Not only size matters: limits to the Law of Three Consonants in French phonology

Benjamin Storme
Université de Lausanne
Sciences du Langage et de l'Information
benjamin.storme@unil.ch

Word count: 10,365

Abstract Grammont's Law of Three Consonants (LTC) states that French schwa is obligatorily pronounced in any CC_C sequence to avoid three-consonant clusters. Although schwa presence has been shown to be sensitive not only to cluster size but also to the nature of consonants in post-lexical phonology, the LTC is still considered as accurate to describe schwa-zero alternations in lexical phonology. The paper uses judgment data from French speakers in France and Switzerland to compare the behavior of schwa in derived words (lexical phonology) and inflected words (post-lexical phonology). The results show that schwa-zero alternations are conditioned not only by cluster size but also by cluster type in lexical phonology. Moreover, the same phonotactic asymmetries among consonant clusters are found in lexical and post-lexical phonologies. The data therefore support a weaker version of the lexical-phonology hypothesis than what is usually assumed for French. Lexical and post-lexical phonologies do not require different phonotactic constraints but only different weights for the same constraints.

Keywords: phonology; morphology; French; schwa; inflection/derivation; judgment data; probabilistic constraint-based grammars

1 Introduction

In French, some morphemes alternate between a form with schwa and a form without schwa. For instance, the noun *demande* 'request' can be realized with a schwa as [dəmɑ̃d] or without schwa as [dmɑ̃d]. Determining the factors that condition the distribution of schwa-zero alternations has been a central topic in French phonology for more than a century. Table 1 provides a non-exhaustive list of the variables that have been reported to play a role in this alternation along with a non-exhaustive list of the sources that document these effects. This table builds largely but not exclusively on Bürki et al. (2011: 3982-3985).

Among these variables, the consonantal context, and in particular the number of consonants surrounding schwa, has received particular attention early on. In his influential treaty on French pronunciation, Grammont (1914: 115-116) states that a preconsonantal schwa is obligatory when preceded by two consonants (CC_C), as illustrated in (1a), but excluded when preceded by a single consonant (C_C), as illustrated in (1b). He calls this generalization the 'loi des trois consonnes' (Law of Three Consonants; LTC) and explains it as a strategy to avoid three-consonant clusters. In (1a), the schwa form is preferred because it makes it possible to avoid the three-consonant cluster [ndm]. In (1b), the schwa-less form is preferred in the absence of three-consonant clusters.

Variables		Source
Segmental variables	Number of surrounding consonants	Bürki et al. (2011)
	Nature of surrounding consonants	Côté (2001); Bürki et al. (2011)
Morphological variables	Grammatical function of following suffix	Dell (1978); Côté (2001)
Prosodic variables	Position in word	Bürki et al. (2011)
	Position wrt prosodic boundaries	Dell (1977); Côté (2001)
	Word position in utterance	Bürki et al. (2011)
	Size of prosodic constituent	Côté (2007)
	Speech rate	Malécot (1976); Bürki et al. (2011)
Lexical variables	Word frequency	Dell (1985); Racine & Grosjean (2002)
	Word identity	Bürki et al. (2011)
Speaker variables	French variety	Gess et al. (2012)
	Speaker identity	Bürki et al. (2011)

Table 1: A non-exhaustive list of variables reported to condition schwa-zero alternations in French.

(1) Grammont's Law of Three Consonants (LTC)

a. Schwa is obligatory in CC_C
  C#C_C [tdm] sept demandes [sɛtdəmɑ̃d] 'seven requests'

b. Schwa is excluded in C_C
  C_C [dm] la demande [ladmɑ̃d] 'the request'

Subsequent works on French schwa have provided a more nuanced view of the LTC. First, the LTC has been found to hold as a gradient rather than a categorical generalization: schwa is not obligatory in CC_C and excluded in C_C but more *likely* in CC_C overall than in C_C (Bürki et al. 2011; Racine & Andreassen 2012; Côté 2012; Hambye & Simon 2012; Hansen 2012). Second, not only the number but also the nature and order of surrounding consonants has been found to be relevant. In C_C, clusters with increasing sonority favor the schwa-less form (Bürki et al. 2011). In CC_C, schwa is more likely to be pronounced if the middle consonant is a stop than if it is a fricative (see Côté 2001: 119 and earlier references therein), as illustrated in (2). Also, schwa is more likely to be pronounced if its absence implies that an obstruent-liquid cluster (OL) is not directly followed by a vowel (Dell 1976; 1985; Côté 2001), as illustrated in (3).

(2) a. Schwa is more likely in CS_C (S=stop)
  C#S_C [tdm] sept demandes [sɛt#d(ə)mɑ̃d] 'seven requests'

b. Schwa is less likely in CF_C (F=fricative)
  C#F_C [tfn] sept fenêtres [sɛt#f(ə)nɛtʁ] 'seven windows'

(3) a. Schwa is more likely in OL_C (O=obstruent, L=liquid)
  O#L_C [kls] chaque leçon [ʃak#l(ə)sɔ̃] 'each lesson'

b. Schwa is less likely in CO_L (O=obstruent, L=liquid)
  C#O_L [spl] douce pelouse [dus#p(ə)luz] 'sweet lawn'

This nuanced view of the LTC has been argued to be relevant in a range of contexts beyond the word-initial context illustrated in (2) and (3): at word boundaries (e.g. *act(e) pénible* 'painful act'), at clitic boundaries (e.g. *Annick l(e) salue* 'Annick greets him'), and at morpheme boundaries before the inflectional future/conditional suffix *-r-* (e.g. *je gard(e)-rai* 'I will keep'; Côté 2001: 85). However, there is one morphological context where the LTC is still considered to hold as a categorical generalization in line with Grammont's strict interpretation in (1). Between stems ending in two consonants and consonant-initial derivational suffixes (CC_-$\text{C}_{\text{derivation}}$), schwa is reported to be categorically pronounced and this regardless of the nature and order of surrounding consonants (Dell 1978; Côté 2001: 85, 109; Côté 2012: 258-259). For instance, schwa is reported to be obligatory in both OL_C and CO_L before derivational suffixes, as illustrated in (4a) and (4b), even though OL_C and CO_L are treated differently in general, as illustrated in (3) for word-initial contexts and in (5) before the inflectional future/conditional suffix *-r-*.

