French liaison is allomorphy, not allophony: Evidence from lexical statistics*

Benjamin Storme

August 26, 2024

To appear in *Morphology*

Abstract

The exact nature of French liaison as a phonological or morphological alternation is still debated. Under the phonological analysis, liaison is allophony: liaison consonants are special phonemes that alternate between a consonant allophone and zero (e.g., [t] ∼ ∅), the zero allophone being derived from the consonant phoneme through deletion (/t/ → ∅). Under the morphological analysis, liaison is allomorphy: liaison words have two underlyingly listed allomorphs, a consonant-final allomorph and a shorter allomorph that lacks this consonant (e.g., *grand* 'great' /gʁɑ̃t, gʁɑ̃/). This paper uses evidence from lexical statistics to arbitrate between these two analyses. The form without liaison consonant (and with deletion, under the phonological analysis) has been found in previous research to become less likely with increasing lexical frequency. The paper shows that this is problematic for the phonological analysis of French liaison, as deletion typically applies more frequently in high-frequency words across languages. The paper further shows, using evidence from a large lexical database, that words involved in liaison alternations generally have lower type frequency but higher token frequency than non-liaison words when phonotactic and morphological effects on lexical frequency are controlled for. This result is in line with the predictions of the morphological analysis, as allomorphy typically involves a relatively small number of words that occur frequently. Due to its empirical nature, this argument constitutes to date one of the strongest arguments in favor of the morphological analysis.

Keywords: French liaison; phonology; morphology; frequency

1 Introduction

Understanding the various types of representations (phonological, morphological, syntactic, etc.) that are involved in language processing is a fundamental aspect of analyzing and explaining linguistic phenomena. One of the key objectives is to attribute linguistic phenomena to the appropriate type of representation. However this task can be especially challenging when dealing with patterns that involve interactions between different levels of linguistic analysis. This paper focuses on French

*I would like to thank Olivier Bonami, the editor of Morphology, and two anonymous reviewers for very helpful feedback and suggestions on this paper. Special thanks to Isabelle Dautriche for giving me access to her list of French monomorphemic words. I am also grateful to the participants in the 2023 meeting of the Phonologie du français contemporain project, the members of the EmpiPhon group at the University of Amsterdam, the participants at the Going Romance 2023 conference in Nijmegen and the participants at NELS 2024 for very helpful feedback. All errors are my own.

liaison, a pattern notorious for its sensitivity to multiple linguistic factors (phonological, lexical, syntactic, stylistic), and proposes a test based on lexical statistics to arbitrate between the two main competing analyses of this pattern as a phonological or a morphological alternation.

In French, some words use a special consonant-final form in some contexts. For instance, the adjective *grand* 'great' is generally pronounced as [gʁɑ̃] but, when preceding a vowel-initial noun, it may be pronounced as [gʁɑ̃t] with a final [t], as in *grand ami* [gʁɑ̃t#ami] 'great friend'. This alternation, known as liaison, has been a central topic in linguistics, due to the complex interaction of factors that shape its use (see Côté 2011 and Bonami & Delais-Roussarie 2021 for an overview). Despite decades of study, it still remains an active area of research (e.g., Boersma & van Leussen 2017; Smolensky, Rosen & Goldrick 2020; Storme 2024c) and some basic issues still remain unresolved.

One of the lingering issues about French liaison concerns the synchronic status of liaison alternations: are they phonological or morphological alternations? In the phonological analysis, liaison is an alternation between *allophones* of a *phoneme* (Schane 1968; Clements & Keyser 1983; Dell 1985; Encrevé 1988; Tranel 1990; Smolensky & Goldrick 2016; Smolensky et al. 2020). Liaison words like *grand* feature a special type of phoneme at their right edge that alternates between a consonant allophone and zero, as depicted in (1a).

(1) Liaison: allophony or allomorphy?

	[gʁɑ̃] ∼ [gʁɑ̃t] alternation
a. Liaison as allophony	/$t_{liaison}$/: [∅, t]
b. Liaison as allomorphy	*grand* 'great': /gʁɑ̃, gʁɑ̃t/

In the morphological analysis, liaison is an alternation between *allomorphs* of a *morpheme* (Gaatone 1978; Klausenburger 1984; Steriade 1999; Bonami & Boyé 2005; Bonami 2011; Storme 2024c). Liaison words like *grand* come with two underlyingly listed allomorphs, a short one and a long one with an extra final consonant, as depicted in (1b).

The two analyses differ in attributing liaison alternations to different linguistic levels: the sound inventory for the phonological analysis (liaison consonants are special *phonemes*) and the lexicon for the morphological analysis (liaison words are special *words*). The debate between the two approaches has not been settled yet, as demonstrated by two recent studies by Smolensky & Goldrick (2016) and Storme (2024c), where the phonological and morphological analyses are adopted, respectively.

This paper contributes to this debate by proposing a new test based on evidence from lexical statistics. This test exploits the fact that allophony and allomorphy have different lexical signatures. Allophonic processes that involve reduction, lenition or deletion, such as French liaison under the phonological analysis, typically apply more frequently in high-frequency words than in low-frequency words (Jurafsky, Bell, Gregory & Raymond 2001; Bell, Brenier, Gregory, Girand & Jurafsky 2009; Bybee 2007; Coetzee & Kawahara 2013). The deletion analysis of liaison in (1a) therefore predicts a negative correlation between the lexical frequency of a liaison word and its rate of liaison: liaison forms (i.e. forms without deletion) should become less likely relative to non-liaison forms (i.e. forms with deletion) as the lexical frequency of the corresponding liaison word increases.

On the other hand, allomorphy is typically limited to a small number of irregular words with high token frequency, such as irregular verbs (Bybee 1985:Chapter 5; Berg 2011). If liaison is allomorphy, as hypothesized in the morphological analysis in (1b), then liaison words should be fewer but occur more frequently than words that do not feature this kind of allomorphy.

The predictions of the phonological and morphological analyses are summarized in (2).

(2) Predictions about the lexical distribution of French liaison

a. Phonological analysis: liaison as allophony
Liaison forms (i.e. without deletion) become less likely relative to non-liaison forms (i.e. with deletion) as the lexical frequency of liaison words increases.

b. Morphological analysis: liaison as allomorphy
Liaison words have lower type frequency but higher token frequency than non-liaison words.

The goal of this paper is to evaluate these predictions. There is evidence available against prediction (2a) of the phonological analysis, with liaison forms becoming *more* likely with increasing lexical frequency (Fougeron, Goldman, Dart, Guélat & Jeager 2001; Kilbourn-Ceron 2017). However, this evidence has not yet been brought to bear on the debate about the underlying representation of liaison.

As for prediction (2b), it has not been tested thoroughly yet. Claims that liaison is limited to closed lexical classes and specific morphological categories (Côté 2011:Section 2.2.1) have not been thoroughly assessed or brought to bear on the debate between the phonological and morphological analyses. The main empirical contribution of this paper is therefore to provide a quantitative test of prediction (2b). The results of this test turn out to also favor the morphological over the phonological analysis.

Section 2 provides some general background on French liaison and presents the phonological and morphological analyses. Section 3 reviews theoretical and empirical arguments that have been made for either analysis and discusses the data that provide evidence against prediction (2a) of the phonological analysis.

The remainder of the paper describes how prediction (2b) of the morphological analysis was tested. A database of liaison and non-liaison words was extracted from Lexique 3.83 (New et al 2001) and used as a testing ground. Three distinct studies were carried out using this database.[1] The first study (Section 4) includes all liaison and corresponding non-liaison words in Lexique 3.83 and controls for the effects of grammatical morphology (i.e. the presence of inflectional suffixes, such as plural *-s*) and consonant identity on lexical frequency. The second and third studies focus on subsets of the database to allow for more controlled tests of the morphological analysis. The second study (Section 5) focuses on masculine singular adjectives such *grand* 'great.masc' or *chanceux* 'lucky.masc', and controls for effects of lexical morphology (e.g., the presence of lexical suffixes, such as *-eux* in *chanceux*) and consonant identity on lexical frequency. The third study (Section 6) focuses on monomorphemic words and controls for the effect of consonant identity on lexical frequency. The results of these three studies are generally compatible with the predictions of the morphological analysis and therefore suggest that French liaison is better understood as allomorphy than allophony synchronically.

2 French liaison: allophony or allomorphy?

Section 2.1 reviews the basics of French liaison. Section 2.2 presents the phonological and morphological analyses, as well as various specific implementations of these analyses.

[1] The data and code are available in Storme (2024b).

2.1 French liaison: the basics

In French, some words show an alternation involving a special consonant-final form. For instance, the masculine adjective *grand* 'great' is usually pronounced as [gʁɑ̃], as shown in (3a,b). But before vowel-initial nouns, it may be pronounced as [gʁɑ̃t] with a final consonant, as shown in (3c). The consonant at the end of this form is called a liaison consonant (Côté 2011; Bonami & Delais-Roussarie 2021).

(3) Liaison alternation in French

		Word1	Word2		Context
a.	*grand*	[gʁɑ̃]	(none)	'great'	(citation form)
b.	*grand monsieur*	[gʁɑ̃]	[məsjø]	'great man'	(before C-initial nouns)
c.	*grand ami*	[gʁɑ̃t]	[ami]	'great friend'	(before V-initial nouns)

Liaison alternations do not apply to all words in French. For instance, the adjective *chouette* 'nice' uses the same consonant-final form [ʃwɛt] across all phonological contexts, as shown in (4). In what follows, the final consonant of [ʃwɛt] in (4) is called a non-liaison consonant and words like *chouette* that are not subject to liaison alternations are called non-liaison words.

(4) Absence of liaison alternation

		Word1	Word2		Context
a.	*chouette*	[ʃwɛt]	(none)	'nice'	(citation form)
b.	*chouette monsieur*	[ʃwɛt]	[məsjø]	'nice man'	(before C-initial nouns)
c.	*chouette ami*	[ʃwɛt]	[ami]	'nice friend'	(before V-initial nouns)

For liaison words such as *grand*, the distribution of the two forms is conditioned by a complex interaction of phonological, syntactic, lexical and stylistic factors (Côté 2011:Section 2.2). A basic generalization is that the word following the liaison form always begins with a vowel. But not all vowel-initial words can trigger liaison in the preceding word. In particular, a specific set of vowel-initial words, *h-aspiré* words, block liaison. Also, liaison is not possible in all syntactic contexts: liaison may be observed between an adjective and a singular noun, as in (3), but not between a singular noun and an adjective. For instance, the liaison consonant [z] is not available between the noun and the adjective in *repas italien* 'Italian meal'. Finally, the use of liaison depends on speech style and register, with the rate of liaison increasing in more elevated styles.

There is a restricted set of liaison consonants in French: [z], [t], [n], [ʁ], [p], and [g] (Bonami & Delais-Roussarie 2021:2110). Examples of liaison words with these consonants are provided in (5). Liaison consonants vary widely in their distribution across words types. Liaison [z] and [t] appear in many lexical items, e.g. in adjectives such (5a) and (5d), and play an important role as grammatical markers in verbs, nouns, and adjectives. More specifically, liaison [z] is a plural marker in nouns and adjectives, as shown in (5b), and a non-third person marker in verbs, as shown in (5c). Liaison [t] is a third-person marker in verbs, as shown in (5e). However, liaison [g] and [p] occur only in a handful of words, such as (5f) and (5g). Liaison [n] and [ʁ] stand in the middle in terms of their frequencies across word types: they occur in adjectives, as illustrated in (5h)[2] and (5i), and for [ʁ] also in infinitives, as illustrated in (5j).

[2] Note that many words with a basic form ending in a nasal vowel (e.g., *bon* [bɔ̃] 'good', *plein* [plɛ̃] 'full') denasalize the vowel in the liaison form (e.g., [bɔn], [plɛn]).

(5) List of liaison consonants

		Examples		
[z]	a.	*gros*	[gʁo] ∼ [gʁoz]	'big.MASC.SG'
	b.	*grands*	[gʁɑ̃] ∼ [gʁɑ̃z]	'great.MASC.PL'
	c.	*suis*	[sɥi] ∼ [sɥiz]	'be.PRES.1SG'
[t]	d.	*grand*	[gʁɑ̃] ∼ [gʁɑ̃t]	'great.MASC.SG'
	e.	*est*	[ɛ] ∼ [ɛt]	'be.PRES.3SG'
[g]	f.	*long*	[lɔ̃] ∼ [lɔ̃g]	'long.MASC.SG'
[p]	g.	*trop*	[tʁo] ∼ [tʁop]	'too much/many'
[n]	h.	*mon*	[mɔ̃] ∼ [mɔ̃n]	'my.MASC.SG'
[ʁ]	i.	*dernier*	[dɛʁnje] ∼ [dɛʁnjeʁ]	'last.MASC.SG'
	j.	*chanter*	[ʃɑ̃te] ∼ [ʃɑ̃teʁ]	'sing.INF'

2.2 Phonological and morphological analyses of French liaison

As previewed in the introduction, there are two broad views about the nature of liaison. In the phonological analysis, liaison is a phonological alternation between a consonant and zero. In the morphological analysis, liaison is a morphological alternation between a shorter and a longer allomorph. This section reviews the two analyses in turn.

In the phonological analysis, liaison words are special because they feature a special kind of *phoneme* at their right edge. Liaison phonemes differ from non-liaison phonemes in having a null allophone in addition to their consonantal allophone. The special phonological status of liaison consonants is indicated as a subscript in what follows. For instance, the liaison consonant /$t_{liaison}$/ has two allophones [t] and ∅ whereas non-liaison /t/ has a single allophone [t], as depicted in (6). Liaison and non-liaison words such as *grand* and *chouette* differ in their underlying segmental make up: *grand* /gʁɑ̃$t_{liaison}$/ ends with /$t_{liaison}$/ whereas *chouette* /ʃwɛt/ ends with /t/.

(6) Liaison as allophony

	Phoneme	Allophones
Liaison consonant	/$t_{liaison}$/	∅, [t]
Non-liaison consonant	/t/	[t]

The exact interpretation of the underlying phonological difference between liaison and non-liaison consonants varies in the literature. The earliest generative accounts analyze liaison consonants as the only true word-final consonants of French whereas final non-liaison consonants are analyzed as being always followed by /ə/ underlyingly (Schane 1968; Dell 1985). For instance, *grand* is analyzed as /gʁɑ̃t/ underlyingly and *chouette* as /ʃwɛtə/. The final /t/ in /gʁɑ̃t/ deletes when not followed by a vowel, resulting in [gʁɑ̃] ∼ [gʁɑ̃t] alternations. The final /ə/ in /ʃwɛtə/ always deletes but protects the preceding /t/ from deleting because the rule of schwa deletion is ordered after the consonant deletion rule. This results in a unique pronunciation [ʃwɛt] for this word.

Later phonological accounts treat both liaison and final non-liaison consonants as word-final segments underlyingly, but assume a special underlying status for liaison consonants to derive the difference in behavior with non-liaison consonants. Liaison consonants are treated either as floating segments (e.g., Encrevé 1988; Tranel 1990; see Tranel 1995 for an overview), as latent segments (Bonami, Boyé & Tseng 2005; Eychenne 2011) or as gradient symbolic representations (Smolensky & Goldrick 2016; Smolensky et al. 2020; Tessier & Jesney 2021). Despite these differences, these

accounts all have in common to treat liaison as a phonological alternation between a consonant and zero.[3]

In the morphological analysis, liaison consonants do not have anything special phonologically but they occur in special *words*: these words have two listed allomorphs differing by the presence/absence of an extra final consonant (e.g., Gaatone 1978; Klausenburger 1984; Steriade 1999; Bonami & Boyé 2005; Bonami 2011; Storme 2024c). For instance, the masculine singular adjective *grand* 'great' has two allomorphs, /gʁɑ̃/ and /gʁɑ̃t/, as depicted in (7). The allomorph /gʁɑ̃t/ features a final consonant /t/ that is absent from the other allomorph /gʁɑ̃/. The distribution of these allomorphs is conditioned by a number of linguistic factors, and in particular by the phonological shape of the following word: the consonant-final allomorph is only selected before vowel-initial words (see Haspelmath & Sims 2010:25-26, Inkelas 2014:Chapter 9.1 on phonologically conditioned allomorphy). By contrast, non-liaison words such as *chouette* come with a single allomorph, as depicted in (7). In this analysis, consonants at the end of non-liaison words like *chouette* and consonants at the end of liaison words like *grand* are identical phonologically: liaison consonants do not have a special phonological status underlyingly. This differs from the phonological approach.

