformity, this correlation will be crucial to explain why liaison consonants might pattern ambiguously 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 coarticulated 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 phonetic ambiguity 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 (see Flemming 2001: 16-23). A large number of studies have shown that C assimilates to V in CV, in particular F2 at consonant 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 in 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 [tʃi] sequences in Japanese. This result suggests that the affricate gets more affricated (the frication noise gets longer) as the following high vowel devoices/reduces in Japanese, 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 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 phonetic ambiguity as in the analysis of prosodic ambiguity (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 undrlying /ti/ as [tsi̥] (candidate (c) in Table 11a) can be attributed to the effect of the two markedness constraints *ti and *tsi 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 11a 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.[17]

[17]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.

UR: /tiʁɑ̃/	*ti 15.81	*tsi 50	IDENT-IO(CONT) 6.16	IDENT-IO(VOI) 6.16	*C_U] 7.82	*H*	Predicted	Observed
a. [tiʁɑ̃]	1					15.81	0.03	0.01
b. [tsiʁɑ̃]		1	1			56.16	0	0
c. [ts̥iʁɑ̃]			1	1		12.33	0.97	0.99

(a) Word with initial /ti/ (*tyran* 'tyrant')

UR: /tʁɑ̃t/	*ti 15.81	*tsi 50	IDENT-IO(CONT) 6.16	IDENT-IO(VOI) 6.16	*C_U] 7.82	*H*	Predicted	Observed
a. [tʁɑ̃t]					1	7.82	1.00	1.00
b. [tʁɑ̃ts]			1		1	13.99	0	0

(b) Word with final /t/ (*trente* 'thirty')

UR: /inosɑ̃/	*ti 15.81	*tsi 50	IDENT-IO(CONT) 6.16	IDENT-IO(VOI) 6.16	*C_U] 7.82	*H*	Predicted	Observed
a. [inosɑ̃]						0	1.00	1.00
b. [i̥nosɑ̃]				1		6.16	0	0

(c) Word with initial /i/ (*innocents* 'innocent')

UR: /gʁɑ̃$_1$, gʁɑ̃t$_2$/	*ti 15.81	*tsi 50	IDENT-IO(CONT) 6.16	IDENT-IO(VOI) 6.16	*C_U] 7.82	*H*	Predicted	Observed
a. [gʁɑ̃$_1$]						0	1.00	1.00
b. [gʁɑ̃t$_2$]					1	7.82	0	0

(d) Word with liaison /t/ (*grand* 'great')

Table 11: Citation forms

Tables 11b and 11c 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. Indeed, 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.

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 11d (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, as shown in Table 12. The forms enclosed in boxes in

Table 12 correspond to the citation forms of the two words. The liaison consonant differs from stable word-final/word-initial consonants in being absent from the corresponding citation form.

		Citation forms	Candidate with affrication	Feature changes	Rate of affrication (Côté 2014)
(a)	Word-final C	[tʁɑ̃**t**], [**i**nosɑ̃]	[tʁɑ̃**ts̥i**nosɑ̃]	2	36.5%
(b)	Liaison C	[gʁɑ̃], [**i**nosɑ̃]	[gʁɑ̃**ts̥i**nosɑ̃]	1	66.0%
(c)	Word-initial C	[vʁɛ], [**ts̥i**ʁɑ̃]	[vʁɛ**ts̥i**ʁɑ̃]	0	99.2%

Table 12: 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. Indeed, 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 two columns of Table 12, 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.

In the constraint-based analysis, dissimilarities with 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 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.

Note that the paradigm uniformity constraints used in the analysis of prosodic ambiguity (RIGHT-ANCHOR-OO and LEFT-ANCHOR-OO) will not play a role here. The reason is that the presence of a liaison consonant between Word1 and Word2 in connected speech is assumed to be penalized by both RIGHT-ANCHOR-OO and LEFT-ANCHOR-OO when there is no pause between the two words. Indeed the liaison consonant is in contact with both word edges in connected speech (the end of Word1 and the beginning of Word2) and, phonetically, there is coarticulation with both the preceding vowel (presence of VC transitions) and the following vowel (presence of CV transitions), whether the liaison consonant 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 13 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 13a), stable final consonants (Table 13b), and stable initial consonants (Table 13c).[18] As in the analysis of prosodic ambiguity, 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 13.