(4) a. Schwa is obligatory in OL_-C$_{\text{derivation}}$ (O=obstruent, L=liquid)
OL_-C$_{\text{derivation}}$ [glm] règlement [ʁɛglə-mɑ̃] ‘regulation’
b. Schwa is obligatory in CO_-L$_{\text{derivation}}$ (O=obstruent, L=liquid)
CO_-L$_{\text{derivation}}$ [ʁdʁ] garderie [gaʁdə-ʁi] ‘Kindergarten’

(5) a. Schwa is more likely in OL_-C$_{\text{inflection}}$ (O=obstruent, L=liquid)
OL_-C$_{\text{inflection}}$ [glʁ] règlera [ʁɛgl(ə)-ʁa] ‘adjust-FUT.3SG’
b. Schwa is less likely in CO_-L$_{\text{inflection}}$ (O=obstruent, L=liquid)
CO_-L$_{\text{inflection}}$ [ʁdʁ] gardera [gaʁd(ə)-ʁa] ‘keep-FUT.3SG’

The view according to which boundaries between stems and derivational suffixes are special compared to word boundaries and to boundaries between stems and inflectional suffixes implies that the phonological grammar may differ quite substantially across strata, with clearly distinct lexical and post-lexical strata.[1] In the post-lexical stratum, a set of phonotactic constraints referencing different types of three-consonant clusters (e.g. *OLC, *COL, etc) would be active, resulting in different patterns of schwa-zero alternations for different types of three-consonant clusters at word boundaries and at stem-suffix boundaries in inflected words, as illustrated in (2), (3), and (5). In the lexical stratum, only a single phonotactic constraint banning three-consonant clusters would be active (*CCC),[2] resulting in a single pattern of schwa-zero alternations for all CC_C sequences at stem-suffix boundaries in derived words, as illustrated in (4).

The hypothesis of clearly distinct lexical and post-lexical phonologies will be referred to as the strong version of the lexical-phonology hypothesis. This paper will test this hypothesis by focusing on (i) obstruent-liquid-consonant sequences (OL_C) and liquid-obstruent-liquid sequences (LO_L) as examples of three-consonant clusters and (ii) derivational and inflectional suffixes as examples of lexical and post-lexical contexts. OL_C and LO_L were chosen because they are reported to favor schwa presence and absence, respectively, as illustrated in (3) and (5). Derivation and inflection were chosen because they both involve suffixation and therefore form a minimal pair for the lexical/post-lexical distinction.

The predictions of the strong version of the lexical-phonology hypothesis are summarized in (6) for the two relevant consonant sequences (OL_C and LO_L) and the two relevant morphological contexts (derivation and inflection). The predictions are summarized in (6a) at the level of the grammar (constraint set in each stratum) and in (6b) at the level of the data (probability distribution of schwa-zero alternations).

(6) Strong version of the lexical-phonology hypothesis: predictions
a. Grammar: phonotactic constraints differ in derivation (lexical stratum) and in inflection (post-lexical stratum)
$\{*\text{CCC}_{\text{derivation}}\}$
$\{*\text{OLC}_{\text{inflection}}, *\text{LOL}_{\text{inflection}}\}$
b. Data: schwa-zero alternations are sensitive to cluster type only in inflection (post-lexical stratum)
$P(\text{ə}|\text{OL\_-C}_{\text{derivation}}) = P(\text{ə}|\text{LO\_-L}_{\text{derivation}})$
$P(\text{ə}|\text{OL\_-C}_{\text{inflection}}) \neq P(\text{ə}|\text{LO\_-L}_{\text{inflection}})$

[1] Following Dell (1978: 7-8), derived words are assumed to belong to the lexical stratum (derived words are lexemes) and inflected words to the post-lexical stratum (inflected words are specific morphosyntactic instantiations of lexemes).

[2] Or equivalently markedness constraints referencing different three-consonant clusters would have different weights in the post-lexical stratum but exactly the same weights in the lexical stratum.

The strong version of the lexical-phonology hypothesis treating the LTC as categorical in the lexicon seems to be assumed in the literature, at least in Dell (1985) and Côté (2001). It was tested by Côté (2012) in a corpus study of Laurentian French: in this corpus, she found no exception to the LTC under its categorical version in the lexicon (Côté 2012: 258). However, as will be further discussed in section 2, the corpus used in this study is probably too small to draw strong conclusions regarding the categorical nature of the LTC at stem-suffix boundaries.

Alternatively, schwa-zero alternations could be gradient and sensitive to cluster type in both lexical and post-lexical strata but closer to categorical in the lexical stratum. According to this view, the same phonotactic constraints against various types of three-consonant clusters would be active in both strata but constraint weights would be more similar (and higher) in the lexical stratum. As a result, deviations from categoricity and differences among different types of three-consonant clusters would be harder to detect in the lexical stratum. This hypothesis will be referred to as the weak version of the lexical-phonology hypothesis. Its predictions are summarized in (7), with (7a) focusing on the predictions at the level of the grammar and (7b) on the predictions at the level of the data.

(7) Weak version of the lexical-phonology hypothesis: predictions

a. Grammar: phonotactic constraints are the same but their weights differ in derivation (lexical stratum) and in inflection (post-lexical stratum)
$\{$*OLC$_{\text{derivation}}$, *LOL$_{\text{derivation}}\}$
$\{$*OLC$_{\text{inflection}}$, *LOL$_{\text{inflection}}\}$

b. Data: schwa-zero alternations are sensitive to cluster type in both derivation (lexical stratum) and inflection (post-lexical stratum)
$P(\text{ə}|\text{OL\_-C}_{\text{derivation}}) \neq P(\text{ə}|\text{LO\_-L}_{\text{derivation}})$
$P(\text{ə}|\text{OL\_-C}_{\text{inflection}}) \neq P(\text{ə}|\text{LO\_-L}_{\text{inflection}})$

The main goal of this paper is to tease apart these two versions of the lexical-phonology hypothesis. This question also has theoretical implications beyond French. Some theories assume that phonotactic asymmetries ultimately reflect perceptual and articulatory asymmetries (e.g. Kawasaki-Fukumori 1992; Flemming 2002) or sonority-driven asymmetries among segments (Clements 1990). According to these theories, the same phonotactic asymmetries should be reflected across the grammar's strata if the same perceptual/articulatory or sonority-driven asymmetries among segments hold across these strata. Under the default assumption that segmental properties are largely independent from morphosyntactic context (e.g. word boundaries, stem-suffix boundaries), these theories of phonotactics are more directly compatible with the weak version of the lexical-phonology hypothesis in (7).