(7) Liaison as allomorphy

	Morpheme	Allomorphs
Liaison word	*grand* 'great'	/gʁɑ̃/, /gʁɑ̃t/
Non-liaison word	*chouette* 'nice'	/ʃwɛt/

Other morphological analyses have been proposed in the literature where liaison consonants are not analyzed (or at least not only) as word-final segments. For instance, Bybee (2001) analyzes liaison consonants as absent from the underlying representation of Word1 but lexically affiliated to a multiple-word construction including both Word1 and Word2 (e.g., *grand ami* 'great friend' is stored in the lexicon as a construction /gʁɑ̃ t ami/). Liaison consonants have also been analyzed as word-initial segments at the beginning of Word2 instead of word-final segments at the end of Word1 (Morin 2003, 2005; Chevrot, Dugua & Fayol 2009). Finally, liaison consonants have also been proposed to be simultaneously word-final and word-initial segments underlyingly (Smolensky & Goldrick 2016; Smolensky et al. 2020; Tessier & Jesney 2021).

These alternative morphological analyses where the liaison consonant is underlyingly attached to Word2 or to a construction involving Word2 will not be considered further in this paper, for both methodological and theoretical reasons.

From a methodological perspective, these analyses do not provide a minimal comparison with the phonological analysis because they do not treat liaison consonants as word-final segments. The lexical frequency of Word2 also has to be taken into consideration when evaluating the predictions of these theories regarding the lexical distribution of liaison, whereas only the lexical frequency of Word1 is relevant for the comparison involving the morphological theory in (7).

From a theoretical perspective, these alternative morphological analyses have been proposed mainly to account for two facts: (i) why liaison consonants may attach prosodically not only to Word1 but also to Word2 across a prosodic break and (ii) why the rate of liaison depends on lexical

[3] Another phonological account has been proposed where liaison consonants are treated as epenthetic (Tranel 1981:251; Morin 1983). However the identity of the liaison consonant varies depending on the word (e.g., *grand* 'great' uses [t] as liaison consonant but *gros* 'big' uses [z]). Moreover not all vowel-final words have a liaison variant (e.g., *vrai* 'true' uses a single vowel-final form [vʁɛ] across the board). As a consequence, the rule of epenthesis must be lexically conditioned. The resulting analysis is then nothing but a notational variant of the morphological analysis, as observed by Klausenburger (1984:27). More generally, Morley (2018) has also argued against the availability of consonant epenthesis in the repertoire of synchronic phonological processes in the world's languages, based on evidence from artificial grammar learning experiments.

properties of both Word1 and Word2. But neither of these facts requires that liaison consonants be lexically affiliated to Word2. The prosodic variability of liaison consonants can be derived as a paradigm uniformity effect assuming a lexical affiliation to Word1 only, as proposed by Storme (2024c) and as further discussed in Section 3.2. Effects of Word2 on the rate of liaison can be derived as due to the way lexical properties of Word2 affect lexical access during speech-production planning, as proposed by Kilbourn-Ceron (2017). Crucially, these effects do not require liaison consonants to belong to Word2 underlyingly.

3 Arguments and data from previous research

Most of the arguments that have been made in favor of the phonological or morphological analysis are based on considerations of theoretical economy, as discussed in Section 3.1. But empirical arguments have also been made. Two empirical arguments from the previous literature are examined critically: one based on the realization of liaison consonants and in favor of the phonological analysis (Section 3.2) and the other on the non-locality of liaison alternations and in favor of the morphological analysis (Section 3.3). Section 3.4 develops a new empirical argument against the phonological analysis of liaison, based on evidence from lexical statistics showing that liaison alternations do no pattern like a typical deletion pattern, contrary to what is assumed under the phonological analysis.

3.1 Arguments based on theoretical economy

One of the main appeals of the phonological analysis is that it makes it possible to maintain a single underlying representation for liaison words: for instance, in this analysis, *grand* corresponds to a single underlying representation (/gʁɑ̃t/ in Schane 1968 and Dell 1985 or /gʁɑ̃t$_{\text{liaison}}$/ with a final liaison phoneme in Encrevé 1988; Tranel 1990; Smolensky & Goldrick 2016). By contrast, the morphological approach requires more complex lexical entries, with *grand* corresponding to two distinct underlying representations /gʁɑ̃, gʁɑ̃t/, in violation of the principle 'one form - one meaning'.

However maintaining a single lexical representation for liaison words also comes at a cost. In the early phonological accounts of Schane (1968) and Dell (1985), this cost is in an increase in the abstractness of underlying representations: for the analysis to work, all words with stable final consonants (e.g., *trente* [tʁɑ̃t] 'thirty') must be assumed to contain a final schwa underlyingly (e.g., /tʁɑ̃tə/), even though this schwa is virtually never pronounced. This abstract schwa is necessary to protect the preceding consonant from undergoing the consonant-deletion rule that gives rise to liaison alternations. In later phonological accounts, protective schwas have been abandoned but at the cost of a more complex phoneme inventory: in these approaches, the set of French phonemes is enriched with a new type of phonemes, floating segments (Encrevé 1988; Tranel 1990) or gradient symbolic representations (Smolensky & Goldrick 2016; Smolensky et al. 2020; Tessier & Jesney 2021) .

An argument based on theoretical economy has also been made in support of the morphological analysis. Gaatone (1978:49) argues that the morphological analysis is superior to the phonological one because it allows for a unified treatment of two types of liaison alternations: (i) liaison alternations of the type *grand* [gʁɑ̃] ∼ [gʁɑ̃t], where the liaison form only differs from the non-liaison form by the presence of an extra final consonant, and (ii) liaison alternations of the type *vieux* [vjø] ∼ *vieil* [vjɛj] 'old.MASC.SG' in (8), where the liaison form [vjɛj] in (8c) differs from the non-liaison form [vjø] along more dimensions and is homophonous with the feminine form *vieille* [vjɛj]. This latter type of liaison alternations clearly involves suppletion (see also Klausenburger 1984:25-30 and

Steriade 1999), and therefore requires an allomorphic treatment. To account for the two types of liaison alternations, the phonological analysis needs to assume two different mechanisms (allophony for *grand* and allomorphy for *vieux/vieil*; e.g., Tranel 1990) whereas the morphological analysis assumes a single mechanism throughout (allomorphy; e.g., Storme 2024c) and is therefore more economical in that sense.

(8) Suppletive liaison in French

		Word1	Word2		Context
a.	*vieux*	[vjø]	(none)	'old'	(citation form)
b.	*vieux monsieur*	[vjø]	[məsjø]	'great man'	(before C-initial words)
c.	*vieil ami*	[vjɛj]	[ami]	'old friend'	(before V-initial words)

To summarize, the phonological analysis allows for more compact lexical entries at the cost of more abstractness in underlying representations or more complex sound inventories. The morphological analysis allows for a unified treatment of different types of liaison (*grand* and *vieux/vieil*) and for more concrete underlying representations, but at the cost of more complex lexical entries for liaison words. In the absence of a precise evaluation metric for comparing the complexity of the two theories, it remains unclear which one should be preferred in terms of theoretical economy.

3.2 The realization of liaison (Tranel 1990, 2000)

Empirical arguments have also played a role in the debate between phonological and morphological analyses of French liaison. This section focuses on data concerning the realization of liaison consonants, which has been used as evidence for the phonological analysis.

Liaison consonants have been found to have a distinct realization from non-liaison consonants, both prosodically (e.g., Encrevé 1988; Tranel 1990; Storme 2024c) and segmentally (e.g., Gaskell, Spinelli & Meunier 2002; Spinelli, Cutler & McQueen 2002; Spinelli, McQueen & Cutler 2003; Nguyen, Wauquier-Gravelines, Lancia & Tuller 2007; Côté 2014; Storme 2024c). In particular, liaison consonants have been found to have a prosodic and segmental realization that is intermediate between the realizations of word-final and word-initial consonants. For instance, in a Word1-Word2 sequence such as *grand ami* 'great friend', liaison consonants can attach either to Word1 or Word2 in the presence of a prosodic break between the two words (e.g., *gran*[t] *ami* or *gran* [t]*ami*). By contrast, non-liaison consonants attach categorically to Word1 (when they are lexically word-final) or to Word2 (when they are lexically word-initial) (Storme 2024c).

The fact that liaison consonants differ from non-liaison consonants in their realization has been used as an argument in favor of the phonological analysis, on the assumption that differences in surface realization imply differences in phonological representations. For instance, Tranel (1990:182-183) and Tranel (2000:50-51) built on the prosodic instability of liaison consonants to motivate different underlying phonological representations for liaison and non-liaison consonants: liaison consonants differ from non-liaison consonants in not being linked to a skeletal slot in the underlying prosodic structure of the word. This underlying phonological property then explains why liaison consonants may attach prosodically to Word1 or Word2 on the surface.

However the assumption that differences in surface realization imply differences in phonological representations is problematic, as surface differences can result from other factors. This point has been made in a number of works on incomplete neutralization, where phonetic differences between phonologically neutralized segments are not attributed to distinct phonological representations but to other mechanisms, such as paradigm uniformity among pronunciation variants of the same morpheme (e.g., Roettger, Winter, Grawunder, Kirby & Grice 2014; Braver 2019).

Storme (2024c) proposed a similar analysis for French liaison, where the special realization of liaison consonants is derived as a paradigm uniformity effect. 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. For instance, the liaison consonant at the end of *grand* may be realized at the beginning of Word2 (*gran* [t]*ami* 'great friend') as a way to enhance the similarity between the contextual variant of Word1 and its citation form, which does not end in [t] (*grand* [gʁɑ̃]). But it may also be realized at the end of Word1 (*gran*[t] *ami*) as a way to enhance the similarity between the contextual variant of Word2 and its citation form, which does not begin with [t] (*ami* [ami]). This account derives different segmental and prosodic behaviors for liaison and non-liaison consonants while assuming the same underlying phonological representations for the two types of consonants. In other words, the fact that liaison consonants have a special realization does not provide an argument for the phonological analysis, as this type of effects can be derived under the morphological analysis enriched with paradigm uniformity.

3.3 Non-locality of liaison (Bürki, Frauenfelder & Alario 2015)

Another empirical argument was proposed by Bürki et al. (2015), this time in favor of the morphological analysis. They proposed to compare the two theories by investigating whether liaison alternations are sensitive to non-local information in the context. Locality is relevant to debates about the representation of liaison on the assumption that phonological processes are generally local whereas allomorph selection may be sensitive to non-local phonological factors. This assumption is crucial in their test and will be discussed later in this section.

The experiment in Bürki et al. (2015) was based on a picture-naming task. Participants were presented with a picture representing an object and were asked to produce a noun phrase to describe that picture, either using *un demi* N 'a half N' (when only half of the object was visible on the picture) or *un immense* N 'a huge N' (when a big object was visible on the picture). Adjectives *demi* and *immense* were chosen because they require the non-liaison and liaison forms of the determiner, respectively: *demi* is consonant-initial and should therefore favor the non-liaison form *un* [ɛ̃] 'a' whereas *immense* is vowel-initial and should favor the liaison form *un* [ɛ̃n] 'an'.

Bürki et al. (2015) found that naming latencies and determiner duration were affected by non-local information in the noun, in particular by whether the adjective and the following noun were consistent along their phonological onsets: naming latencies and determiner duration were found to be longer when the phonological onsets were inconsistent (the adjective starts with a vowel and the noun by a consonant or the other way around) than when they were consistent (the adjective and the noun both start with a vowel or with a consonant). These results show that liaison alternations are sensitive to non-local information.

Bürki et al. (2015) interpreted these results as supporting the morphological analysis, on the assumption that a phonological rule would only be sensitive to local information. In the model they propose, both the adjective and noun activate the corresponding determiner form and the selection is completed faster when the two words send activation to the same determiner form.

This argument is clearly empirical, contrary to the arguments discussed in Section 3.1. However it crucially relies on the assumption that a phonological rule has to be local, as acknowledged by Bürki et al. (2015:433). However phonological processes may be non-local. For instance, in Ineseño Chumash, there is a regressive pattern of sibilant harmony, whereby all sibilants are required to agree for anteriority in a word, regardless of the distance between them (Applegate 1972). This can be illustrated with the word /ha-s-xintila-waʃ/ → [haʃxintilawaʃ] 'his former gentile name', where /s/ changes to [ʃ] in the first syllable of the word under the non-local influence of /ʃ/ in the last

syllable of the word.

If the phonological process responsible for liaison alternations can apply both locally and non-locally, then there is a way to predict longer naming latencies and a longer determiner duration in case of a phonological mismatch between the onsets of the adjective and the noun: this happens when the output determined locally (e.g., the liaison form *un* [ɛ̃n] in *un immense* 'a huge') conflicts with the output determined non-locally (e.g., the non-liaison form *un* [ɛ̃] in *un ... ballon* 'a ... ball').

Another assumption made by Bürki et al. (2015) is that allomorph selection is more likely to involve a non-local relation between the trigger and the target than a phononological rule. However most cases of allomorphy reported in the literature also imply a local relation between the trigger and the target (Dolatian 2023). So it is unclear that non-locality can be used to arbitrate between the two analyses.

3.4 A new empirical argument: lexical-frequency effects on liaison

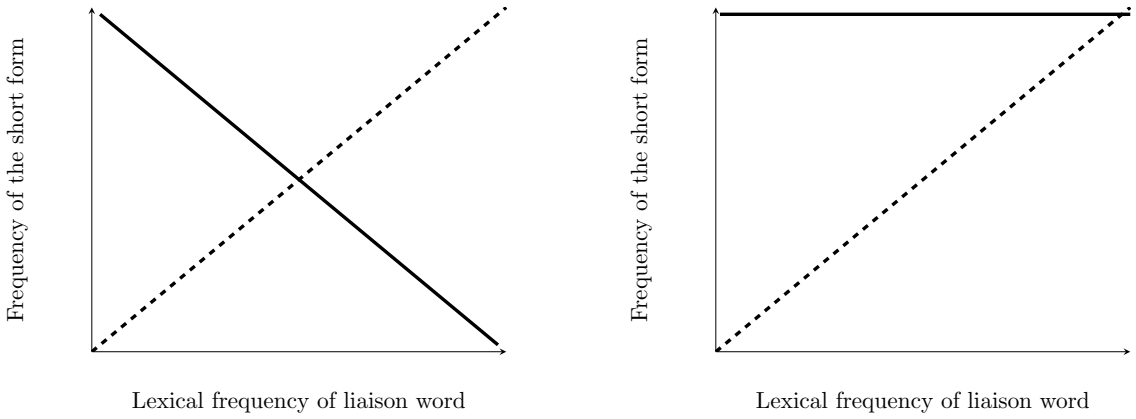
There is another empirical argument to be made against the phonological analysis, based on the way segment deletion interacts with lexical frequency across languages. Words are more likely to undergo reduction processes at the segmental level (shortening, lenition, deletion) when they have higher probability (Jurafsky et al. 2001; Bell et al. 2009). For instance, word-final /t,d/-deletion is more likely in high frequency words than in low frequency words in English (Bybee 2007:chapter 9; Coetzee & Kawahara 2013) and in Dutch (Goeman 1999:179-184). Similarly, the likelihood of word-final /s/-deletion in Spanish increases with lexical frequency (File-Muriel 2010). This correlation between reduction and lexical frequency has been explained in terms of general principles of speech production, either as manifesting a pressure towards efficient communication (e.g., Aylett & Turk 2004) or as reflecting constraints on lexical access (e.g., Bell et al. 2009).