URs: /gʁɑ̃$_1$, gʁɑ̃t$_2$/ +/inosɑ̃/ CFs: [gʁɑ̃]+[inosɑ̃]	*VV 16.43	*ti 15.81	*tsi 50	ID-OO (cont) 1.21	ID-OO (voi) 2.81	ID-IO (cont) 6.16	ID-IO (voi) 6.16	*H*	Predicted	Observed
a. [gʁɑ̃$_1$inosɑ̃]	1							16.43	0.15	0.15
b. [gʁɑ̃t$_2$inosɑ̃]		1						15.81	0.29	0.29
c. [gʁɑ̃ts$_2$inosɑ̃]			1			1		56.16	0	0
d. [gʁɑ̃ts$_2$͡inosɑ̃]					1	1	1	15.14	0.56	0.56

(a) Word1 with liaison /t/ + Word2 with initial /t/ (*grand innocent* 'great innocent')

URs: /tʁɑ̃t/ +/inosɑ̃/ CFs: [tʁɑ̃t]+[inosɑ̃]	*VV 16.43	*ti 15.81	*tsi 50	ID-OO (cont) 1.21	ID-OO (voi) 2.81	ID-IO (cont) 6.16	ID-IO (voi) 6.16	*H*	Predicted	Observed
a. [tʁɑ̃tinosɑ̃]		1						15.81	0.63	0.64
b. [tʁɑ̃tsinosɑ̃]			1	1		1		57.37	0	0
c. [tʁɑ̃ts͡inosɑ̃]				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/ +/tiʁɑ̃/ CFs: [vʁɛ]+[ts͡iʁɑ̃]	*VV 16.43	*ti 15.81	*tsi 50	ID-OO (cont) 1.21	ID-OO (voi) 2.81	ID-IO (cont) 6.16	ID-IO (voi) 6.16	*H*	Predicted	Observed
a. [vʁɛtiʁɑ̃]		1		1	1			19.83	0	0.01
b. [vʁɛtsiʁɑ̃]			1		1	1		58.97	0	0
c. [vʁɛts͡iʁɑ̃]						1	1	12.33	1.00	0.99

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

Table 13: Connected-speech variants

The analysis derives all three realizations attested in Côté (2014) for liaison words, as shown in Table 13a: absence of liaison consonant (candidate (a)), liaison without affrication

[18] 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 13a, candidate (c) in Table 13b and Table 13b) or not (candidate (c) in Table 13a, candidate (d) in Table 13b and Table 13b) 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.

(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 13b vs. candidate a in Table 13b). 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 13b) but not in the liaison word (candidate (d) in Table 13a), 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 13c). 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 13c) 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 13a): indeed, 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 predicts that the ambiguous status of French liaison in (8) follows as a necessary consequence of the proposed constraint set, and this regardless of the framework for probabilistic grammars (Stochastic OT, Noisy HG, or MaxEnt).

(8) Statistical implicational generalizations derived by the analysis

a. P(No affrication|Liaison C) $\leq$ P(No affrication|Final C)

b. P(Affrication|Liaison C) $\leq$ P(Affrication|Initial C)

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

The demonstration is more involved than in the case of prosodic ambiguity (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 (8a) holds because the candidate with an unaffricated liaison consonant (candidate (b) in Table 13a) fares worse within the candidate set for liaison words than the candidate with an unaffricated final consonant (candidate (a) in Table 13b) does within its own candidate set. And, similarly, generalization (8b) holds because the candidate with an affricated liaison consonant (candidate (d) in Table 13a) fares worse within its candidate set than the candidate with an affricated initial consonant (candidate (c) Table 13c) 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 (8) 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 /t/ correlates with a reduction of /i/ in /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 two quasi-minimal pairs 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.[19] Annotations were done manually in Praat (Boersma & Weenink 2021). /t/ duration was used as 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 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.

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).