This point can be illustrated more concretely by considering explanations that have been proposed for asymmetries among three-consonant clusters in French. The phonotactic constraints that drive French schwa-zero alternations in three-consonant sequences have been argued to ultimately have perceptual and sonority-driven motivations (Côté 2001: 137-152). For instance, Côté proposed that schwa is more likely to appear after a medial stop (CS_C) than after a medial fricative (CF_C), as illustrated in (2), because stops have weaker perceptual cues than fricatives to signal place of articulation and therefore are more in need of vocalic support. She also proposed that schwa is more likely to occur in OL_C than in CO_L, as illustrated in (3) and (5), because, in the absence of schwa, OL_C features a local sonority peak (the medial liquid) that does not correspond to a syllable peak, in violation of the sonority sequencing principle. This problem is solved if schwa is pronounced in OL_C. By contrast, CO_L already satisfies the sonority sequencing principle in the absence of schwa and therefore schwa presence is not as crucial. In both cases (CF_C vs. CS_C and OL_C vs. CO_L), the explanation refers to phonetic and phonological asymmetries among segments. Because these segmental asymmetries are expected to hold across the grammar's strata (e.g. a stop should have weaker internal cues than a fricative regardless of morphosyntactic context), then the same phonotactic asymmetries among clusters should be observed across strata, in line with the weak version of the lexical-phonology hypothesis in (7).

In addition to the strong and weak lexical-phonology hypotheses, two further hypotheses will be tested in this paper. As was just mentioned, obstruent-liquid clusters are reported to be more strongly avoided before consonants than before vowels, due to the sonority sequencing principle (Dell 1976; Côté 2001). Moreover, three-consonant clusters are reported to be more strongly avoided than two-consonant clusters (Bürki et al. 2011). In other words, this means that schwa should be more likely to occur in OL_C than in LO_L and in LO_L than in C_C. The predictions of this hypothesis on cluster markedness are summarized in (8a) at the level of the grammar and in (8b) at the level of the data.

(8) Hypothesis about cluster markedness
  a. Grammar:
  $w(\text{*OLC}) > w(\text{*LOL}) > w(\text{*CC})$
  b. Data:
  $P(\text{ə}|\text{OL\_C}) > P(\text{ə}|\text{LO\_L}) > P(\text{ə}|\text{C\_C})$

Finally, schwa is reported to be more likely to appear in CC_C when the consonant cluster crosses the boundary between a stem and a derivational suffix than when it crosses the boundary between a stem and an inflectional suffix (Dell 1978; Côté 2001; 2012). However, no such asymmetry between derivation and inflection is reported for two-consonant clusters. For instance, Côté (2001: 85) describes schwa as excluded (or at least unlikely) in C_C in both derived and inflected words, without reporting any difference in schwa likelihood in the two cases (e.g. *fruiterie* 'fruit store' and *je gâterai* 'I will spoil'). If this is correct, this means that a phonotactic constraint referencing a three-consonant cluster should have a larger weight (w) at the boundary between a stem and a derivational suffix (lexical stratum) than at the boundary between a stem and an inflectional suffix (post-lexical stratum). This asymmetry would not extend to two-consonant clusters. The predictions of this hypothesis are summarized in (9a) at the level of the grammar (constraint weights) and in (9b) at the level of the data.

(9) Hypothesis about the effect of suffix's function on cluster markedness
  a. Grammar: phonotactic constraints banning three-consonant clusters have larger weights at derivational than at inflectional boundaries, but this does not extend to phonotactic constraints against two-consonant clusters
  $w(\text{*}C_1C_2C_{3\text{derivation}}) > w(\text{*}C_1C_2C_{3\text{inflection}})$
  $w(\text{*}C_1C_{2\text{derivation}}) = w(\text{*}C_1C_{2\text{inflection}})$
  b. Data: schwa is more likely to break a three-consonant cluster at derivational than at inflectional boundaries, but this does not extend to two-consonant clusters
  $P(\text{ə}|C_1C_2\_\text{-}C_{3\text{derivation}}) > P(\text{ə}|C_1C_2\_\text{-}C_{3\text{inflection}})$
  $P(\text{ə}|C_1\_\text{-}C_{2\text{derivation}}) = P(\text{ə}|C_1\_\text{-}C_{2\text{inflection}})$

The hypothesis according to which three-consonant clusters are more marked in lexical phonology than in post-lexical phonology is compatible with the observation that three-consonant clusters are more easily tolerated at the edges of higher morphosyntactic domains in post-lexical phonology in French (Dell 1977). For instance, schwa is more likely to occur in [st_d] between a noun and its complement within the same noun phrase, as illustrated in (10a), than at the boundary between two phrases, as illustrated in (10b) (Dell 1977: 151).

(10) a. Schwa is more likely between a noun (N) and a prepositional phrase (PP)
Elle met [la$_{\mathrm{D}}$ list([ə])$_{\mathrm{N}}$ d'artistes$_{\mathrm{PP}}$]$_{\mathrm{NP}}$ dans sa poche.
'She puts the list of artists in her pocket.'
b. Schwa is less likely between a noun phrase (NP) and a prepositional phrase (PP)
Elle [met [la liste d'artist([ə])]$_{\mathrm{NP}}$ [dans sa poche]$_{\mathrm{PP}}$]$_{\mathrm{VP}}$.
'She puts the list of artists in her pocket.'

Côté (2001: 129-132) proposed a prosodic characterization for the kind of asymmetries illustrated in (10a) and (10b), with schwa being less likely at the edge of higher prosodic domains due to strengthening and lengthening effects in these positions. Strengthening and lengthening of consonants at the edge of high prosodic domains make the presence of schwa more superfluous for the sake of consonant identification (Côté 2001: 146-151). The prosodic analysis accounts for the asymmetry in (10a) and (10b) under the reasonable assumption that the schwa in (10a) occurs inside a phonological phrase (*la liste d'artistes*) whereas the schwa in (10b) occurs between two phonological phrases (*la liste d'artistes* and *dans la poche*). However this analysis does not extend to the asymmetry between derivation and inflection, as there is no prosodic boundary below the word level in French. If derivational suffixes are found to favor schwa presence more than inflectional suffixes in CC_C at the stem-suffix boundary, as hypothesized in (9), this means that domain effects cannot be all reduced to prosody and that there are genuine effects of morphosyntactic domains on schwa-zero alternations, as originally formulated by Dell (1977).