Under the phonological analyses of French liaison reviewed in Section 2.2, liaison alternations arise from a deletion process: liaison words have an underlying final consonant that may undergo deletion. Under the default assumption that consonant deletion is governed by the same general speech production principles across segment types, phonological analyses of French liaison therefore predict a *positive* correlation between deletion and lexical frequency: the short form of a liaison word (which is derived through consonant-final deletion under the phonological account) should become more frequent as the lexical frequency of that word increases, as schematized by the dashed lines in Figure 1.

However, research on French liaison has actually found (i) a *negative* correlation between these two variables prevocalically, as schematized by the solid line in Figure 1a, and (ii) an *absence* of correlation non-prevocalically, where only the short form (without liaison) is used, as schematized by the solid line in Figure 1b.

Prevocalically, the non-liaison form has been found to become *less* likely relative to the liaison form as lexical frequency increases (Fougeron et al. 2001; Kilbourn-Ceron 2017). Kilbourn-Ceron (2017:143-145) reports that this effect holds even for liaison words that belong to open-class categories, namely plural nouns in PlNoun-Adj sequences (e.g., *les pas* [z] *enjoués* 'the cheerful steps') and singular adjectives in Adj-Noun sequences (e.g., *petit* [t] *ami* 'little friend'). Also effects of word length, lexical identity, speaker, and lexical frequency of the following word were carefully controlled for by Kilbourn-Ceron (2017) in assessing the role of the lexical frequency of liaison words. This means that there is a genuine effect of the lexical frequency of liaison words on the rate of liaison, with higher lexical frequency correlating with a higher frequency of use of the liaison form.

To illutrate this point more concretely, consider two liaison words with different lexical frequencies: the high-frequency liaison word *gros* [gʁo] ∼ [gʁoz] 'big' (with 180.91 occurrences per million



(a) Prevocalically (b) Prepausally and preconsonantally

Figure 1: Schematized correlation between the lexical frequency of liaison words and the frequency of use of the short form (without liaison), as predicted under the phonological analysis (dashed lines) and as observed in corpus data (solid lines).

words; New, Brysbaert, Veronis & Pallier 2007) and the lower-frequency adjective *constant* [kɔ̃stɑ̃] ∼ [kɔ̃stɑ̃t] 'constant' (with 1.88 occurrences per million words; New et al. 2007). Kilbourn-Ceron's (2017) results imply that *gros* is more likely to appear under its liaison form than *constant*. The phonological analysis of liaison as a reductive process incorrectly predicts the opposite, namely that *constant* should be more likely to appear under its liaison form than *gros*.

The effect of lexical frequency on the rate of liaison has been established for prevocalic positions. However, outside of this context (i.e. prepausally and preconsonantally), no such effect has been reported: only the non-liaison form (e.g., *grand* [gʁɑ̃] 'great') is used, as schematized in Figure 1b. This is also problematic for an analysis in terms of deletion. For instance, in English, word-final /t,d/-deletion is probabilistic and positively correlated with lexical frequency in all contexts: prevocalically, preconsonantally, and prepausally (Coetzee & Kawahara 2013:Figure 4).

These facts are problematic for phonological analyses of liaison, as they strongly suggest that liaison does not involve deletion and is therefore not a reductive process. A reviewer argues that the phonological account of liaison consonants as floating segments (e.g., Encrevé 1988; Tranel 1990) might be immune from this criticism, if one assumes that floating segments are not sensitive to the same speech-production constraints as non-floating segments, for which the effect of lexical frequency on reduction has been documented. However there does not seem to be a principled way to exclude floating segments from the set of segments whose reduction/deletion is conditioned by lexical frequency. Indeed, schwa is also typically analyzed as a floating segment by proponents of the floating-segment analysis for liaison consonants (e.g., Tranel 1987). Yet schwa deletion patterns like a regular reduction process: schwa is more likely to delete in more frequent words (e.g., Eychenne 2019). A proponent of the floating-segment analysis would need to explain why the deletion of floating segments sometimes patterns like a regular reduction process (in the case of schwa) and sometimes not (in the case of liaison consonants).

The fact that the liaison form never surfaces outside of prevocalic contexts is more compatible with an analysis of liaison as an external sandhi: the liaison form is a special form of a word that is only used in prevocalic contexts for phonotactic reasons (to break a hiatus) or for morphological reasons (to signal the presence of a morpheme), as argued by Storme (2024a). Kilbourn-Ceron (2017:chapter 4) has proposed an interpretation of lexical-frequency effects on liaison under the external-sandhi hypothesis. In her analysis, an external sandhi process arises at the junction

between two words when those two words are planned in the same speech-production planning window. And two words are more likely to be planned together if they are easy to retrieve from memory. Higher lexical frequency facilitates word-form retrieval, and hence liaison words with higher lexical frequency (e.g., *gros* 'big' as compared to *faux* 'false') are more likely to appear under their sandhi/liaison form.

The interpretation of liaison as an external sandhi process that does not involve reduction/deletion is compatible with the morphological analysis. In this analysis, the liaison and non-liaison forms are not derived phonologically from a single underlying representation but are both listed underlyingly. The liaison form is used before vowels for phonotactic or morphological reasons. In other words, liaison is a phonologically/morphologically optimizing pattern of suppletive allomorphy (see Inkelas 2014:Chapter 9 on phonologically optimizing suppletive allomorphy).

To the author's knowledge, patterns of suppletive allomorphy have not been reported to be sensitive to lexical frequency in the same way as reduction processes. Hence, the fact that the longer allomorph (the liaison form) becomes more likely as lexical frequency increases in French, in violation of the strong cross-linguistic tendency for more frequent words to be shorter (Zipf 1929; Levshina 2022), is not necessarily problematic for the morphological analysis. However it opens an interesting question for future research about how the principles that drive external sandhi (i.e. high-frequency words are more likely to use a sandhi form) interact with lexical frequency effects on word length (i.e. high-frequency words are more likely to be shorter). French presents an interesting case study, as these two factors conflict: prevocalically, the liaison/sandhi form is phonologically and/or morphologically more optimal but also longer than the non-liaison form. In French, the conflict is solved in favor of the phonological/morphological constraints over the frequency effect. Levshina (2022:Section 4.2) provides similar evidence from Arabic and Russian for the hypothesis that frequency effects might be overriden by other linguistic factors in determining word length.

4 Testing the morphological analysis: Study 1

The fact that French liaison does not behave like a phonological reductive process, as shown in Section 3.4, provides only *indirect* evidence for the morphological analysis. The goal of Sections 4-6 is to provide a *direct* test of this analysis, using again evidence from lexical statistics.

As previewed in the introduction, this test is based on the observation that allomorphy is typically limited to the high-frequency range of the lexicon: it affects a relatively small number of words that occur very frequently (Bybee 1985:Chapter 5; Berg 2011). For instance, irregular verbs across languages tend to display suppletive stem allomorphy (cf. English *go/went*, French *vont/allaient/iront* 'they go/went/will go') and also happen to have low type frequency and high token frequency within languages. By contrast, words that do not display allomorphy (e.g., English *dance* or French *danser* 'to dance') are distributed more uniformly across the lexical-frequency range.

If French liaison is a morphological alternation, then the prediction in (9) should hold.

(9) Prediction of the morphological analysis
Liaison words have lower type frequency but higher token frequency than non-liaison words.

In line with the first part of prediction (9), liaison has been argued to involve only closed lexical classes and specific morphological categories (Côté 2011:Section 2.2.1). However this claim has not been tested thoroughly and in a controlled way. Section 4 tests prediction (9) while controlling for the effect of phonotactics and grammatical morphology on lexical frequency. Section 4.1 presents the dataset used as a testing ground and explains how liaison and additional phonological and

morphological information were coded. Section 4.2 presents the results of the two statistical analyses that were conducted on this dataset. Section 4.3 concludes with a discussion. The dataset and R script are available in Storme (2024b) (`study1-data.csv` and `study1-script.R`, respectively).

4.1 Dataset and coding

4.1.1 Dataset: Lexique 3.83

To test prediction (9), one must minimally have two types of information: a list of liaison and non-liaison words along with the number of times each word occurs in a corpus. To the author's knowledge, there is as of now no such list available. However it is possible to build one using Lexique 3.83, a large open-access French lexicon (New, Pallier, Brysbaert & Ferrand 2004; New & Pallier 2023).

Lexique 3.83 includes information about token frequencies of individual words, making it an ideal resource for testing hypotheses about lexical statistics. In this study, the frequency measure based on a corpus of subtitles (column `freqfilm2`) was used, as it reasonably aligns with spoken speech. Word frequency is measured in fpm (frequency per million words) in Lexique 3.83, with the smallest frequency being 0.01.

4.1.2 An orthographic definition of liaison

Lexique 3.83 lacks information regarding the liaison status of words but does offer an orthographic and phonological transcription (see columns `ortho` and `phon`, respectively). Typically, the liaison status of a word can be determined by comparing these two representations. The phonological form, which employs a slightly modified version of the X-SAMPA transcription system, corresponds to the citation form of the word. For example, the word *grand* is phonologically transcribed as [gʁɑ̃] in Lexique 3.83,[4] indicating the absence of liaison. Liaison consonants are consistently present in the orthographic form but omitted from the phonological representation provided by Lexique 3.83. Consequently, it becomes feasible to identify liaison words as those displaying a disparity between these two forms. More specifically, a liaison word is generally characterized by the following attributes:

- its graphic form ends with a graphic laison consonant (*n, r, t, d, s, p, g, x,* or *z*) and
- the phonological form of its citation form does not end with the corresponding phonological liaison consonant in [n, ʁ, t, z, p, g].

This definition accurately describes the bulk of liaison words. For instance, it correctly identifies *grand* as a liaison word. In this case, the graphic form of the word ends with a consonant in the set of graphic liaison consonants (specifically *d*) but its phonological form [gʁɑ̃] does not end with the corresponding consonant (specifically [t]). There is a small set of liaison words that do not fit this definition. These are the liaison words with a graphic liaison consonant that matches the final consonant of their citation form. This occurs when their stem ends in a consonant that belongs to the set of phonological liaison consonants (e.g., *chantent* 'sing.3PL' [ʃɑ̃t] ∼ [ʃɑ̃tət], *ronces* 'brambles' [ʁɔ̃s] ∼ [ʁɔ̃səz]). An exception must be made for these words.

It is important to acknowledge that the definition of liaison used in this paper is orthographic and therefore highly inclusive. Some words that are characterized as liaison words in this study are actually seldom manifested as such, and only occur under their liaison forms in specific registers.

[4] Transcriptions from Lexique 3.83 are rendered in the International Phonetic Alphabet in this paper.

For example, the liaison form of the verb *chantent* 'sing.3PL' [ʃɑ̃tət] is mostly limited to formal speech. Consequently, the definition of liaison employed is notably comprehensive, encompassing more words than are commonly observed in casual speech. Nevertheless, a specific systematic restriction was imposed. Singular nouns were excluded from the corpus because liaison is categorically disallowed after singular nouns, as detailed in Côté (2011:Section 2.1.1).

An inclusive definition is desirable for two reasons. First, words that do not behave as liaison words in casual speech (e.g., *chantent*) may behave as liaison words in formal speech and therefore should not be entirely ignored. Second, an inclusive definition of liaison imposes a more stringent test of the morphological analysis, as it inflates the number of liaison words, making it harder to corroborate the hypothesis that liaison words are fewer in the lexicon.

4.1.3 Controlling for phonotactics and inflectional morphology

The liaison status of a word is not the only factor that is likely to influence lexical frequency. Two additional factors were controlled for in Study 1. The first factor is the identity of the final consonant. Consonants have different type and token frequencies in French, including when they appear in word-final position (see Malécot 1974:161 for data on token frequency). To control for potential phonotactic effects on lexical frequency, only non-liaison words ending in one of the consonants attested as liaison consonants (i.e. [n, ʁ, t, z, p, g]) were included in the final corpus. Furthermore, the identity of the final consonant was recorded for each word.

Another factor that necessitates control is the morphological composition of words, specifically the final morpheme within the word. Some words share a common final morpheme, which means they cannot be treated as independent entities. For example, *fruits* and *voitures* both exhibit a liaison variant ending in [z]. But, in both cases, this final [z] corresponds to the same plural suffix, alternating with zero (see Dell 1985:183, Eychenne 2011, Storme 2024a for analyses of liaison /z/ as a plural morpheme).

In Study 1, only the impact of *inflectional* suffixes such as plural *-s* on lexical frequency was controlled for. The list of inflectional suffixes identified in Study 1 is available in the Appendix. Concretely, plural nouns like *fruits* 'fruits' and *voitures* 'cars' were counted as instances of the same plural morpheme and therefore collectively contributed one unit for the type frequency analysis. For the token-frequency analysis, their counts in the corpus of subtitles were added up to obtain the overall token frequency of the corresponding inflectional suffix. Words that do not end in an inflectional suffix were treated as independent lexical items, contributing each one unit to the type frequency.

It is particularly important to control for inflection when assessing the morphological analysis of liaison as there is evidence that inflectional suffixes are stored independently in the mental lexicon and segmented by language users in online language processing. For instance, Leminen, Leminen, Lehtonen, Nevalainen, Ylinen, Kimppa, Sannemann, Mäkelä & Kujala (2011) found that words formed with an inflectional suffix (such as the plural suffix) activate a brain area that is involved in morphological structure building, i.e. the left temporal area, whereas words formed with a derivational suffix do not. They argue that it supports a fully decompositional analysis for inflectional morphology specifically (see also Leminen, Smolka, Duñabeitia & Pliatsikas 2019).

There is another, more practical reason for not controlling the effect of derivational suffixes, such as suffix *-eux* in *glorieux* 'glorious' and *chanceux* 'lucky'. Lexique 3.83 does not provide consistent information on derivational morphology.[5] To explore this aspect, the reader is referred to Study 2,

[5] There is some information about the derivational status of words in Lexique 3.83 (see column `morphoder`) but it is not always consistent. For instance, although the adjective *abassourdissant* clearly contains a derivational suffix *-ant*, this suffix is not identified as such in Lexique 3.83.

which is a smaller-scale study that includes semi-automatized control for lexical morphology, and Study 3, where only monomorphemic words are considered.

4.2 Results

4.2.1 Type-frequency analysis

Among the 142,694 words listed in Lexique 3.83, 78,538 were found to be liaison words, using the orthographic definition described in Section 4.1. This means that more than half of the words in Lexique 3.83 are liaison words. Although this might appear to contradict the prediction that liaison words have a lower type frequency than non-liaison words, this raw figure should be interpreted cautiously. It does not account for the morphological relationships between many words, especially those ending with the same inflectional suffix. For instance, among these 78,538 liaison words, 34,388 are plural nouns or adjectives ending with the same plural suffix [-z], such as *fruits* 'fruits' and *voitures* 'cars'.

The picture is quite different when inflectional morphology and phonotactics are controlled for. Figure 2 and Table 1 show, in graphical and numerical form, respectively, the number of liaison and non-liaison words in Lexique 3.83 when words sharing the same inflectional suffix collectively contribute only one unit to the type frequency, in line with the hypothesis that inflectional morphology is decompositional (see Section 4.1.3). To control for phonotactic effects on lexical frequency, words are also grouped by the identity of the final consonant.