Two statistical analyses were conducted. A Bayesian logistic regression was fit to the

[19] I am grateful to Marie-Hélène for making the data available to me.

data using the brms package in R, with Vowel (present, absent) as as 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 3. 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 4. 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 phonetic ambiguity 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/. Indeed, 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 paradigm uniformity. By contrast, word-initial /ti/ sequences already undergo affrication in 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

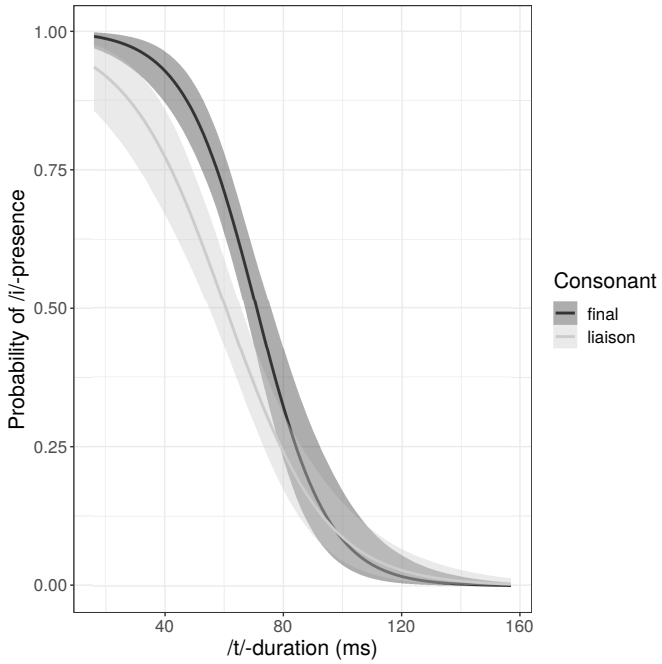


Figure 3: Posterior probability of /i/-deletion as a function of /t/-duration and consonant type (liaison, final)

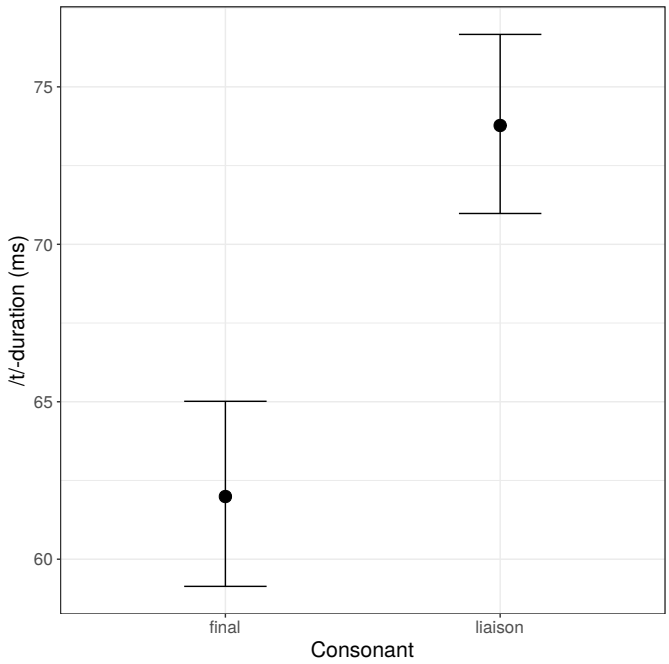


Figure 4: Posterior distribution of consonant duration (ms) for stable word-final /t/ and liaison /t/ (mean and 95% CI)

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 present (=Q1). The other main research question is about which factors affect the presence of liaison consonants (=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 14.

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

Table 14: 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 15 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 mainly on representations (a richer phoneme inventory and/or a richer lexicon) or on computation (constraint interaction), as indicated in columns 2-4 of Table 15.

	Research question	Enriched phoneme inventory	Enriched lexicon	Constraint interaction
Floating consonants[a]	Q1	yes	no	no
Lexical constructions[b]	Q1, Q2	no	yes	no
Gradient symbolic representations[c]	Q1, Q2	yes	yes	no
Paradigm uniformity[d]	Q1	no	no	yes
Production planning[e]	Q2	no	no	yes

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

Table 15: 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?