The rest of the paper is organized as follows. Section 2 motivates the use of judgment data for testing these three hypotheses and presents the study's methods. Section 3 presents the results, focusing in turn on the statistical analysis of the judgment data and on the grammatical modeling with probabilistic constraint-based grammars. Section 4 discusses the results. Section 5 concludes.

2 Methods

2.1 Judgment task

The present study uses speakers' metalinguistic judgments as primary data, following a long tradition in linguistics (Schütze & Sprouse 2013; Schütze 2016; Myers 2017) and in the study of French schwa (Dell 1985; Côté 2001; Racine & Grosjean 2002; Racine 2007). However, because many recent studies on French schwa are based on speech corpora instead of judgments (e.g. Bürki et al. 2011; Racine & Andreassen 2012; Côté 2012; Eychenne 2019), the use of judgment data may require some justification. First, corpus data suffer from a problem of data sparsity that is expected to be particularly acute in the present case. Indeed, the hypotheses tested in this study bear on words that feature very specific morphological and phonological properties, namely inflected and derived words with OL_C and LO_L clusters at stem-suffix boundaries. Corpora would probably need to be very large in order to feature enough occurrences of the relevant words. For instance, the PFC corpus for Laurentian French used in Côté (2012) contain 2,530 contexts for schwa but only three instances of CC_C at the boundary between a stem and the inflectional future/conditional suffix *-r-* (Côté 2012: 258). Second, to the author's knowledge, available corpora of French speech do not provide the morphological information necessary to readily test the hypotheses that are central in the present study. In particular, the PFC corpus does not provide information about whether a word-internal schwa is morpheme-internal or at a morpheme boundary (see Racine et al. 2016 for a description of the variables that were coded for schwa in PFC).

The present study's design was inspired by a previous study by Racine (2007; 2008) that used metalinguistic judgments to estimate the likelihood of words' schwa forms and schwa-less forms in French (see also Racine & Grosjean 2002: 312-313). More specifically, participants were asked to rate how likely they would be to pronounce schwa variants and schwa-less variants for a set of

115 words. The task was slightly different from that used by Racine. In the present study, the task corresponds to a judgment of relative frequency whereas participants in Racine's study were asked to rate the absolute frequency of each variant independently. A judgment of relative frequency was used because it makes it possible to directly obtain the information most relevant for the research question of interest, namely the estimated relative frequency of the two variants. In Racine's work, an extra-step is needed to calculate the relative frequency of the two variants from their individual frequencies (see Racine 2007: 127). Following Racine (2007), the judgments were elicitated using a seven-point Likert scale, with 1 indicating a categorical preference for the schwa variant (e.g. *garderie*), 7 indicating a categorical preference for the schwa-less variant (e.g. *gard'rie*), and 4 indicating no preference for either form. An example is shown in Figure 1.

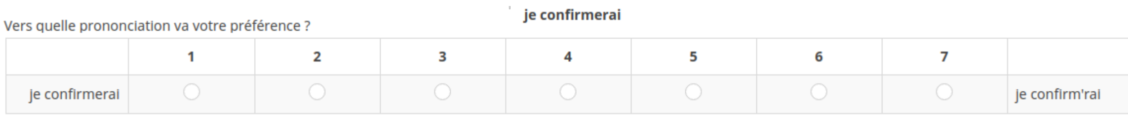

Figure 1: Judgment task.

2.2 Participants

21 Swiss French speakers (8 females, 13 males; recruited among students at a Swiss university) and 34 French speakers from France (27 females, 7 males; recruited online via the CNRS's platform RISC) participated in the study online, using the LimeSurvey platform (LimeSurvey 2012). Based on previous research, French speakers from Switzerland are expected to rate higher schwa-less variants as compared to speakers from France (Racine 2007). Participants from both origins were tested to control whether this difference in the baseline rate of schwa production interferes with the LTC. The participants provided their informed consent to participate in the research and agreed to make their data available online. No sensitive information about participants was collected.

2.3 Experimental items and fillers

Two variables were manipulated to construct the experimental items: Cluster (with three levels: OL_C, LO_L, C_C) and Morphology (with two levels: derivation, inflection). C_C stands for any two-consonant cluster, LO_L for liquid-obstruent-liquid clusters, and OL_C for obstruent-liquid-consonant cluster. Inflected words all featured the future suffix *-r-* because this suffix is to the author's knowledge the only consonant-initial inflectional suffix in French. Inflected words were all presented with a subject pronoun preceding them (e.g. *je chanterai* 'I will sing') to ensure that they were correctly identified as inflected words. Derived words were presented without any additional information (e.g. *garderie*).

Four of the six experimental conditions included 15 words whereas the two remaining ones included 14 words.[3] There was therefore a total of 88 experimental items in the study. Table 2 illustrates each condition using items that were featured in the stimulus set. 27 filler items were used in addition. The fillers featured schwa in morpheme-internal position, mostly in the first syllable of words (e.g. *chemin* 'path'). Experimental items and fillers used in the study are listed in the paper's Appendix.

For each word, the schwa variant was conveyed using the word's graphic form (e.g. *garderie*). The graphic form always contains an *e* corresponding to the schwa phone [ə]. The schwa-less

[3] This difference is due to an error when typing the stimuli in the online platform. However this error is not problematic because the statistical analysis does not require the same number of observations per condition.

		Suffix derivational	inflectional
Cluster	C_C	*biscuit-e-rie* (14)	*chant-e-rai* (15)
	LO_L	*concierg-e-rie* (15)	*gard-e-rons* (15)
	OL_C	*souffl-e-rie* (14)	*règl-e-ra* (15)

Table 2: Experimental items.

variant was conveyed by replacing the *e* by the apostrophe (e.g. *gard'rie*). The order of presentation of the experimental items and fillers was randomized.