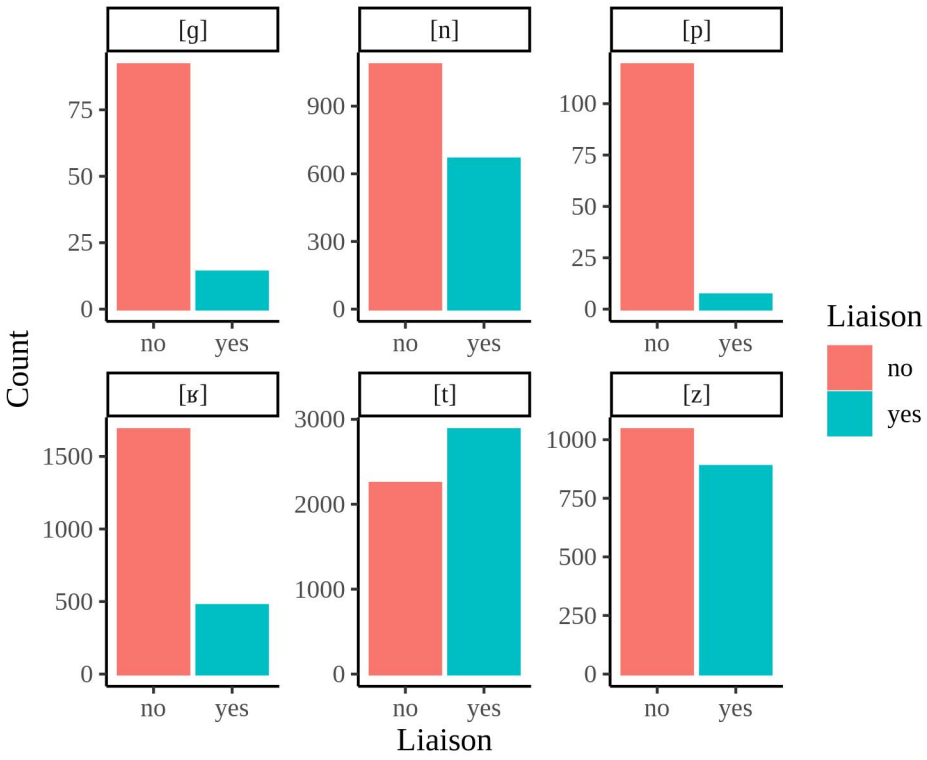


Figure 2: Descriptive statistics for the type-frequency analysis (Study 1): count of liaison and non-liaison words in Lexique 3.83 as a function of the identity of the word-final consonant. Word count is determined after the segmentation of inflectional suffixes, with words sharing an inflectional suffix counting as one.

	Liaison		Sum
FinalConsonant	no	yes	
[g]	92	14	106
[n]	1084	666	1750
[p]	119	7	126
[ʁ]	1685	473	2158
[t]	2247	2879	5126
[z]	1043	886	1929
Sum	6270	4925	11195

Table 1: Descriptive statistics for the type-frequency analysis (Study 1): count of liaison and non-liaison words in Lexique 3.83 as a function of the identity of the word-final consonant. Word count is determined after the segmentation of inflectional suffixes, with words sharing an inflectional suffix counting as one.

Descriptively, liaison words appear to be fewer than non-liaison words in Lexique 3.83 when inflectional morphology and phonotactics are controlled for, and this appears to hold for each consonant, except for [t]. To check whether this generalization is statistically robust, a logistic regression was fit to the data, using the package for Bayesian statistics `brms` (Bürkner 2017) in R (R Core Team 2021). A Bayesian approach was adopted (rather than a frequentist approach) because it provides more intuitive and meaningful inferences (Kruschke & Liddell 2018). The default parameters and priors set in `brms` for logistic regressions were used. For inference criteria, the posterior probability of the hypothesis was used: if this probability was higher than 0.95, then compelling evidence was considered to be provided for the hypothesis.

The dependent variable was the Liaison status (yes/no) of the final morpheme after segmentation of grammatical suffixes (with 'no' as the reference level). The independent variable was the dummy-coded variable Final consonant (reference level = [g]). Due to the way Final consonant was coded, a negative intercept means that liaison morphemes ending in [g] are fewer than non-liaison morphemes ending in [g]. The estimates for the effect of Final consonant vary by consonant and indicate by how much this difference must be adjusted for the other consonants attested as liaison consonants ([z, t, ʁ, p, n]). The log odds of liaison words ending in consonants other than [g] were obtained by adding up the intercept and the estimate of Final consonant for each consonant.

The estimated log odds of liaison words are shown in Table 2 for each consonant, along with the corresponding 90% Credibility Interval (CI), and the posterior probability that this log odds is negative. A negative sign for the estimated log odds of liaison words means that liaison words are fewer than non-liaison words in the lexicon, in line with the morphological analysis of liaison. All the estimates shown in Table 2 were obtained using the `hypothesis()` function in `brms`.

For all consonants except [t], the log odds of liaison words was found to be negative with a posterior probability larger than 0.95, meaning that there is very strong evidence that liaison words have *lower* type frequency than non-liaison words. These results are in line with the predictions of the morphological analysis of liaison. However, for words ending in [t], the posterior probability of the hypothesis was found to equal 0, meaning that there is very strong evidence that liaison words ending in [t] *do not* have lower type frequency than non-liaison words ending in [t]. As shown in Table 2, the log odds of liaison words was actually found to be positive for [t], meaning that liaison words ending in [t] have *higher* type frequency than non-liaison words ending in [t]. This unexpected result will be further discussed in Section 4.3.

Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Post.Prob
logodds(LIAISON=yes, [g]) < 0	-1.91	0.30	-2.42	-1.43	1
logodds(LIAISON=yes, [n]) < 0	-0.49	0.05	-0.57	-0.41	1
logodds(LIAISON=yes, [t]) < 0	0.25	0.03	0.20	0.29	0
logodds(LIAISON=yes, [ʁ]) < 0	-1.27	0.05	-1.36	-1.18	1
logodds(LIAISON=yes, [z]) < 0	-0.16	0.05	-0.24	-0.09	1
logodds(LIAISON=yes, [p]) < 0	-2.91	0.40	-3.60	-2.29	1

Table 2: Inferential statistics for the type-frequency analysis (Study 1): log odds of liaison words for each consonant (estimate, estimated error and 90% Credibility Interval) and posterior probability that this log odds is negative. A negative log odds means that liaison words are fewer than non-liaison words in the lexicon.

4.2.2 Token-frequency analysis

Although liaison words are fewer in the lexicon than non-liaison words, they appear to be more frequently used in the corpus of movie subtitles, as shown in Figure 3. Figure 3 shows the distribution of token frequencies for liaison and non-liaison words in the corpus after controlling for phonotactics and inflectional morphology. Token frequencies were calculated from corpus counts, assuming that words ending in the same inflectional suffix count as a single lexical item. Corpus counts were obtained by multiplying the frequencies provided in Lexique 3.83 (per million of words) by 100. The choice of 100 is due to the smallest word frequency in Lexique 3.83 being equal to 0.01. The corpus counts of inflectional suffixes were obtained by adding up the corpus counts of all words that end with these inflectional suffixes. Token frequencies (on the x-axis) are expressed using the Zipf scale (Van Heuven, Mandera, Keuleers & Brysbaert 2014) in Figure 3. The Zipf scale is a logarithmic transformation of lexical frequency, and ranges from 1 (very rare words) to 7 (highly frequent words). This scale was used in this figure instead of raw corpus counts to enhance readability and clarity.

Figure 3 shows that, overall, the frequency distribution of words follows the typical power-law relationship observed by Zipf (1949), with many words occurring rarely (with Zipf values below 4) and a small set of words occurring frequently (with Zipf values above 4). Although liaison and non-liaison words both follow this general trend, their frequency distributions are not identical: except for [t], liaison words are less represented than non-liaison words in the low-frequency range of the lexicon. This is in line with the hypothesis that liaison words have higher token frequency than non-liaison words.

The data in Table 3 are also in line with this hypothesis. Table 3 shows the average token count of liaison and non-liaison words in the corpus of movie subtitles, by the identity of the final consonant. On average, liaison words appear to be represented by a larger number of tokens in the corpus than non-liaison words, in line with the prediction of the morphological analysis that liaison words have higher token frequency than non-liaison words. Note that this also holds for [t]. This might seem to contradict the observation that liaison words in [t] outnumber non-liaison words in [t] in the low frequency range of the lexicon (see Figure 3). However this is not necessarily a contradiction if liaison words in [t] also outnumber non-liaison words in [t] in the *high* frequency range of the lexicon.

To check whether the generalization that liaison words are on average more frequent than non-liaison words is statistically robust, a negative binomial regression was fit to the word count data, with the LIAISON status (yes/no) of the word and the identity of the FINALCONSONANT and their interaction as independent variables. A negative binomial distribution is characterized by two

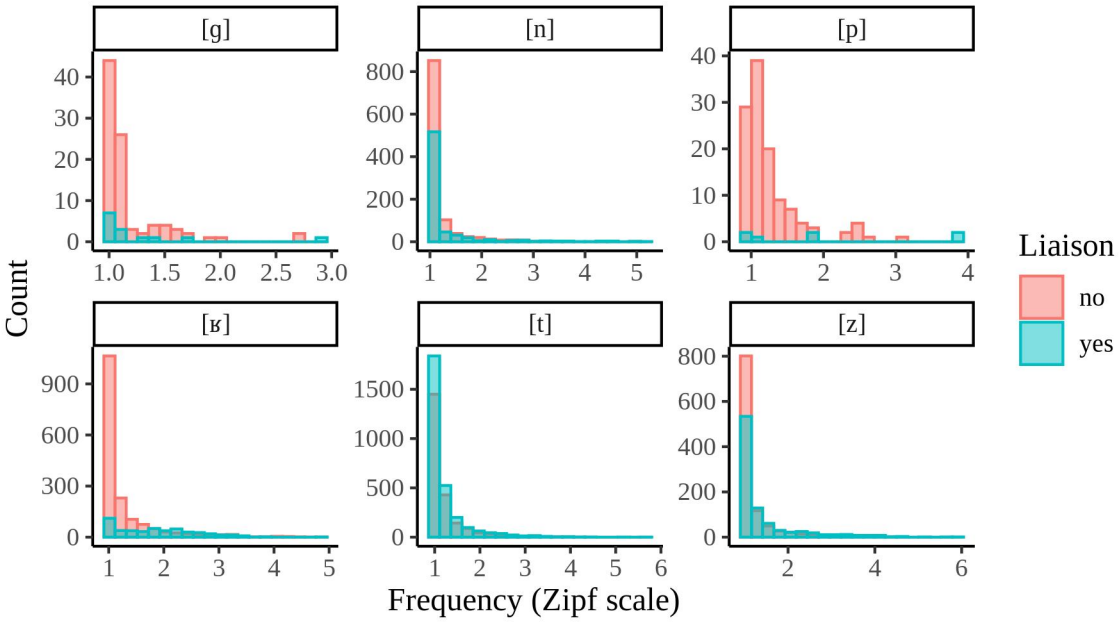


Figure 3: Descriptive statistics for the token-frequency analysis (Study 1): distribution of token frequencies (in Zipf; from 1=very rare to 7=very frequent) for liaison and non-liaison words in the corpus of movie subtitles from Lexique 3.83 as a function of the identity of the word-final consonant. The frequency values on the x-axis were binned into 20 intervals. Word count is determined after the segmentation of inflectional suffixes, with words sharing an inflectional suffix counting as a single lexical item.

	Liaison	
FinalConsonant	no	yes
[g]	178	620
[n]	1114	8222
[p]	349	21410
[ʁ]	2529	3934
[t]	472	3116
[z]	165	21765

Table 3: Descriptive statistics for the token-frequency analysis (Study 1): average token count of liaison and non-liaison words in Lexique 3.83 as a function of the identity of the word-final consonant. Word count is determined after the segmentation of inflectional suffixes, with words sharing an inflectional suffix counting as a single lexical item.

parameters: λ 'lambda', which describes the mean number of occurrences of an event, and ϕ 'phi', which characterizes the dispersion of the data (see Winter & Bürkner 2021:11-13). A negative binomial distribution is suitable to model word count data, as it is a discrete distribution that only takes integers as values (contrary to the normal distribution, which assumes a continuous dependent variable). The negative binomial regression was fit with the `brms` package (Bürkner 2017) in R, using the default paramaters and priors for this kind of regression.

If liaison words have higher token frequency than non-liaison words, we expect the token count of liaison words to be characterized by a higher λ in the corpus as compared to non-liaison words. The difference between the λ parameters corresponding to liaison and non-liaison words is shown in Table 4 for each consonant, along with the 90% credibility interval for that difference, and the posterior probability that this difference is positive. A positive sign means that λ for liaison words is larger than λ for non-liaison words, and therefore that liaison words occur more frequently than non-liaison words in the corpus, in line with the prediction of the morphological analysis of liaison.

The difference between the λ parameters of liaison and non-liaison words was found to be positive for all consonants, with a posterior probability greater than 0.95. This means that there is very strong evidence that liaison words have higher token frequency on average than non-liaison words. This result is in line with the predictions of the morphological analysis of liaison.

Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Post.Prob
$\lambda_{\text{Liaison=yes, [g]}} - \lambda_{\text{Liaison=no, [g]}} > 0$	1.48	0.90	0.12	3.10	0.97
$\lambda_{\text{Liaison=yes, [n]}} - \lambda_{\text{Liaison=no, [n]}} > 0$	2.00	0.14	1.77	2.24	1.00
$\lambda_{\text{Liaison=yes, [t]}} - \lambda_{\text{Liaison=no, [t]}} > 0$	1.89	0.08	1.76	2.02	1.00
$\lambda_{\text{Liaison=yes, [ʁ]}} - \lambda_{\text{Liaison=no, [ʁ]}} > 0$	0.44	0.14	0.21	0.68	1.00
$\lambda_{\text{Liaison=yes, [z]}} - \lambda_{\text{Liaison=no, [z]}} > 0$	4.88	0.13	4.67	5.09	1.00
$\lambda_{\text{Liaison=yes, [p]}} - \lambda_{\text{Liaison=no, [p]}} > 0$	4.79	1.51	2.83	7.56	1.00

Table 4: Inferential statistics for the token-frequency analysis (Study 1): difference between the λ parameters of the negative binomial regression for liaison and non-liaison words (estimate, estimated error and 90% Credibility Interval) and posterior probability that this difference is positive. A positive difference means that liaison words occur more frequently than non-liaison words in the corpus.

4.3 Discussion

Overall liaison words were found to have lower type frequency but higher token frequency than non-liaison words, after controlling for the effect of grammatical morphology and phonotactics on lexical frequency. This result is compatible with the prediction of the morphological analysis of liaison as allomorphy, allomorphy being typically limited to a relatively small number of high-frequency words.

There was one unexpected result however: liaison words ending in [t] were found to have *higher* type frequency than non-liaison words ending in [t] in Lexique 3.83. Also, while the average token frequency of words ending in [t] was found to be larger for liaison than for non-liaison words, liaison words outnumbered non-liaison words in the low-frequency range of the lexicon (see Figure 3).

One possible reason for this result is the failure to sufficiently control for morphological effects on lexical frequency. Study 1 controlled for the effect of *grammatical* morphology but not *lexical* morphology, as described in Section 4.1. For instance, words that share a derivational suffix, such as adjectives *appétiss-ant* 'appetizing' and *odor-ant* 'fragrant', were treated as independent lexical units when calculating word counts, on the assumption that derivational suffixes are not stored independently in the mental lexicon. The same applies to compounds such as *aigre-doux* 'sweet and sour', which were treated as independent lexical units from their constituents, i.e. *aigre* 'sour' and *doux* 'sweet'. However this assumption is probably an oversimplification, in particular for productive derivational suffixes. Suppose that derivational suffixes with a liaison form, such as *-ant* [ɑ̃]-[ɑ̃t], have low type frequency, as predicted by the morphological analysis of liaison, but are highly productive. Then there should be many derived words built using these suffixes (*appétiss-ant*, *odor-ant*, etc.), and these words should contribute to an increase of the type frequency of liaison words. Under these conditions, failure to adequately control for the effect of lexical morphology on word-type frequency will result in a spurious inflation of the number of liaison words.