In this classification, the paradigm uniformity analysis belongs to the latter category because it treats the ambiguity of French liaison as emerging from the interaction of conflicting similarity requirements (faithfulness to the citation form of Word1 vs. faithfulness to the citation form of Word2), while assuming traditional underlying phonological and lexical representations for liaison consonants and liaison words. The classification of the other analyses will be motivated further below.

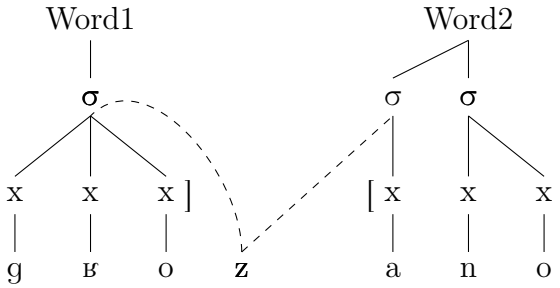
Sections 7.1-7.3 review the first three analyses listed in Table 15 and highlight how they compare to the paradigm uniformity analysis in accounting for Q1. Section 7.4 concludes by sketching how insights from Kilbourn-Ceron's production planning hypothesis can be combined with the paradigm uniformity analysis to get a more comprehensive model of French liaison accounting for both Q1 and Q2.

7.1 The approach using 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 (9). 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).

(9) 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 ambiguous realization of liaison consonants before vowel-initial words. However they differ from the paradigm uniformity analysis because they build the ambiguous behavior of liaison consonants directly into their phonological representations. These approaches indeed require to enrich the phoneme inventory of French 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 ambiguities (prosodic and phonetic). By contrast, the approach using floating segments currently only accounts for prosodic ambiguity. Indeed, 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 phonetically in this context, as documented in Section 2.2. 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: one projects a skeletal slot whereas the other does not, as shown in (9). For Quebec French, a rule of affrication that is sensitive to this underlying difference could then be conceived. 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 The approach using 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 /gʁɑ̃$_{\text{Adj}}$ t $X_{\text{V-initial N}}$/ 'great N', where $X_{\text{V-initial}}$ 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 14 (e.g. the rate of liaison is higher for liaison words with higher lexical frequency), Bybee mentioned in passing that they can also account for the prosodic ambiguity 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 it possible to say *un élé phant* [ɛ̃n#ele 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 limit as the approach using floating consonants in that it does not account for phonetic ambiguity. 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* [pɔk œ tavjɔ̃]). Bybee (2001) sketches an explanation for why it 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 actually clearly refers to principles of paradigm uniformity among morphologically related forms.

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 comes from the fact that multiple-word constructions are stored in the lexical-construction analysis whereas they are not in the paradigm uniformity analysis. Altough 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 4 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.

7.3 The approach using gradient symbolic representations

In the approach using gradient symbolic representations, liaison consonants are characterized by an activity degree that is smaller than that of stable consonants (Smolensky & Goldrick 2016; Smolensky et al. 2020). This lower activity degree is the mechanism that explains why liaison consonants do not always surface. Moreover, 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*).

This analysis is represented in (10), where the degree of activity of liaison consonants is indicated as a subscript (for stable consonants, the degree of activity is always equal to 1). A word like *ami* [ami] 'friend' is stored with all the possible liaison consonants that can be attached to it as first segment. When the words *petit* and *ami* are combined, the activity level of liaison /t/ increases, allowing it to surface. Because the /t/ is underlyingly present in both words, it can surface either at the end of Word1 or at the beginning of Word2. The presence of word-specific activity degrees makes it possible to model some of the lexical effects on liaison presence reported in Table 14.

(10) Liaison consonants as gradient phonemes affiliated to both Word1 and Word2 (Smolensky & Goldrick 2016)
/pətit$_{\underline{0.48}}$/ + /{t$_{\underline{0.09}}$, z$_{\underline{0.09}}$, n$_{\underline{0.09}}$}ami/

Like the two approaches previously reviewed, the approach using gradient symbolic representations does not account for the phonetic ambiguity of French liaison. Indeed, the activity degree of phonemes determines how likely they are to surface but does not determine their segmental or phonetic realization. Another mechanism is required in addition to derive the phonetic ambiguity of French liaison.