2.4 Data analyses

While it is common practise to analyze ordinal data such as Likert-scale data as metric variables using linear regression, Liddell & Kruschke (2018) show that this can lead to a number of errors, including false alarms (i.e. detecting an effect that is not real), misses (i.e. failure to detect real effects), and even inversions of effects (i.e. the order of the means according to the metric scale is opposite to the true ordering of the means). One of the major problem with this metric approach is that it assumes that the distance between the response categories is equidistant whereas this might not be necessarily the case. For instance, in the case of a seven-point Likert scale, the distance between 5 and 6 might be treated differently from the distance between 2 and 3 in the participant's mind, even though the two distances are both equal to one on a metric scale. To avoid these issues, Liddell & Kruschke (2018) recommend to analyze Likert-scale data using ordinal instead of linear regression models.

In this paper, the judgment data were modeled using the ordinal cumulative model (Bürkner & Vuorre 2019: 78-79). The cumulative model assumes that the observed ordinal response variable derives from the categorization of a latent continuous unobserved variable. In the present study, the ordinal variable is the rating of the preference for the schwa or schwa-less variant along the seven-point scale. The latent variable is the participant's underlying opinion about the relative frequency of the two variants. To model this categorization in the case of a seven-point Likert scale, the cumulative model assumes that there are six thresholds which partition the latent variant variable into seven ordered categories (1, 2, ..., 6, 7). The model provides estimates both for the different conditions' means along the latent continuous variable and the position of the six thresholds. The reader is referred to Bürkner & Vuorre (2019) for further details.

The analysis of the judgment data was also supplemented with a linguistic analysis using probabilistic constraint-based grammars. This additional analysis is motivated by the fact that we ultimately care about the linguistic system that underlies participants' behavior. And this linguistic system can be characterized as a constraint-based grammar. A constraint-based analysis is used because, as noted by Durand & Laks (2000: 32), constraints provide a very intuitive interpretation of the LTC as caused by a general markedness constraint *CCC banning three-consonant clusters. Also recent theoretical papers have modeled schwa-zero alternations using probabilistic constraint-based grammars (Bayles et al. 2016; Smith & Pater 2020). Among the different families of probabilistic constraint-based grammars, MaxEnt was chosen (Hayes & Wilson 2008) because it is easy to implement and has been shown to provide a good fit to linguistic data compared to alternative frameworks (e.g. Smith & Pater 2020).[4]

[4] MaxEnt has been argued to make problematic typological predictions compared to other frameworks for probabilistic constraint-based grammars (Anttila & Magri 2018). However because this paper is not directly concerned with typology, this issue will be left aside.

A Bayesian approach was adopted (rather than a frequentist approach) for inferring the parameters of both the ordinal regression and the probabilistic grammars. This choice was motivated by the fact that Bayesian inference yields outcomes that are intuitive and easy to interpret. In particular, it provides a posterior distribution for all the model's parameters and combinations of parameter values given the data. This makes it very easy to test any hypothesis about the parameter values and about differences between parameter values. Also, Bayesian approaches virtually always converge to accurate values of the parameters (Liddell & Kruschke 2018).

3 Results

3.1 *Ordinal regression*

3.1.1 Description of analysis

A Bayesian hierarchical ordinal cumulative regression was fit to the seven-point Likert-scale data as a function of dummy-coded factors Morphology (reference level 'derivation'), Cluster (reference level OL_C), and Origin (reference level 'France') and all their interactions, using Stan (Carpenter et al. 2017) and the `brms` package (Bürkner 2017) in R (R Core Team 2020). The model included the maximal random effect structure justified by the study's design (Barr et al. 2013), allowing the effects and their interactions to vary by participant (Morphology, Cluster) and by word (Origin). The probit link function was used in order to apply a cumulative model assuming the latent variable to be normally distributed (Bürkner & Vuorre 2019: 84). The default priors of the `brms` package were used. Equal variances were assumed for the unobserved variables that underlie the observed ordinal variable.

Four sampling chains with 4000 iterations with a warm-up period of 2000 iterations for each chain were run, resulting in a total of 8000 samples. To avoid initialization at too small or too large values, initial values for the MCMC sampler were set to zero.[5]

For all relevant parameters, their mean and 95% credibility interval (CI) according to the model's posterior distribution are reported. In the analysis, the parameters concern the latent unobserved continuous variable corresponding to participants' opinion about the likelihood of schwa absence. Due to the way the Likert-scale was set up, greater values correspond to a greater likelihood of schwa deletion (according to the participants). For testing hypotheses about the difference Δ between two conditions, Franke & Roettger (2019)'s recommendations were followed. The posterior probability that this difference is larger than zero ($\Delta > 0$) is reported. If this probability is close to 1 and furthermore zero is outside of the posterior 95% CI for Δ, compelling evidence is considered to be provided for the hypothesis that posits the existence of a difference between the relevant conditions.

3.1.2 Description of results

Figure 2 shows the posterior distribution (mean and 95% CI) of each response category (1, 2, ..., 6, 7) for all cells in the factorial design. This posterior distribution was calculated using Equation 5 in Bürkner & Vuorre (2019: 79). This equation expresses the probability of each response category k as a function of the predictors, their corresponding regression coefficients, and the thresholds τ_k and τ_{k-1} inferred along the latent continuous variable.

Differences between clusters in derived words. Participants were found to rate schwa absence as more likely in LO_L than in OL_C in derivation and there is compelling evidence for this difference for participants from both France ($\mathbb{E}(\mu_{\text{French, LO\_L, der}} - \mu_{\text{French, OL\_C, der}}) = 0.54$, CI=

[5] This issue is discussed by Paul Bürkner on the Stan forums (https://discourse.mc-stan.org/t/initialization-error-try-specifying-initial-values-reducing-ranges-of-constrained-values-or-reparameterizing-the-model/4401).