This point can be made more concrete by showing how failing to control for grammatical morphology in Study 1 would change the relative type frequency of liaison and non-liaison words. Figure 4 is based on the same data as Figure 2 but treats inflected words sharing the same inflectional suffix as independent lexical units. Liaison words appear now to have much higher type frequency than non-liaison words for the three word-final consonants [ʁ], [t], and [z] that occur in inflectional suffixes (see the Appendix). This increase in type frequency for liaison words is entirely due to the failure to control for the morphological parenthood between inflected words sharing the same inflectional suffix.

This comparison demonstrates the importance to carefully control for morphological relations among words when testing the morphological analysis of French liaison. Study 1 only controlled for the effect of *grammatical* morphology (inflection) on lexical frequency. But the effect of *lexical* morphology (derivation, composition) should also be controlled for.

5 Controlling for the effect of lexical morphology: Study 2

Study 2 tests the predictions of the morphological analysis while controlling for the effect of lexical morphology on lexical frequency. To do this, the study focuses on a subset of the dataset used in Study 1 involving adjectival liaison. Adjectival liaison refers here to liaison alternations involving masculine singular adjectives. The main reason to focus on adjectival liaison is that masculine singular adjectives are the main case of *lexical* morphemes involved in liaison alternations. These alternations are observed in prenominal positions, as in (10). The corresponding adjectival morphemes are either free morphemes, like *grand* [gʁɑ̃] ∼ [gʁɑ̃t] in (10a), or bound morphemes, like the derivational suffix *-eux* [ø] ∼ [øz] in (10b).

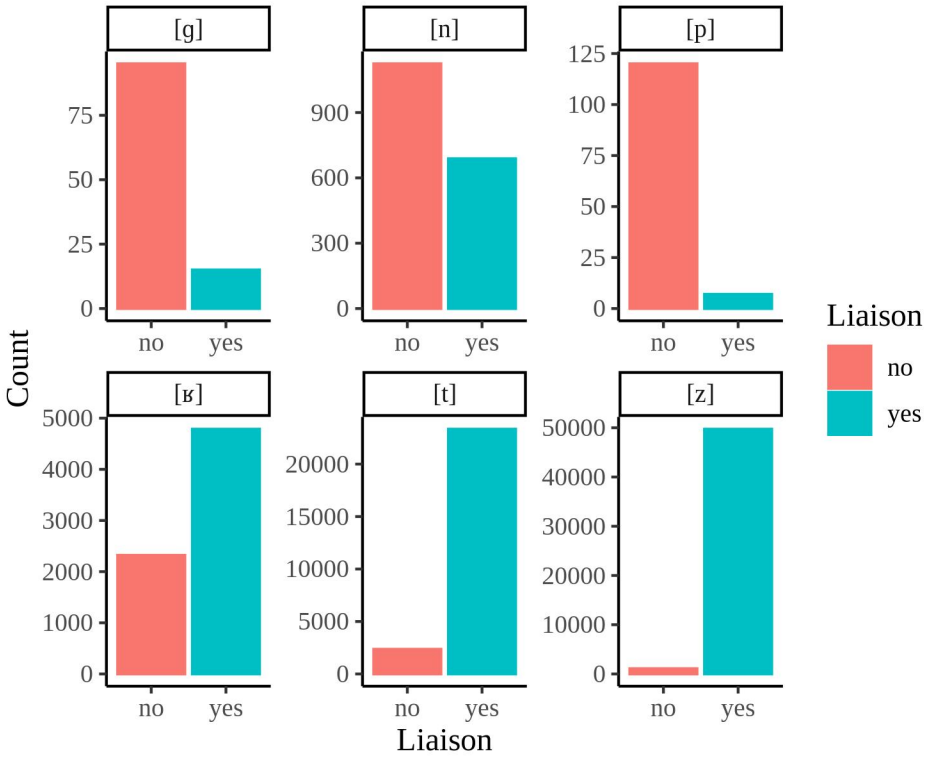


Figure 4: Count of liaison and non-liaison words in Lexique 3.83 as a function of the identity of the word-final consonant when grammatical morphology is not controlled for.

(10) Adjectival liaison (= liaison in masculine singular adjectives)

a. Free adjectival morphemes

	Word1	Word2	
grand monsieur	[gʁɑ̃]	[məsjø]	'great man'
grand ami	[gʁɑ̃t]	[ami]	'great friend'

b. Bound adjectival morphemes

	Word1	Word2	
heureux mariage	[øʁ-ø]	[maʁjaʒ]	'happy wedding'
heureux événement	[øʁ-øz]	[evɛnəmɑ̃]	'happy event'

The morphological analysis of liaison predicts that liaison adjectival morphemes such as (10) have lower type frequency but higher token frequency than non-liaison adjectival morphemes such as *chouette* 'nice'. As in Study 1, these hypotheses will be tested while controlling in addition for the effect of phonotactics on lexical frequency.

The question whether adjectival liaison has low type frequency has been touched upon before, but inconclusively, due to the absence of quantitative assessment. Bybee (2001:179-180) writes that adjectival liaison is limited to a small set of masculine singular adjectives (she lists only six adjectives). If true, this would clearly support the morphological analysis. However Morin (2005:11-12) notes that adjectival liaison is actually available for many more adjectives, in particular in the formal speech register (see also Côté 2011:Section 2.1.1). Study 2 will make it possible to settle this question using quantitative evidence.

Section 5.1 describes the dataset used in Study 2, explaining how adjectival morphemes were

segmented. Section 5.2 presents the results. Section 5.3 concludes with a discussion of the results. The dataset and R script are available in Storme (2024b) (`study2-data.csv` and `study2-script.R`, respectively).

5.1 Dataset and coding

5.1.1 Focusing on masculine singular adjectives

A subset of the dataset from Study 1 was used in Study 2, including only adjectives that are relevant to test hypotheses about adjectival liaison. Adjectival liaison involves masculine singular adjectives. Therefore unambiguously feminine adjectives such as *heureuse* [øʁøz] 'happy.SG.FEM' and unambiguously plural adjectives such as *grands* [gʁɑ̃] ∼ [gʁɑ̃z] 'great.PL.MASC' were discarded. Adjectives that are invariable for both gender and number (e.g., *marron* 'brown') were also discarded, as they are typically derived from singular nouns through conversion and therefore are not primarily adjectival morphemes.[6] However adjectives that are ambiguous as to their gender or as to their number (but not both) were not discarded. This includes masculine liaison adjectives with a liaison form ending in [z] (e.g., *heureux* 'happy'), which are number ambiguous (they use the same form in the singular and in the plural), and singular non-liaison adjectives with a graphic form ending in *-e* (e.g., *chouette*), which are gender ambiguous (they use the same form in the masculine and in the feminine). These adjectives were included to have as much data as posible. However it should be borne in mind that they do not have separate lexical entries for their different interpretations. Therefore the token frequencies provided in Lexique 3.83 for them *overestimate* their actual token frequencies as masculine singular adjectives. Note that this overestimation is expected to affect both liaison and non-liaison adjectives, as there are systematic ambiguities in both sets of adjectives (e.g., *heureux* can be used both in the singular and in the plural and *chouette* can be used both as a masculine and as a feminine adjective). This problem does not affect the predictions about *type* frequencies, as those are independent from token frequencies.

After this selection, the dataset of adjectival liaison was manually checked. Six words that were incorrectly identified as adjectives involving liaison were excluded. The resulting dataset includes 4357 adjectives. The Python and R scripts that were used to build the dataset for adjectival liaison can be found in Storme (2024b) under the name `1-extract-liaison.py` and `prepare-study2-data.R`.

There are restrictions on the type of adjectives that can occur prenominally in casual speech (Garrigues 1997) and therefore are susceptible to appear under a liaison form. However, as mentioned in the introduction to Section 5, these restrictions are much less stringent in the formal register, where virtually any adjective can appear prenominally (Morin 2005:11-12). Study 2 therefore adopts a maximally inclusive view of liaison adjectives by not discarding adjectives that typically do not appear prenominally in casual speech. This inclusive view of liaison allows for a more stringent test of the morphological analysis: if the type frequency of liaison adjectives is found to be smaller than that of non-liaison adjectives in this very inclusive dataset, then it should be even smaller in a more restricted dataset based on casual speech.

5.1.2 Morphological segmentation

The set of final morphemes occurring in the corpus was identified manually by a French native speaker (the author). This set contains both bound morphemes (e.g., suffixes like *-ais* in *français* 'French') and free morphemes (e.g., the adjective *gris* in the compound adjective *blanc-gris* 'white

[6] Singular nouns do not allow for liaison forms (Côté 2011:Section 2.2).

gray' or the adjective *fin* in the prefixed adjective *extra-fin* 'very fine'). A Python script was then used to automatically extract the final morpheme of each adjective in the corpus. This script is available in Storme (2024b) under the name `2-extract-final-adj-morpheme.py`. The segmentation procedure is described in more details in the paragraphs below.

Adjectives were not required to be segmentable into meaningful units in order to qualify as morphologically complex. For instance, the adjective *belliqueux* 'aggressive, warlike' was analyzed as containing a liaison suffix *-eux* [ø(z)] even though the stem *belliqu-* is not attested independently. In other words, the view of morphological segmentation that was adopted here is distributional rather than semantic.

Stems were assumed to contain at least one syllable (e.g., *pieux* [pjø] 'pious' was not analyzed as containing the suffix *-eux* added to a stem *pi-* [pj-]), with a few exceptions where a monosyllabic stem could be clearly identified (e.g., *boueux* [bwø] 'muddy' is based on *boue* [bu] 'mud').

A distributional definition of morphological segmentation was adopted mainly for convenience. It was often difficult to decide whether a word is morphologically segmentable into semantically transparent meaningful units. For instance, the adjective *capillaire* 'relative to hair' might be treated as morphologically simple, in the absence of an independent base corresponding to the sequence *capill-* and despite the presence of the suffix *-aire*. However *capill-* could also be treated as a bound allomorph of the noun *cheveu* 'hair', in which case *capillaire* becomes decomposable into meaningful units. In the absence of a comprehensive morphological database of French morphologically complex words,[7] a distributional criterion was applied: *capillaire* was analyzed as morphologically complex because it contains the independently attested adjectival suffix *-aire* (see *budgétaire*, *caniculaire*) and the resulting stem *capill-* is at least monosyllabic.

Note that these principles were not applied blindly, and some exceptions were set when an adjective was clearly not featuring a suffix. That was the case of the adjective *anti-calcaire*. This adjective is not derived from a base *calc-* using the adjectival suffix *-aire*, but from the noun *calcaire* through prefixation. That was also the case of the adjective *avant* (as in *la roue avant* 'the front wheel'). This adjective is not derived from a base *av-* using the adjectival suffix *-ant*, but from the preposition/adverb *avant* 'before' through conversion.

5.2 Results

5.2.1 Type-frequency analysis

Figure 5 and Table 5 show, in graphical and numerical form, respectively, the number of liaison and non-liaison adjectival morphemes in Lexique 3.83 as a function of the identity of the final consonant. There are no adjectival morphemes ending in [p] in Lexique 3.83, which explains why [p] is not discussed here.

Descriptively, liaison adjectival morphemes appear to be fewer than non-liaison adjectival morphemes in Lexique 3.83 and this appears to hold for each consonant, except for [z]. To check whether this generalization is statistically robust, a logistic regression was fit to the data, using the same methods as in Study 1 (see Section 4.2). The estimated log odds of liaison adjectival morphemes are shown in Table 6 for each consonant, along with the corresponding 90% Credibility Interval (CI), and the posterior probability that this log odds is negative. A negative sign for the estimated log odds of liaison adjectival morphemes means that liaison adjectival morphemes are

[7] A database called MorphoLex (Mailhot, Wilson, Macoir, Deacon & Sánchez-Gutiérrez 2020) was recently developed, based on the definition of the morpheme as the smallest meaning bearing unit. However it does not contain all the adjectives in the corpus. Also, it is not always consistent. For instance, *adjacent* is analyzed as complex (*ad-jac-ent*) even though it is not clearly decomposable into independently meaningful units.

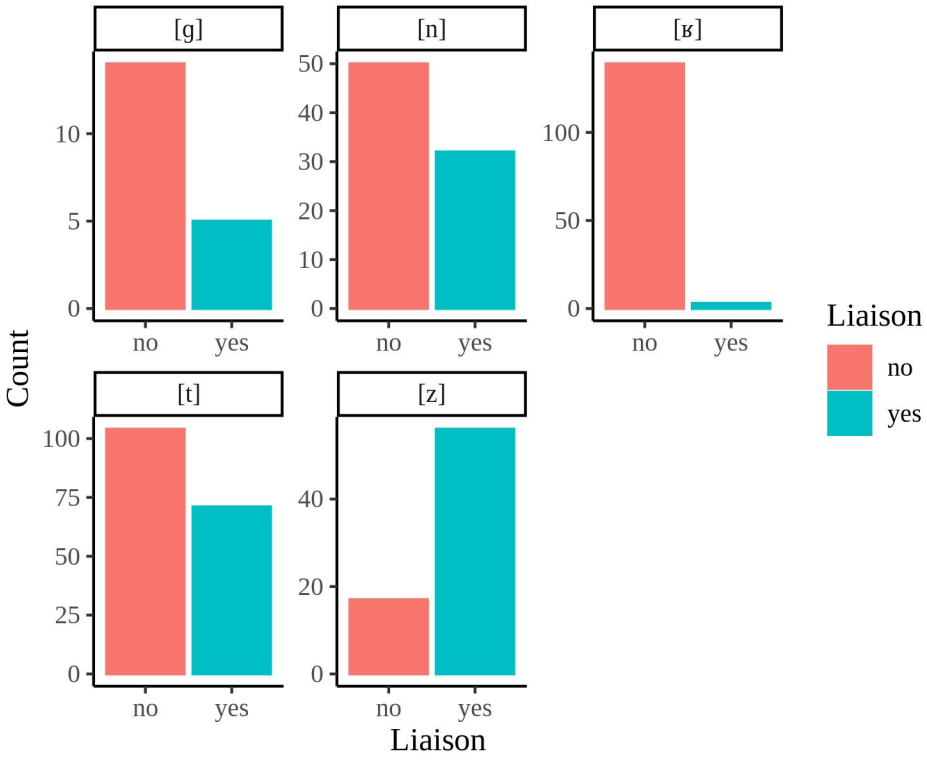


Figure 5: Descriptive statistics for the type-frequency analysis (Study 2): count of liaison and non-liaison adjectival morphemes in Lexique 3.83 as a function of the identity of the last consonant in the morpheme.

	Liaison		Sum
FinalConsonant	no	yes	
[g]	14	5	19
[n]	50	32	82
[ʁ]	139	3	142
[t]	104	71	175
[z]	17	56	73
Sum	324	167	491

Table 5: Descriptive statistics for the type-frequency analysis (Study 2): count of liaison and non-liaison adjectival morphemes in Lexique 3.83 as a function of the identity of the last consonant in the morpheme.

fewer than non-liaison adjectival morphemes in the lexicon, in line with the morphological analysis of liaison.

For all consonants except [z], the log odds of liaison adjectival morphemes was found to be negative with a posterior probability larger than 0.95, meaning that there is very strong evidence that liaison adjectival morphemes have *lower* type frequency than non-liaison adjectival morphemes. These results are in line with the predictions of the morphological analysis of liaison.