However the approach using gradient symbolic representations has the same advantage over the paradigm uniformity analysis as the lexical-construction approach in being able to account for lexical effects on the rate of liaison.

7.4 Towards a comprehensive model of French liaison

The paradigm uniformity analysis provides a more comprehensive account of the *realization* of liaison consonants than alternative analyses. It can indeed derive both prosodic ambiguity and phonetic ambiguity. Yet it does not account for the factors that affect the *rate* of liaison, 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 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 uniformity analysis by indexing output-output faithfulness constraints to properties that are relevant for production planning (see Pater 2007 on constraint indexation). Indeed, 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 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 weighing output-output faithfulness constraints more 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) in French to account for the greater resistance of *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 16 illustrates this approach using a toy example based on the analysis of prosodic ambiguity 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 16 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-$\text{OO}_{\text{low predictability}}$). Furthermore, this constraint has a higher weight than the constraint that

URs: /gʁɑ̃$_1$, gʁɑ̃t$_2$/+/ano/ Bases: [gʁɑ̃]+[ano]	*H*	Predicted
a. [gʁɑ̃$_1$ ano]	2.63	0.73
b. [gʁɑ̃t$_2$ ano]	4.78	0.08
c. [gʁɑ̃ t$_2$ano]	4.00	0.19
d. [gʁɑ̃$_1$l ano]	54.78	0.00
e. [gʁɑ̃$_1$ lano]	54.00	0.00

(a) *grand anneau* 'large ring'
$w_{\text{Left-Anchor-OO}_{\text{high predictability}}} = 4.00$

URs: /gʁɑ̃$_1$, gʁɑ̃t$_2$/+/ami/ Bases: [gʁɑ̃]+[ami]	*H*	Predicted
a. [gʁɑ̃$_1$ ami]	2.63	0.51
b. [gʁɑ̃t$_2$ ami]	4.78	0.06
c. [gʁɑ̃ t$_2$ami]	2.80	0.43
d. [gʁɑ̃$_1$l ami]	54.78	0.00
e. [gʁɑ̃$_1$ lami]	52.80	0.00

(b) *grand ami* 'great friend'
$w_{\text{Left-Anchor-OO}_{\text{high predictability}}} = 2.80$

Table 16: How an analysis including indexed output-output faithfulness constraints accounts jointly for the rate and realization of liaison

evaluates the pronunciation of words with high contextual predictability such as *ami* (Left-Anchor-OO$_{\text{high predictability}}$). Table 16 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*)[20] while still accounting for the prosodic ambiguity 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 16 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 push the liaison consonant away from it). And the same is expected to hold for the *phonetic* 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

Liaison consonants have been shown in previous research to pattern ambiguously 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 phonological underlying status of liaison consonants. Rather it can be derived 'for free' from the observation that liaison words come under two variants (with and without liaison) and from independently motivated principles of uniformity among paradigmatically related forms (contextual variants of a word and the corresponding citation form). Also, the ambiguous behavior of liaison consonants can be derived without positing lexical constructions or massive allomorphy in the lexicon. It is sufficient to assume that only liaison words have two listed allomorphs. An explicit implementation of the analysis in a probabilistic constraint-based grammar was proposed and shown to be able to derive both prosodic ambiguity and phonetic ambiguity of French

[20]These numbers are only indicative and do not represent the actual rates of liaison for these specific word sequences.

liaison. 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. 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 not ambiguous in themselves but 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 phonetic ambiguity 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 in the future. First, it should be extended to account for patterns of phonetic ambiguity 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 ambiguous 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 further test 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.

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. Chicester, 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.

Delattre, Pierre C. 1947. La liaison en français, tendances et classification. *The French Review* 21:148–157.

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.

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 Jeager. 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.

Meinschaefer, 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. 2017. The *loi de position* and the acoustics of French mid vowels. *Glossa: a journal of general linguistics* 2.

Storme, Benjamin. 2019. Contrast enhancement as motivation for closed syllable laxing and open syllable tensing. *Phonology* 36:303–340.

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.7445766.

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.