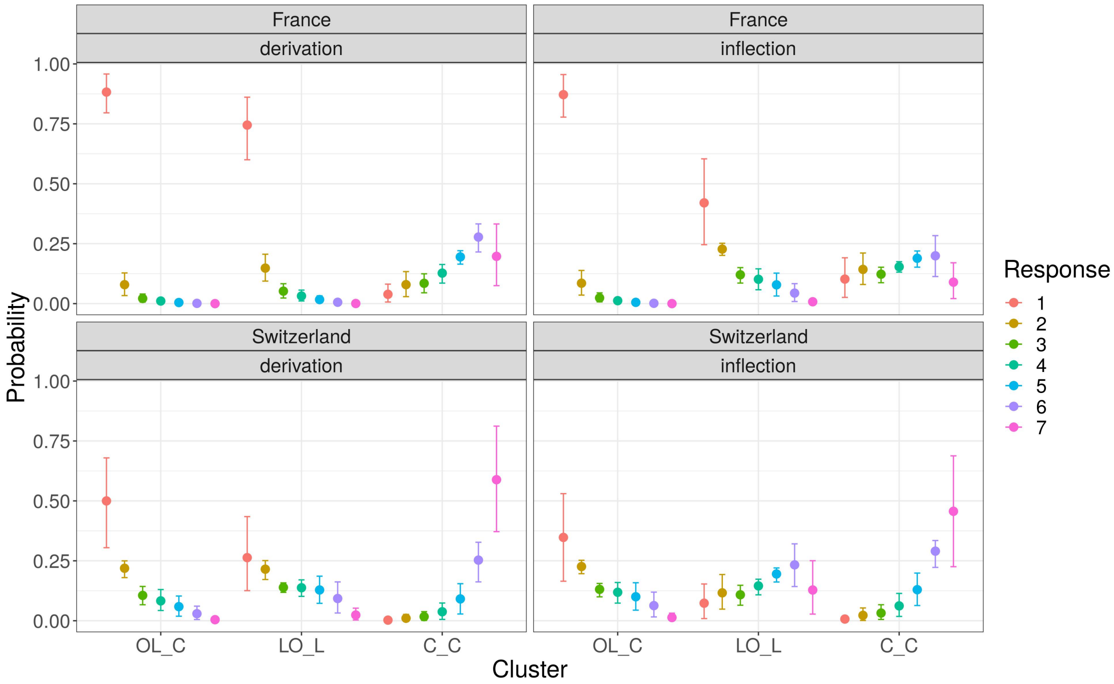

Figure 2: Posterior distribution (mean and 95% CI) of each of the seven response categories as a function of Morphology, Cluster, and Origin.

$[0.15, 0.93]$, $P(\Delta > 0) = 1$) and Switzerland ($\mathbb{E}(\mu_{\text{Swiss, LO\_L, der}} - \mu_{\text{Swiss, OL\_C, der}}) = 0.65$, CI$=[0.25, 1.08]$, $P(\Delta > 0) = 1$).[6] Participants were found to rate schwa absence as more likely in C_C than in LO_L in derivation and there is compelling evidence for this difference for participants from both France ($\mathbb{E}(\mu_{\text{French, C\_C, der}} - \mu_{\text{French, LO\_L, der}}) = 2.49$, CI$=[2.02, 2.99]$, $P(\Delta > 0) = 1$) and Switzerland ($\mathbb{E}(\mu_{\text{Swiss, C\_C, der}} - \mu_{\text{Swiss, LO\_L, der}}) = 2.28$, CI$=[1.74, 2.85]$, $P(\Delta > 0) = 1$).

Differences between clusters in inflected words. Participants were also found to rate schwa absence as more likely in LO_L than in OL_C in inflection and there is compelling evidence for this difference for participants from both France ($\mathbb{E}(\mu_{\text{French, LO\_L, inf}} - \mu_{\text{French, OL\_C, inf}}) = 1.37$, CI$=[0.92, 1.78]$, $P(\Delta > 0) = 1$) and Switzerland ($\mathbb{E}(\mu_{\text{Swiss, LO\_L, inf}} - \mu_{\text{Swiss, CO\_L, inf}}) = 1.11$, CI$=[0.63, 1.57]$, $P(\Delta > 0) = 1$). Participants were found to rate schwa absence as more likely in C_C than in LO_L in inflected words and there is compelling evidence for this difference for participants from both France ($\mathbb{E}(\mu_{\text{French, C\_C, inf}} - \mu_{\text{French, LO\_L, inf}}) = 1.10$, CI$=[0.74, 1.45]$, $P(\Delta > 0) = 1$) and Switzerland ($\mathbb{E}(\mu_{\text{Swiss, C\_C, inf}} - \mu_{\text{Swiss, LO\_L, inf}}) = 1.07$, CI$=[0.67, 1.48]$, $P(\Delta > 0) = 1$).

It can be concluded that there is sufficient evidence to support the hypothesis that both the number and nature of surrounding consonants matters in both derived and inflected words, with OL_C being judged as more likely to feature schwa than LO_L and LO_L more likely to feature schwa than C_C in both derived and inflected words.

Differences between derivation and inflection. For OL_C clusters, participants were found to rate schwa absence as more likely in inflection than in derivation, but there is compelling evidence for this difference for participants from Switzerland ($\mathbb{E}(\mu_{\text{Swiss, OL\_C, inf}} - \mu_{\text{Swiss, OL\_C, der}}) = 0.40$, CI$=[0.05, 0.74]$, $P(\Delta > 0) = 0.99$) but not for participants from France ($\mathbb{E}(\mu_{\text{French, OL\_C, inf}} - \mu_{\text{French, OL\_C, der}}) = 0.05$, CI$=[-0.29, 0.39]$, $P(\Delta > 0) = 0.62$). For LO_L clusters, participants were found to rate schwa absence as more likely in inflection than in derivation, and there is compelling for this difference for participants from both France ($\mathbb{E}(\mu_{\text{French, LO\_L, inf}} - \mu_{\text{French, LO\_L, der}}) = 0.88$, CI$=[0.48, 1.28]$, $P(\Delta > 0) = 1$) and Switzerland ($\mathbb{E}(\mu_{\text{Swiss, OL\_C, inf}} - \mu_{\text{Swiss, OL\_C, der}}) = 0.86$,

[6] Notation $\mathbb{E}()$ is a shorthand for the expectation (mean) of the posterior distribution of interest.