However, for adjectival morphemes ending in [z], the posterior probability of the hypothesis was found to equal 0, meaning that there is very strong evidence that liaison adjectival morphemes

Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Post.Prob
logodds(LIAISON=yes, [g]) < 0	-1.09	0.52	-1.97	-0.28	0.99
logodds(LIAISON=yes, [n]) < 0	-0.45	0.23	-0.83	-0.08	0.97
logodds(LIAISON=yes, [t]) < 0	-0.38	0.16	-0.64	-0.13	0.99
logodds(LIAISON=yes, [ʁ]) < 0	-3.99	0.62	-5.10	-3.08	1.00
logodds(LIAISON=yes, [z]) < 0	1.22	0.28	0.76	1.68	0.00

Table 6: Inferential statistics for the type-frequency analysis (Study 2): log odds of liaison adjectival morphemes for each consonant (estimate, estimated error and 90% Credibility Interval) and posterior probability that this log odds is negative. A negative log odds means that liaison adjectival morphemes are fewer than non-liaison adjectival morphemes in the lexicon.

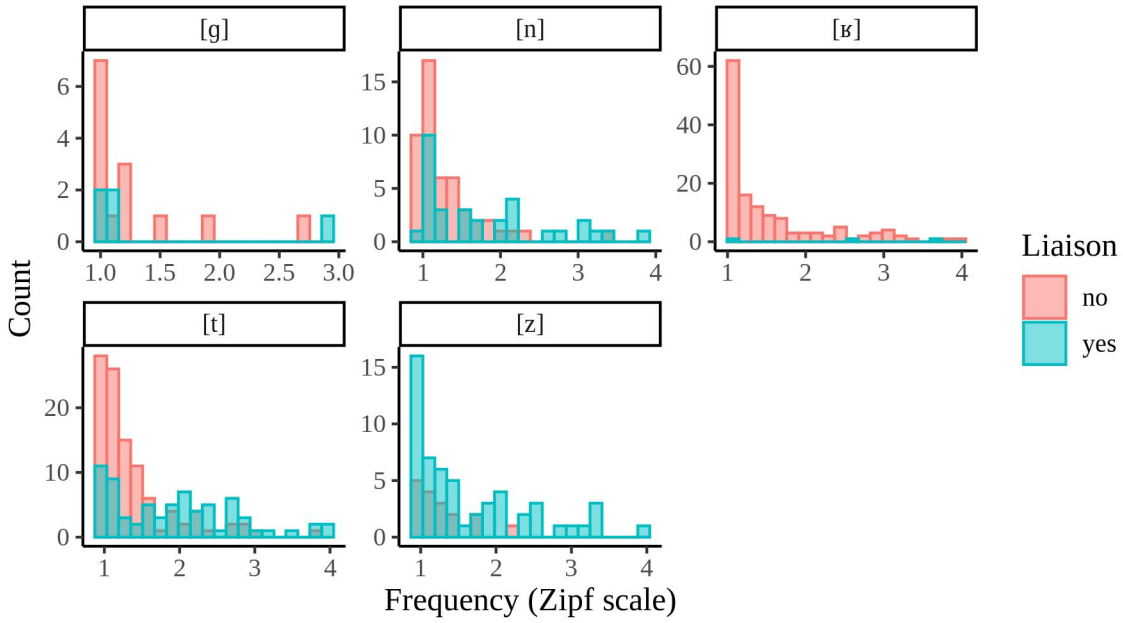


Figure 6: Descriptive statistics for the token-frequency analysis (Study 2): distribution of token frequencies (in Zipf; from 1=very rare to 7=very frequent) for liaison and non-liaison adjectival morphemes in the corpus of movie subtitles from Lexique 3.83 as a function of the identity of the last consonant in the morpheme. The frequency values on the x-axis were binned into 20 intervals.

ending in [z] *do not* have lower type frequency than non-liaison words ending in [z]. As shown in Table 6, the log odds of liaison adjectival morphemes was actually found to be positive for [z], meaning that liaison morphemes ending in [z] have *higher* type frequency than non-liaison morphemes ending in [z]. This unexpected result will be further discussed in Section 5.3.

5.2.2 Token-frequency analysis

Although liaison adjectives are fewer in the lexicon than non-liaison adjectives, they appear to be more frequently used in the corpus of movie subtitles, as shown in Figure 6 and Table 7.

Figure 6 shows the distribution of token frequencies for liaison and non-liaison adjectival morphemes in the corpus after controlling for phonotactics. Token frequencies (on the x-axis) are expressed using the Zipf scale (Van Heuven et al. 2014), as in Study 1. Figure 6 shows that, over-

	Liaison	
FinalConsonant	no	yes
[g]	458	1592
[n]	632	5333
[ʁ]	2588	15290
[t]	1177	6231
[z]	212	3873

Table 7: Descriptive statistics for the token-frequency analysis (Study 1): average token count of liaison and non-liaison adjectival morphemes in Lexique 3.83 as a function of the identity of the last consonant in the morpheme.

all, the frequency distribution of adjectival morphemes follows the typical power-law relationship observed by Zipf (1949), with many morphemes occurring rarely (with Zipf values below 4) and a small set of morphemes occurring frequently (with Zipf values above 4). Although liaison and non-liaison adjectival morphemes both follow this general trend, their frequency distributions are not identical: except for [z], liaison adjectival morphemes are less represented than non-liaison adjectival morphemes in the low-frequency range of the lexicon. This is in line with the hypothesis that liaison adjectival morphemes have higher token frequency than non-liaison adjectival morphemes.

The data in Table 7 are also in line with this hypothesis. Table 7 shows the average token count of liaison and non-liaison adjectival morphemes in the corpus of movie subtitles, by the identity of the final consonant. On average, liaison adjectival morphemes appear to be represented by a larger number of tokens in the corpus than non-liaison adjectival morphemes, in line with the prediction of the morphological analysis that liaison morphemes have higher token frequency than non-liaison morphemes. Note that this also holds for [z]. This might seem to contradict the observation that liaison words in [z] outnumber non-liaison words in [z] in the low frequency range of the lexicon (see Figure 6). However this is not necessarily a contradiction if liaison words in [z] also outnumber non-liaison words in [z] in the *high* frequency range of the lexicon.

To check whether the generalization that liaison adjectival morphemes occur on average more frequently than non-liaison adjectival morphemes is statistically robust, a negative binomial regression was fit to the word count data, using the same methods as in Study 1 (see Section 4.2). The difference between the λ parameters corresponding to liaison and non-liaison adjectival morphemes is shown in Table 8 for each consonant, along with the 90% credibility interval for that difference, and the posterior probability that this difference is positive. A positive sign means that λ for liaison adjectival morphemes is larger than λ for non-liaison adjectival morphemes, and therefore that liaison adjectival morphemes occur more frequently than non-liaison adjectival morphemes in the corpus, in line with the prediction of the morphological analysis of liaison.

The difference between the λ parameters of liaison and non-liaison adjectival morphemes was found to be positive for all consonants, with a posterior probability larger than 0.95, except for [g]. This means that there is strong evidence that liaison morphemes have higher token frequency on average than non-liaison morphemes. This result is in line with the predictions of the morphological analysis of liaison. For [g], the posterior probability of the hypothesis did not reach the 0.95 threshold (it is equal to 0.89). This null result can probably be explained by the fact that there is a small number of adjectives ending in [g] in Lexique 3.83 (there are only 19 such adjectives, as shown in Table 7).

Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Post.Prob
$\lambda_{\text{LIAISON=yes, [g]}} - \lambda_{\text{LIAISON=no, [g]}} > 0$	1.61	1.47	-0.52	4.20	0.89
$\lambda_{\text{LIAISON=yes, [n]}} - \lambda_{\text{LIAISON=no, [n]}} > 0$	2.16	0.55	1.30	3.11	1.00
$\lambda_{\text{LIAISON=yes, [t]}} - \lambda_{\text{LIAISON=no, [t]}} > 0$	1.67	0.37	1.07	2.28	1.00
$\lambda_{\text{LIAISON=yes, [ʁ]}} - \lambda_{\text{LIAISON=no, [ʁ]}} > 0$	2.92	2.10	0.42	6.89	0.98
$\lambda_{\text{LIAISON=yes, [z]}} - \lambda_{\text{LIAISON=no, [z]}} > 0$	2.79	0.68	1.59	3.81	1.00

Table 8: Inferential statistics for the token-frequency analysis (Study 2): difference between the λ parameters of the negative binomial regression for liaison and non-liaison adjectival morphemes (estimate, estimated error and 90% Credibility Interval) and posterior probability that this difference is positive. A positive difference means that liaison adjectival morphemes occur more frequently than non-liaison adjectival morphemes in the corpus.

5.3 Discussion

Overall liaison adjectives were found to have lower type frequency but higher token frequency than non-liaison adjectives, after controlling for the effect of lexical morphology and phonotactics on lexical frequency. This result is compatible with the prediction of the morphological analysis of liaison as allomorphy, allomorphy being typically limited to a relatively small number of high-frequency words.

There was one unexpected result however: liaison words ending in [z] were found to have *higher* type frequency than non-liaison words ending in [z] in Lexique 3.83. This result is unlikely to be due to a lack of morphological control, as lexical morphology was carefully controlled for in Study 2.

A possible avenue to consider is the difference in phonotactic markedness between word-final [z] and intervocalic [z], and how this difference might have affected the diachronic development of liaison and non-liaison words. Word-final voiced obstruents are phonotactically marked (Gordon 2016:Chapter 2). In French, this pattern is attested as a statistical tendency synchronically, and affects in particular fricatives such as [z] in prepausal positions (Jatteau, Vasilescu, Lamel, Adda-Decker & Audibert 2019). Diachronically, this bias against word-final [z] should result in a reinterpretation of [z]-final words as [s]-final words, and therefore in a decrease of the type frequency of non-liaison words ending in [z] over time.

On the other hand, voiceless consonants are phonotactically marked intervocalically (Gordon 2016:151). In French, intervocalic [s] was voiced intervocalically and became [z] in the history of French (Vaissière 1996), as can be seen by comparing French *rose* [ʁozə] and its Latin etymon *rosa* [rosa]. Liaison consonants often occur intervocalically (Storme 2024a), and this context should favor the voicing of liaison [s], resulting in an increase of the type frequency of liaison words ending in [z] over time. The diachronic voicing of the liaison consonant is reflected synchronically in [s]-[z] alternations in the paradigm of the adjective *gros* 'big': the liaison form of *gros* ends in [z], as in *le gros arbre* [lə#gʁoz#aʁbʁ] 'big tree', whereas the feminine form *grosse* ends in [s], as in *la grosse araignée* [la#gʁos#aʁeɲe] 'big spider'.

These two opposite effects, devoicing of final [z] and voicing of intervocalic [s], might explain the unexpected higher type frequency of liaison words ending in [z] as compared to non-liaison words ending in [z]. To explain why there is no similar asymmetry for the other liaison consonants, it must be assumed that the phonotactic markedness of the corresponding consonants was less dependent on the context.

Finally, it should also be borne in mind that the dataset of adjectival liaison used in Study

2 is very inclusive and treats as liaison adjectives adjectives that might never be attested in a liaison context in casual speech. As mentioned in Section 5.1.1, adjectival liaison is only observed prenominally but not all adjectives are allowed to appear in this position in spoken French. This means that the present study might overestimate the number of actual liaison adjectives, including for [z]. However it is striking that, for five of the six liaison consonants, the number of liaison adjectival morphemes is smaller than that of corresponding non-liaison morphemes, despite the very inclusive definition of liaison words adopted in this paper.

6 Testing the morphological analysis against monomorphemic words: Study 3

In Study 2, the effect of lexical morphology on word frequency was controlled for by focusing on a subset of the French lexicon including only adjectival morphemes. Study 3 further controls for the effect of morphology by only considering words that consist of a single morpheme. Section 6.1 describes the dataset of monomorphemic words used in Study 3. Section 6.2 presents the results. Section 6.3 concludes with a brief discussion.

6.1 Dataset and coding

Monomorphemic words were extracted from Lexique 3.83 using a list from Mahowald, Dautriche, Gibson & Piantadosi (2018). This list contains 6,728 words identified as monomorphemic by a French native speaker.[8] These words are listed under their phonological form. However, for testing the morphological analysis of liaison, the graphic form is also needed. To obtain this, Lexique 3.83 was filtered to retain only words whose phonological forms are included in the list of monomorphemic words from Mahowald et al. (2018). The resulting lexicon included both monomorphemic words (e.g., *beau* 'beau') and their polymorphemic homophones (e.g., *beau-x* 'beautiful-PL'). Polymorphemic homophones were then removed by discarding all words ending with the grammatical suffixes listed in the Appendix. Additionally, words that do not end in a consonant involved in liaison alternations and singular nouns were removed, as in Study 1 and Study 2. The final list contained 1,460 words.

It should be noted that words that are identified as monomorphemic in Study 3 were not necessarily identified as such in Study 2. For instance, *mondain* 'socialite' is treated as polymorphemic (*mond-ain*) in Study 2 but as monomorphemic in Study 3. The reason for this discrepancy lies in the use of two distinct definitions of what constitutes a morpheme in the two studies. In Study 2, a distributional definition of the morpheme was adopted (see Section 5.1.2). According to this definition, *mondain* is identified as bimorphemic because the suffix *-ain* can be identified in other words (*métropolitain*, *napolitain*, etc.). By contrast, the list of monomorphemic words in Mahowald et al. (2018) is based on the definition of the morpheme as the smallest meaningful unit. Under this definition, *mondain* 'socialite' can be analyzed as monomorphemic in so far as its meaning does not reflect the meaning of the base *monde* 'world'.

The differing definitions of a morpheme used in Study 2 and Study 3 should not be viewed as problematic. Instead, this variation allows for testing the morphological hypothesis under different theoretical assumptions. If the predictions of the morphological hypothesis are consistent regardless of the definition of a morpheme, it strengthens the evidence supporting the hypothesis.

[8] I am grateful to Isabelle Dautriche for giving me access to this dataset.

6.2 Results

6.2.1 Type-frequency analysis

Figure 7 and Table 9 show, in graphical and numerical form, respectively, the number of liaison and non-liaison monomorphemic words in Lexique 3.83 as a function of the identity of the final consonant.

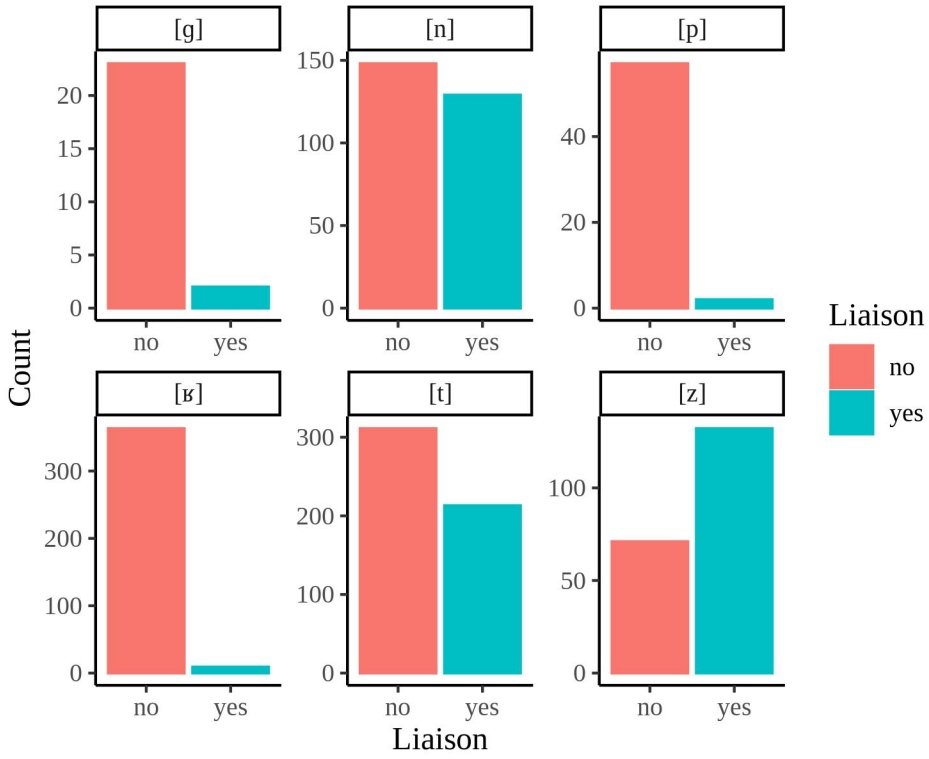


Figure 7: Descriptive statistics for the type-frequency analysis (Study 3): count of liaison and non-liaison monomorphemic words in Lexique 3.83 as a function of the identity of the word-final consonant.