CI= $[0.43, 1.30]$, $P(\Delta > 0) = 1$). An unexpected result was obtained for C_C clusters. In this context, participants were found to rate schwa absence as less likely in inflection than in derivation, with compelling evidence from participants from both France ($\mathbb{E}(\mu_{\text{French, C\_C, inf}} - \mu_{\text{French, C\_C, der}}) = -0.51$, CI= $[-0.81, -0.19]$, $P(\Delta > 0) = 0$) and Switzerland ($\mathbb{E}(\mu_{\text{Swiss, C\_C, inf}} - \mu_{\text{Swiss, C\_C, der}}) = -0.35$, CI= $[-0.69, 0.01]$, $P(\Delta > 0) = 0.03$). This result is unexpected in two ways. The literature on French indeed does not report any asymmetry between inflection and derivation in schwa likelihood for two-consonant clusters (e.g. Côté 2001: 85). Moreover, if an asymmetry was to observed, one would expect it to go in the opposite direction, with derivation favoring schwa presence more than inflection. Indeed, this is what has been observed for three-consonant clusters in the present study and this is also what is expected under the general hypothesis that schwa is more likely at the boundary of lower morphosyntactic domains (Dell 1977; Côté 2001).

Because this paper is primarily interested in the behavior of three-consonant clusters, the sonority profile of C_C sequences was not controlled for when selected the experimental items for derived and inflected words. However two-consonant clusters with increasing sonority have been shown to favor the absence of schwa as compared to two-consonant clusters without increasing sonority (Bürki et al. 2011). Therefore one must check that the unexpected result found for C_C cannot be attributed to uncontrolled differences in sonority in the two sets of experimental items. Both sets of inflected words and derived words featured more two-consonant clusters with increasing sonority, as shown in Table 3. But the asymmetry was slightly stronger in derived words than in inflected words (10/4 vs. 10/5) and this might have slightly favored schwa absence in derived words as compared to inflected words.

		Suffix derivational	 inflectional
Sonority	increasing	10	10
		tm, tʁ, ʃm, nʁ, kl, mʁ, sl, dl, ʒm, fʁ	nʁ, kʁ, ʒʁ, tʁ, nʁ, pʁ, tʁ, dʁ, mʁ, sʁ
	non-increasing	4	5
		lm, ʁt, tt, ʃt	lʁ, lʁ, ʁʁ, ʁʁ, jʁ

Table 3: Sonority profile of C_C clusters in the experimental items as a function of morphology. The following sonority scale from Côté (2009) is assumed, from the least to most sonorous: stops < fricatives < nasals < liquids < glides.

To test the effect of sonority in this condition, a further analysis was run. The same ordinal regression was run but replacing the three-level variable Cluster with a four-level variable distinguishing C_C sequences with and without increasing sonority. The two models were compared using the difference in the expected log pointwise predictive density (EPLD) for the two models (Vehtari et al. 2017), with the function `loo_compare()` in the `brms` package. The model including sonority was not found to be better than the simpler model that does not include it according to this criterion ($\Delta = -2$, $SE = 5.1$). Furthermore, for speakers from France, derived words were still found to favor schwa absence as compared to inflected words both when sonority was increasing ($\mathbb{E}(\mu_{\text{French, C\_C}_{\text{increase}}\text{, inf}} - \mu_{\text{French, C\_C}_{\text{increase}}\text{, der}}) = -0.50$, CI= $[-0.85, -0.16]$, $P(\Delta < 0) = 1$) and when sonority was not increasing ($\mathbb{E}(\mu_{\text{French, C\_C}_{\text{noincrease}}\text{, inf}} - \mu_{\text{French, C\_C}_{\text{noincrease}}\text{, der}}) = -0.50$, CI= $[-1.03, -0.02]$, $P(\Delta < 0) = 0.97$). For Swiss speakers, the same result was found, but without compelling evidence. All in all, these results mean that the asymmetry between derived and inflected words in C_C found in this study cannot be explained away as an effect of sonority.

Remark on the effect of origin (France vs. Switzerland). Differences between participants from France and Switzerland were not systematically tested because this question is peripheral to the paper's main research question. However, one can note that participants from Switzerland appear

$$\begin{array}{ccccc} \mathrm{OL\_C_{der}} & > & \mathrm{LO\_L_{der}} & > & \mathrm{C\_C_{der}} \\ \vee_{\mathrm{Sw}} \; \parallel_{\mathrm{Fr}} & & \vee & & \wedge \\ \mathrm{OL\_C_{inf}} & > & \mathrm{LO\_L_{inf}} & > & \mathrm{C\_C_{inf}} \end{array}$$

Table 4: Summary of results: probability of the schwa variant as a function of Cluster, Morphology, and Origin (France vs. Switzerland).

on Figure 2 to be more likely to accept the schwa-less variant than participants from France (they have overall a larger proportion of 1 responses in OL_C and LO_L and a larger proportion of 7 responses in C_C). This is line with the conclusions reached by Racine (2007; 2008) based on judgment data and by Racine et al. (2016) (among others) based on production data showing more schwa deletion in word-initial syllables for Swiss speakers. However, as shown in the results above, participants from France and Switzerland behaved similarly with respect to the hypotheses studied in this paper, except for the treatment of OL_C in derived vs. inflected words. This similarity between the two groups of participants is line with Racine (2007)'s observation that speakers from France and Switzerland differ in their baseline rates of schwa production but otherwise follow the same general principles for schwa-zero alternations.

Summary of the results. The study's results are summarized in Table 4, with $>$ indicating a greater estimated likelihood of the schwa variant. The analysis of the judgment data support the hypothesis that not only the size of consonant clusters but also the type of three-consonant clusters matters for schwa-zero alternations even in derived words. Furthermore, the same asymmetry is observed in derived and inflected words, with OL_C clusters favoring schwa variants more than LO_L clusters according to the participants' judgments. There is also an effect of cluster size, as C_C clusters were judged to be more likely to feature schwa deletion than OL_C and LO_L clusters. These results therefore support the weak version of the lexical-phonology hypothesis and are in line with substance-based theories of phonotactics predicting that the same phonotactic asymmetries should hold across the language's strata.

In accordance with the hypothesis that schwa presence at stem-suffix boundaires is more likely in derived than in inflected words, schwa presence was generally judged more likely in OL_C and LO_L in derived than in inflected words (although there was no compelling evidence for this effect for the participants from France in the case of OL_C). However, this hypothesis was contradicted by the results for two-consonant sequences (C_C). In this context, inflection was unexpectedly found to favor schwa presence and this unexpected effect could not be explained away as a result of uncontrolled asymmetries in the clusters' sonority profiles. These results suggest an important difference in the patterning of three-consonant clusters and two-consonant clusters: three-consonant clusters are more tolerated in inflection than in derivation whereas two-consonant clusters are more tolerated in inflection than in derivation.