	Liaison		Sum
FinalConsonant	no	yes	
[g]	23	2	25
[n]	148	129	277
[p]	57	2	59
[ʁ]	363	9	372
[t]	311	213	524
[z]	71	132	203
Sum	973	487	1460

Table 9: Descriptive statistics for the type-frequency analysis (Study 3): count of liaison and non-liaison monomorphemic words in Lexique 3.83 as a function of the identity of the word-final consonant.

Descriptively, among monomorphemic words, liaison words appear to be fewer than non-liaison words, and this appears to hold for each consonant, except for [z]. Note that [z] was also found to be the only consonant that did not conform to the predictions of the morphological analysis in Study 2.

To check whether this generalization is statistically robust, a logistic regression was fit to the data, using the same methods as in Studies 1 and 2. The results are shown in Table 10. The log odds of liaison words were found to be negative for all consonants except [z], meaning that liaison words are generally fewer in the lexicon than non-liaison words for these consonants. However, for [n], the posterior probability of the morphological hypothesis did not reach 0.95. As a consequence, the evidence in favor of the morphological analysis is not particularly strong for words ending in [n]. For [t], [ʁ], [p] and [g], the posterior probability was found to equal 1, meaning that the evidence in favor of the morphological hypothesis is very strong for words ending in these consonants. As in Study 2, for words ending in [z], the evidence goes against the predictions of the morphological analysis, as there are more liaison words than non-liaison words ending in this consonant (the log odds of liaison words is positive). This problematic result will be discussed in Section 6.3.

Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Post.Prob
logodds(LIAISON=yes, [g]) < 0	-2.66	0.82	-4.15	-1.46	1.00
logodds(LIAISON=yes, [n]) < 0	-0.14	0.12	-0.34	0.06	0.87
logodds(LIAISON=yes, [t]) < 0	-0.38	0.09	-0.53	-0.23	1.00
logodds(LIAISON=yes, [ʁ]) < 0	-3.75	0.35	-4.34	-3.22	1.00
logodds(LIAISON=yes, [z]) < 0	0.62	0.15	0.39	0.86	0.00
logodds(LIAISON=yes, [p]) < 0	-3.60	0.79	-5.07	-2.46	1.00

Table 10: Inferential statistics for the type-frequency analysis (Study 3): log odds of liaison monomorphemic words for each consonant (estimate, estimated error and 90% Credibility Interval) and posterior probability that this log odds is negative. A negative log odds means that liaison monomorphemic words are fewer than non-liaison monomorphemic words in the lexicon.

6.2.2 Token-frequency analysis

Figure 8 shows the distribution of token frequencies for liaison and non-liaison monomorphemic words in the corpus of movie subtitles after controlling for phonotactics. Token frequencies (on the x-axis) are expressed using the Zipf scale (Van Heuven et al. 2014), as in Study 1 and Study 2. For monomorphemic words endings in [t] and [n], liaison words are clearly less represented among the low-frequency range than non-liaison words, as expected under the morphological analysis. For [z], the situation is a bit more complex: liaison words are more represented in the low-frequency range of the lexicon than non-liaison words, but also in the high-frequency range. Among monomorphemic words ending with the other consonants ([p], [ʁ], and [g]), there are very few liaison words, so the results are less clear. Table 11 shows the average token count of liaison and non-liaison monormorphemic words in the corpus of movie subtitles, by the identity of the final consonant. On average, liaison monomorphemic words appear to be represented by a larger number of tokens in the corpus than non-liaison monomorphemic words, as expected under the morphological analysis, except for words ending in [ʁ].

To check whether this generalization is statistically robust, a negative binomial regression was fit to the word count data, using the same methods as in Study 1 (see Section 4.2). The difference between the λ parameters corresponding to liaison and non-liaison monomorphemic words is shown

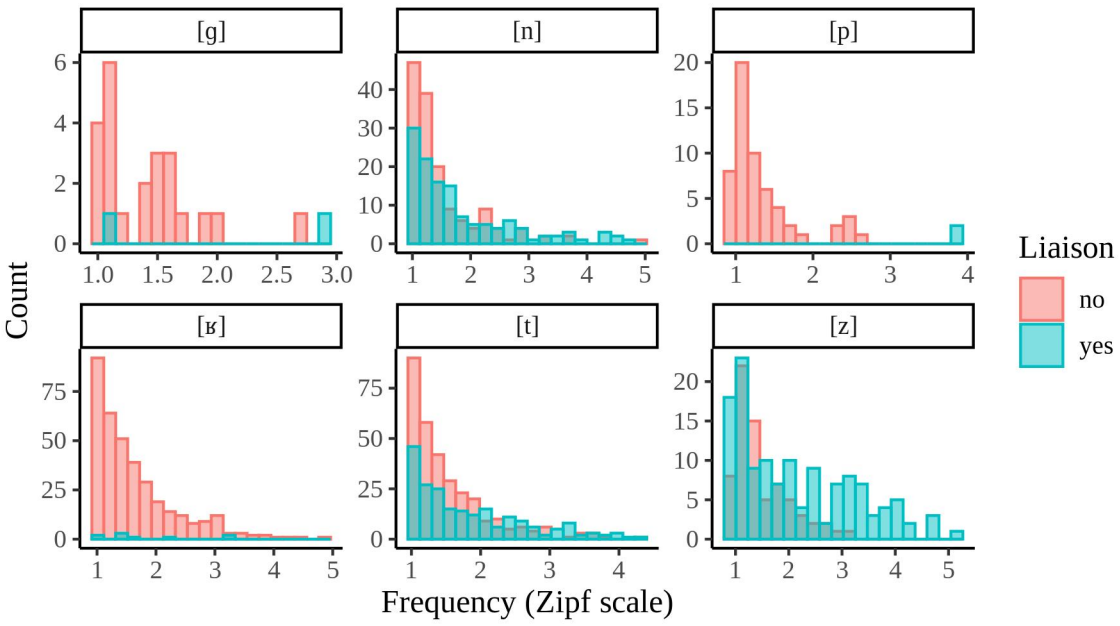


Figure 8: Descriptive statistics for the token-frequency analysis (Study 3): distribution of token frequencies (in Zipf; from 1=very rare to 7=very frequent) for liaison and non-liaison monomorphemic words in the corpus of movie subtitles from Lexique 3.83 as a function of the identity of the word-final consonant. The frequency values on the x-axis were binned into 20 intervals.

in Table 8 for each consonant, along with the 90% credibility interval for that difference, and the posterior probability that this difference is positive.

The difference between the λ parameters of liaison and non-liaison monomorphemic words was found to be positive for all consonants, with a posterior probability larger than 0.95, except for [ʁ]. This means that, for five out of the six liaison consonants, there is strong evidence that liaison monomorphemic words have higher token frequency on average than non-liaison monomorphemic words. This result is in line with the predictions of the morphological analysis of liaison. For [ʁ], the posterior probability of the hypothesis is equal to 0.38, meaning that liaison and non-liaison words ending in this consonant have similar token frequencies on average.

	Liaison	
FinalConsonant	no	yes
[g]	400	3979
[n]	6829	17141
[p]	356	74277
[ʁ]	5379	3424
[t]	2199	5975
[z]	693	36519

Table 11: Descriptive statistics for the token-frequency analysis (Study 3): average token count of liaison and non-liaison monomorphemic words in Lexique 3.83 as a function of the identity of the word-final consonant.

Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Post.Prob
$\lambda_{\text{LIAISON=yes, [g]}} - \lambda_{\text{LIAISON=no, [g]}} > 0$	3.71	2.59	0.55	8.70	0.98
$\lambda_{\text{LIAISON=yes, [n]}} - \lambda_{\text{LIAISON=no, [n]}} > 0$	0.92	0.26	0.49	1.36	1.00
$\lambda_{\text{LIAISON=yes, [t]}} - \lambda_{\text{LIAISON=no, [t]}} > 0$	1.00	0.19	0.69	1.32	1.00
$\lambda_{\text{LIAISON=yes, [ʁ]}} - \lambda_{\text{LIAISON=no, [ʁ]}} > 0$	-0.14	0.84	-1.33	1.36	0.38
$\lambda_{\text{LIAISON=yes, [z]}} - \lambda_{\text{LIAISON=no, [z]}} > 0$	3.95	0.33	3.41	4.48	1.00
$\lambda_{\text{LIAISON=yes, [p]}} - \lambda_{\text{LIAISON=no, [p]}} > 0$	6.90	2.54	3.84	11.83	1.00

Table 12: Inferential statistics for the token-frequency analysis (Study 3): difference between the λ parameters of the negative binomial regression for liaison and non-liaison monomorphemic words (estimate, estimated error and 90% Credibility Interval) and posterior probability that this difference is positive. A positive difference means that liaison monomorphemic words occur more frequently than non-liaison mnomorphemic words in the corpus.

6.3 Discussion

The results are generally compatible with the predictions of the morphological hypothesis, with liaison monomorphemic words having lower type frequency but higher token frequency than non-liaison words. There were two exceptions. Liaison words ending in [z] were found to have higher type frequency that non-liaison words ending in the same consonant. The same result was obtained in Study 2. The same phonotactic explanation as discussed in Section 5.3 could be proposed here. The other exception concerns monomorphemic words ending in [ʁ]: liaison words ending in [ʁ] were not found to have higher token frequency than non-liaison words ending in [ʁ]. Note though that the number of monomorphemic liaison words ending in [ʁ] is quite small (9), and this might explain why there is no statistically robust asymmetry.

7 Conclusion

Liaison is a hotly debated topic in French linguistics, with two opposing views about the underlying status of this alternation. According to the phonological view, liaison is an alternation between allophones of a single phoneme. According to the morphological view, liaison is an alternation between allomorphs of a single morpheme.

This paper has contributed to this debate by proposing a new kind of evidence to arbitrate between the two analyses. This evidence, based on lexical statistics, favors the morphological analysis overall. First, liaison does not pattern like a typical phonological deletion process, as the frequency of the liaison form does not decrease but increases with increasing lexical frequency. Second, liaison alternations pattern overall like allomorphy as they affect a comparatively small number of morphemes with comparatively high token frequency.

This latter result was obtained in three studies that compared the type and token frequencies of liaison words and non-liaison words while controlling for phonotactics (i.e. the identity of the final consonant) and for various morphological factors (presence of inflectional suffixes in Study 1, presence of lexical morphemes in Study 2, and morphological complexity in Study 2). In these three studies, liaison words were generally found to have lower type frequency but higher token frequency than non-liaison words, as expected under the morphological analysis. This result is particularly remarkable as these studies adopted a highly inclusive view of liaison. The number of words that are actually liaison words (i.e. may appear under their liaison forms) is likely to be even smaller in

a corpus that reflects more accurately the use of liaison alternations in spontaneous speech.

To the author's knowledge, the argument provided in this paper in favor of the morphological analysis is one of the few clearly empirical arguments that have been proposed in the debates on French liaison. Arguments have mainly been based on theoretical economy. There have been other empirical arguments, but they are mostly inconclusive. An empirical argument in favor of the phonological analysis was proposed by Tranel (1990, 2000), based on the observation that liaison consonants pattern differently from non-liaison consonants in their realization. However, as explained in Section 3.3, this pattern can be derived as a paradigm uniformity effect in the morphological analysis. An empirical argument in favor of the morphological analysis was also proposed by Bürki et al. (2015), based on the observation that the processing of liaison is sensitive to non-local properties of the context. However, as explained in Section 3.3, it is unclear that non-locality can be used as an argument for the morphological analysis.

The test proposed in this paper could be applied to other types of alternations beyond French liaison. Another obvious case in French is the [ə]-[ɛ] alternation. This alternation has been analyzed as allophonic by Dell (1985:198-214), with /ə/ being the phoneme and [ɛ] an allophone limited to closed syllables. However this alternation is phonologically less minimal than other mid-vowel alternations conditioned by syllable structure in French (e.g., [e]-[ɛ], [ø]-[œ], [o]-[ɔ]; see Eychenne 2014; Storme 2017, 2021), as it involves two feature changes (rounding and height). One possibility would be that [ɛ]-[ə] actually involves allomorphy whereas other mid-vowels alternations are truly allophonic. This hypothesis could be tested using the same kind of lexical evidence as in this paper. If the [ə]-[ɛ] alternation is allomorphic but the other mid-vowel alternations are allophonic, then words with [ə]-[ɛ] alternations should have lower type frequency but higher token frequency than words featuring the other mid-vowel alternations.

Appendix

Grammatical suffixes in Study 1 were identified using a combination of phonological information and morphological information from Lexique 3.83. The morphological information is provided in three columns in Lexique 3.83: part of speech can be found under `cgram`, gender under `genre`, and number under `nombre`.

In all non-verbal parts of speech (nouns, adjectives, pronouns, and determiners) and in past participles, there is only one grammatical suffix: the plural morpheme. Other grammatical features (e.g., gender) are not expressed by a suffix but by stem alternations (e.g., *beau/belle* 'beautiful.MASC/FEM' [bo]/[bɛl], *le/la* 'the.MASC/FEM' [lə]/[la], *ce/cette* 'this.MASC/FEM' [sə]/[sɛt], etc.). The plural morpheme is a liaison morpheme that alternates between [z] and ∅. This alternation is illustrated in (11). All words ending in liaison [z] which belong to the aforementioned parts of speech and are listed as plural in Lexique 3.83 (i.e. the column `nombre` has value `p`) were identified as featuring a plural suffix.

(11) Plural suffix

a. In nouns

	Word1	Word2	
fruits frais	[fʁɥi]	[fʁɛ]	'fresh fruits'
fruits exotiques	[fʁɥi-z]	[egzotik]	'tropical fruits'

b. In adjectives

	Word1	Word2	
grandes femmes	[gʁɑ̃d]	[fam]	'great women'
grandes amies	[gʁɑ̃d-z]	[ami]	'great friends.FEM'

c. In pronouns

	Word1	Word2	
elles sont	[ɛl]	[sɔ̃]	'they.FEM are'
elles ont	[ɛl-z]	[ɔ̃]	'they.FEM have'

d. In determiners

	Word1	Word2	
les Français	[le]	[fʁɑ̃sɛ]	'the French'
les Allemands	[le-z]	[almɑ̃]	'the Germans'

e. In past participles

	Word1	Word2	
envoyés vers	[ɑ̃vwaje]	[vɛʁ]	'sent towards.'
envoyés à	[ɑ̃vwaje-z]	[a]	'sent to.'

Verbal grammatical morphology is more diverse, but remains limited to a small number of consonant-final suffixes: third person, non-third person, present participle, and infinitive. Most of these suffixes are also involved in liaison alternations. The third person is expressed in most conjugations by a liaison suffix alternating between ∅ and [t], as illustrated in (12a).[9] Non-third persons (i.e. first and second person) are expressed in most conjugations by a liaison suffix alternating between ∅ and [z], as illustrated in (12b).[10] The present participle is expressed by a liaison suffix alternating between [ɑ̃] and [ɑ̃t], as illustrated in (12c). The infinitive is marked by a liaison suffix alternating between [e] and [eʁ] in the first conjugation group, as illustrated in (12d), and by a non-liaison suffix [ʁ] in the second and third conjugation groups, as illustrated in (12e). These suffixes were identified using the column `cgram` in Lexique 3.83.

(12) Verbal suffixes

a. Third-person suffix ∅ ∼ [t]

	Word1	Word2	
Il veut deux cafés.	[vø]	[dø]	'He wants two coffees.'
Il veut un café.	[vø-t]	[ɛ̃]	'He wants a coffee.'

b. Non-third-person suffix ∅ ∼ [z]

	Word1	Word2	
Je veux deux cafés.	[vø]	[dø]	'I want two coffees.'
Je veux un café.	[vø-z]	[ɛ̃]	'I want a coffee.'

c. Present participle suffix [ɑ̃] ∼ [ɑ̃t]

	Word1	Word2	
voulant deux cafés.	[vulɑ̃]	[dø]	'wanting two coffees.'
voulant un café.	[vulɑ̃t]	[ɛ̃]	'wanting a coffee.'

d. Infinitive suffix [e] ∼ [eʁ] (first conjugation group)

[9] This suffix is not used in all cases. For instance, the third person is not expressed by any overt marker in the present tense for verbs of the first conjugation group (e.g., *il chante* [ʃɑ̃t] 'he sings').

[10] This suffix is not used in all cases. For instance, the first person singular is not expressed by any overt marker in the present tense for verbs of the first conjugation group (e.g., *je chante* [ʃɑ̃t] 'I sing').

	Word1	Word2	
acheter deux cafés	[aʃəte]	[dø]	'to buy two coffees'
acheter un café	[aʃəte-ʁ]	[ɛ̃]	'to buy a coffee'

e. Infinitive suffix [ʁ] (second and third conjugation groups)

	Word1	Word2	
boire deux cafés	[bwa-ʁ]	[dø]	'to drink two coffees'
boire un café	[bwa-ʁ]	[ɛ̃]	'to drink a coffee'

References

Applegate, Richard. 1972. *Ineseño Chumash grammar*. Berkeley, CA: University of California PhD thesis.

Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech* 47(1). 31–56.

Bell, Alan, Jason M Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111. https://doi.org/10.1016/j.jml.2008.06.003.

Berg, Thomas. 2011. A diachronic frequency account of the allomorphy of some grammatical markers. *Journal of Linguistics* 47(1). 31–64. https://doi.org/10.1017/S002222671000023X.

Boersma, Paul & Jan-Willem van Leussen. 2017. Efficient evaluation and learning in multilevel parallel constraint grammars. *Linguistic Inquiry* 48(3). 349–388. https://doi.org/10.1162/ling_a_00247.

Bonami, Olivier. 2011. French liaison: shape alternation, inflection, or both? `http://www.llf.cnrs.fr/sites/llf.cnrs.fr/files/biblio/Surrey-liaison-slides.pdf`. Accessed: 2023-09-08.

Bonami, Olivier & Gilles Boyé. 2005. Construire le paradigme d'un adjectif. *Recherches Linguistiques de Vincennes* 34. 77–98. https://doi.org/10.4000/rlv.1370.

Bonami, Olivier, Gilles Boyé & Jesse Tseng. 2005. Sur la grammaire des consonnes latentes. *Langages* 158. 89–100. https://doi.org/10.3917/lang.158.0089.

Bonami, Olivier & Élizabeth Delais-Roussarie. 2021. La liaison. In Anne Abeillé & Danièle Godard (eds.), *La grande grammaire du français*, 2109–2218. Arles: Actes Sud.

Braver, Aaron. 2019. Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology* 36(1). 1–36. https://doi.org/10.1017/S0952675719000022.

Bürki, Audrey, Ulrich H Frauenfelder & F-Xavier Alario. 2015. On the resolution of phonological constraints in spoken production: Acoustic and response time evidence. *Journal of the Acoustical Society of America* 138(4). EL429–EL434. https://doi.org/10.1121/1.4934179.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. https://doi.org/10.18637/jss.v080.i01.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, Joan L. 2001. Frequency effects on French liaison. In Joan L. Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 337–359. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.17byb.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301571.001.0001.

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(3). 557–596. https://doi.org/10.1017/S0305000908009124.

Clements, George N. & Samuel Jay Keyser. 1983. *CV Phonology: A generative theory of the syllable*. Cambridge, MA: The MIT Press.

Coetzee, Andries W. & Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language & Linguistic Theory* 31(1). 47–89. https://doi.org/10.1007/s11049-012-9179.

Côté, Marie-Hélène. 2011. French liaison. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, vol. 5, 1–26. Malden, MA: Wiley Blackwell. https://doi.org/10.1002/9781444335262.wbctp0112.

Côté, Marie-Hélène. 2014. Liaison et assibilation en français laurentien. In Christiane Soum-Favaro, Annelise Coquillon & Jean-Pierre Chevrot (eds.), *La liaison : Approches contemporaines*, 9–32. Berne: Peter Lang. https://doi.org/10.3726/978-3-0352-0204-5.

Dell, François. 1985. *Les règles et les sons*. Paris: Hermann. Second edition.

Dolatian, Hossep. 2023. Output-conditioned and non-local allomorphy in Armenian theme vowels. *The Linguistic Review* 40(1). 1–42. https://doi.org/10.1515/tlr-2022-2104.

Encrevé, Pierre. 1988. *La liaison avec et sans enchaînement: Phonologie tridimensionnelle et usages du français*. Paris: Seuil.

Eychenne, Julien. 2011. La liaison en français et la théorie de l'optimalité. *Langue française* 169. 79–101. https://doi.org/10.3917/lf.169.0079.

Eychenne, Julien. 2014. Schwa and the loi de position in Southern French. *Journal of French Language Studies* 24(2). 223–253. https://doi.org/10.1017/S0959269513000173.

Eychenne, Julien. 2019. On the deletion of word-final schwa in Southern French. *Phonology* 36(3). 355–389. https://doi.org/10.1017/S0952675719000198.

File-Muriel, Richard J. 2010. Lexical frequency as a scalar variable in explaining variation. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 55(1). 1–25. https://doi.org/10.1353/cjl.0.0065.

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. https://aclanthology.org/2001.jeptalnrecital-long.15/.

Gaatone, David. 1978. Forme sous-jacente unique ou liste d'allomorphes? (à propos des consonnes de liaison en français). *Linguistics* 214. 33–54. https://doi.org/10.1515/ling.1978.16.214.33.

Garrigues, Mylène. 1997. Une méthode de désambiguïsation locale nom/adjectif pour l'analyse automatique de textes. *Langages* 31(126). 60–78. https://www.persee.fr/doc/lgge_0458-726x_1997_num_31_126_1777.

Gaskell, M. Gareth, Elsa Spinelli & Fanny Meunier. 2002. Perception of resyllabification in French. *Memory & Cognition* 30(5). 798–810.

Goeman, Antonie. 1999. *T-deletie in Nederlandse dialecten*: Vrije Universiteit dissertation.

Gordon, Matthew K. 2016. *Phonological typology*, vol. 1. Oxford: Oxford University Press.

Haspelmath, Martin & Andrea Sims. 2010. *Understanding morphology*. London: Routledge. https://doi.org/10.4324/9780203776506.

Inkelas, Sharon. 2014. *The interplay of morphology and phonology*. Oxford: Oxford University Press.

Jatteau, Adèle, Ioana Vasilescu, Lori Lamel, Martine Adda-Decker & Nicolas Audibert. 2019. "Gra[f]e!" Word-final devoicing of obstruents in Standard French: An

acoustic study based on large corpora. In *Proceedings Interspeech 2019*, 1726–1730. https://doi.org/10.21437/Interspeech.2019-2329.

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language* 45. 229–254. https://doi.org/10.1075/tsl.45.13jur.

Kilbourn-Ceron, Oriana. 2017. *Speech production planning affects variation in external sandhi*: McGill University dissertation.

Klausenburger, Jurgen. 1984. *French liaison and linguistic theory*. Wiesbaden: Steiner.

Kruschke, John K. & Torrin M. Liddell. 2018. Bayesian data analysis for newcomers. *Psychonomic bulletin & review* 25(1). 155–177. https://doi.org/10.3758/s13423-017-1272-1.

Leminen, Alina, Miika Leminen, Minna Lehtonen, Päivi Nevalainen, Sari Ylinen, Lilli Kimppa, Christian Sannemann, Jyrki P Mäkelä & Teija Kujala. 2011. Spatiotemporal dynamics of the processing of spoken inflected and derived words: A combined EEG and MEG study. *Frontiers in Human Neuroscience* 5. 66. https://doi.org/10.3389/fnhum.2011.00066.

Leminen, Alina, Eva Smolka, Jon A. Duñabeitia & Christos Pliatsikas. 2019. Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex* 116. 4–44. https://doi.org/10.1016/j.cortex.2018.08.016.

Levshina, Natalia. 2022. Frequency, informativity and word length: Insights from typologically diverse corpora. *Entropy* 24(2). https://doi.org/10.3390/e24020280.

Mahowald, Kyle, Isabelle Dautriche, Edward Gibson & Steven T. Piantadosi. 2018. Word forms are structured for efficient use. *Cognitive Science* 42(8). 3116–3134. https://doi.org/https://doi.org/10.1111/cogs.12689.

Mailhot, Hugo, Maximiliano A Wilson, Joël Macoir, S Hélène Deacon & Claudia Sánchez-Gutiérrez. 2020. MorphoLex-FR: A derivational morphological database for 38,840 French words. *Behavior Research Methods* 52. 1008–1025. https://doi.org/10.3758/s13428-019-01297-z.

Malécot, André. 1974. Frequency of occurrence of French phonemes and consonant clusters. *Phonetica* 29(3). 158–170. https://doi.org/10.1159/000259468.

Morin, Yves-Charles. 1983. De la (dé)nasalisation et de la marque de genre en français. *Lingua* 61. 133–156. https://doi.org/10.1016/0024-3841(83)90030-X.

Morin, Yves-Charles. 2003. Remarks on prenominal liaison consonants in French. In Stefan Ploch (ed.), *Living on the Edge. 28 papers in honour of Jonathan Kay*, 385–400. Berlin: de Gruyter. https://doi.org/10.1515/9783110890563.385.

Morin, Yves Charles. 2005. La liaison relève-t-elle d'une tendance à éviter les hiatus? Réflexions sur son évolution historique. *Langages* 158(2). 8–23. https://doi.org/10.3917/lang.158.0008.

Morley, Rebecca L. 2018. Is phonological consonant epenthesis possible? A series of artificial grammar learning experiments. *Phonology* 35(4). 649–688. https://doi.org/10.1017/S0952675718000234.

New, Boris, Marc Brysbaert, Jean Veronis & Christophe Pallier. 2007. The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28. 661–677. https://doi.org/10.1017/S014271640707035X.

New, Boris & Christophe Pallier. 2023. Lexique. `http://www.lexique.org/`.

New, Boris, Christophe Pallier, Marc Brysbaert & Ludovic Ferrand. 2004. Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers* 36(3). 516–524. https://doi.org/10.3758/BF03195598.

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 Pilar Prieto, Joan Mascaró & Marie-Josep Solé (eds.), *Segmental and prosodic issues*

in Romance phonology, 3–23. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.282.03ngu.

R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. 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. https://doi.org/10.1016/j.wocn.2014.01.002.

Schane, Sanford. 1968. *French phonology and morphology*. Cambridge, MA: MIT Press.

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. https://roa.rutgers.edu/content/article/files/1552_smolensky_1.pdf.

Smolensky, Paul, Eric Rosen & Matthew Goldrick. 2020. Learning a gradient grammar of French liaison. In Hyunah Baek, Chikako Takahashi & Alex Hong-Lun Yeung (eds.), *Proceedings of the 2019 annual meeting on phonology*, 1–12. Washington, DC: Linguistic Society of America. https://doi.org/10.3765/amp.v8i0.4680.

Spinelli, Elsa, Anne Cutler & James M. McQueen. 2002. Resolution of liaison for lexical access in French. *Revue française de linguistique appliquée* 7. 83–96. https://doi.org/10.3917/rfla.071.0083.

Spinelli, Elsa, James M. McQueen & Anne Cutler. 2003. Processing resyllabified words in French. *Journal of Memory and Language* 48(2). 233–254. https://doi.org/10.1016/S0749-596X(02)00513-2.

Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison. In Jean-Marc Authier, Barbara E. Bullock & Lisa A. Reed (eds.), *Formal perspectives in Romance linguistics. Selected papers from the 28th linguistic symposium on Romance languages*, 243–270. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.185.18ste.

Storme, Benjamin. 2017. The *loi de position* and the acoustics of French mid vowels. *Glossa: a journal of general linguistics* 2. https://doi.org/10.5334/gjgl.300.

Storme, Benjamin. 2021. Evidence for suffix cohesion in french. *Journal of French Language Studies* 31(1). 75–103. https://doi.org/10.1017/S0959269520000198.

Storme, Benjamin. 2024a. French liaison and hiatus avoidance. *Radical: A journal of phonology* 6. 1–30. https://radical.cnrs.fr/wp-content/uploads/2024/08/Storme_2024.pdf.

Storme, Benjamin. 2024b. French liaison is allomorphy, not allophony: Evidence from lexical statistics (Data and code). https://doi.org/10.5281/zenodo.12733801.

Storme, Benjamin. 2024c. Paradigm uniformity effects on French liaison. *Natural Language and Linguistic Theory* https://doi.org/10.1007/s11049-023-09596-z. (Advance online publication).

Tessier, Anne-Michelle & Karen Jesney. 2021. Learning French liaison with gradient symbolic representations: Errors, predictions, consequences. In Ryan Bennett, Richard Bibbs, Mykel Loren Brinkerhoff, Max J. Kaplan, Stephanie Rich, Amanda Rysling, Nicholas Van Handel & Maya Wax Cavallaro (eds.), *Proceedings of the 2020 Annual Meeting on Phonology*, https://doi.org/10.3765/amp.v9i0.4940.

Tranel, Bernard. 1981. *Concreteness in generative phonology. Evidence from French*. Berkeley and Los Angeles: University of California Press.

Tranel, Bernard. 1987. French schwa and nonlinear phonology. *Linguistics* 25(5). 845–866. https://doi.org/doi:10.1515/ling.1987.25.5.845.

Tranel, Bernard. 1990. On suppletion and French liaison. *Probus* 2. 87–104. https://doi.org/10.1515/prbs.1990.2.2.169.

Tranel, Bernard. 1995. French final consonants and nonlinear phonology. *Lingua* 95(1-3). 131–167. https://doi.org/10.1016/0024-3841(95)90104-3.

Tranel, Bernard. 2000. Aspects de la phonologie du français et la théorie de l'optimalité. *Langue Française* 126. 39–72.

Vaissière, Jacqueline. 1996. From Latin to Modern French: On diachronic changes and synchronic variations. *AIPUK, Arbetisberitche, Institut für Phonetik und digitale Sprachverarbeitung, Universität Kiel* (31). 61–74.

Van Heuven, Walter JB, Pawel Mandera, Emmanuel Keuleers & Marc Brysbaert. 2014. SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology* 67(6). 1176–1190. https://doi.org/10.1080/17470218.2013.85052.

Winter, Bodo & Paul-Christian Bürkner. 2021. Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and Linguistics Compass* 15(11). e12439. https://doi.org/https://doi.org/10.1111/lnc3.12439.

Zipf, George Kingsley. 1929. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 40. 1–95. https://doi.org/10.2307/310585.

Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort.* Cambridge, MA: Addison-Wesley Press.