3.2 Probabilistic constraint-based grammars

To supplement the analysis of judgment data with a more linguistically meaningful interpretation, the data were also analyzed using probabilistic constraint-based grammars. In this framework, the likelihood of schwa presence/absence in the different experimental conditions can be directly interpreted in terms of constraint weights. This makes it possible to interpret the judgment data in terms of the relative strengths of phonotactic constraints against consonant clusters. In this section, the judgment data were aggregated across all participants and words. In other words, the grammar that was inferred is the average grammar across participants and words. Although participants from France and Switzerland probably have different grammars because they have different baseline rates for schwa production, they generally show the same asymmetries between consonant clusters, as summarized in Table 4. Therefore, the weights for the corresponding phonotactic constraints should be ordered in the same way for both groups.

3.2.1 Description of analysis

For the constraint-based analysis, the response variable (the 7-point Likert scale) was transformed into a binary variable (schwa presence vs. absence). The reason for this transformation is that constraint-based grammars are designed as models of language production (a form is produced or not) rather than as models of metalinguistic judgment. In language production, a form is produced or not. In judgment data, a form may receive a gradient judgment of acceptability and this does not directly translate into a binary choice (unless binary judgments are collected). However constraint-based grammars may be used and are often used to model judgment data (e.g. Boersma & Hayes 2001 on dark and light /l/ in English, Smith & Pater 2020 on schwa-zero alternations in French). If the judgment data are not binary, this requires applying a transformation that binarizes the data (e.g. Boersma & Hayes 2001: 82). In this paper, the following transformation was applied. Words that received ratings strictly above 4 were treated as categorically favoring the schwa-less variant. Words that received ratings strictly below 4 were treated as categorically favoring the schwa variant. Words that received a rating equal to 4 were randomly assigned to one or the other category.

Two constraint-based grammars were fit to the transformed data aggregated across participants and words, using MaxEnt as grammatical framework (Hayes & Wilson 2008). Two grammars were constructed to represent the weak and strong lexical-phonology hypotheses, respectively. The first grammar had a different markedness constraint for each of the six cluster-suffix combinations (*OLC$_{\text{inf}}$, *LOL$_{\text{inf}}$, *CC$_{\text{inf}}$, *OLC$_{\text{der}}$, *LOL$_{\text{der}}$, *CC$_{\text{der}}$), allowing for OL_C and LO_L to behave differently in derived words. The second grammar was identical except that it had a single *CCC constraint for derived words, in accordance with the hypothesis that the Law of Three Consonants is categorical in this context (*OLC$_{\text{inf}}$, *LOL$_{\text{inf}}$, *CC$_{\text{inf}}$, *CCC$_{\text{der}}$, *CC$_{\text{der}}$).[7] Table 5 shows how the first grammar assigns different constraint violations for OL_C and LO_L in derived words. Table 6 shows how the two clusters are treated identically in derived words in the second grammar.

Both grammars included a faithfulness constraint protecting against schwa epenthesis (Dep(V)). This analysis assumes that the schwa-less variant is the underlying form and the schwa variant is derived through epenthesis. This is the classic analysis of French schwa at morpheme boundaries (Dell 1985). However this choice is not crucial to the analysis.

For both analyses, the constraint weights were inferred using a Bayesian binomial regression implemented in `rjags` (Plummer 2016). To help with model convergence, one of the weights was set to a constant value of 1 (the weight of *CC$_{\text{inf}}$). Following Goldwater & Johnson (2003), a Gaussian prior with mean equal to zero was chosen for all other constraint weights. Informally, this prior specifies that zero is the default weight for constraints (which means that the constraint has no effect on the output). The variance of the Gaussian prior was set to 1,000. Three MCMC chains were used with 100,000 samples and a thinning interval of 10 (which means that every 10th value in the chain was kept in the final MCMC sample while all other values were discarded). The first 5,000 samples of each chain were used for burn-in (which means they were also discarded).

[7] In this paper, markedness hierarchies are set up as *scale-partition* constraint families and not as *stringency* constraint families (see Smith & Moreton 2012 for a discussion of these two approaches). In the stringency approach, there would be one markedness constraint banning specific clusters (e.g. *OLC) and a general markedness constraint banning all CCC clusters (*CCC) instead of two specific markedness constraints (*OLC, *LOL). Similarly, in the stringency approach, there would be a morphologically indexed markedness constraint (e.g. *OLC$_{\text{der}}$) and a general markedness constraint that does not depend on morphological domains (*OLC) instead of two morphologically indexed markedness constraints (*OLC$_{\text{der}}$, *OLC$_{\text{inf}}$). Specific constraints were chosen in all cases so as not to bias the analysis in one way or the other (e.g. OLC is not *a priori* assumed to be more marked than LOL, clusters are not *a priori* assumed to be more marked in derivation than in inflection). Constraint weights only (and not constraint violations) will determine whether one context is more marked than the other.

/kõsjɛʁʒ-ʁi/	*OLC$_{der}$	*OLC$_{inf}$	*LOL$_{der}$	*LOL$_{inf}$	*CC$_{der}$	*CC$_{inf}$	Dep(V)
[kõsjɛʁʒʁi]			1				
[kõsjɛʁʒəʁi]							1

LO_-L$_{derivation}$: *concierge-rie* ‘caretaker’s lodge’

/sufl-ʁi/	*OLC$_{der}$	*OLC$_{inf}$	*LOL$_{der}$	*LOL$_{inf}$	*CC$_{der}$	*CC$_{inf}$	Dep(V)
[suflʁi]	1						
[sufləʁi]							1

OL_-C$_{derivation}$: *souffle-rie* ‘wind tunnel’

Table 5: Constraint violations for schwa-zero alternations in derived and inflected words, as a function of the nature of surrounding consonants (OL_C vs. LO_), according to the weak lexical-phonology hypothesis (constraints *OLC$_{der}$ and *LOL$_{der}$ are distinct).

/kõsjɛʁʒ-ʁi/	*CCC$_{der}$	*OLC$_{inf}$	*LOL$_{inf}$	*CC$_{der}$	*CC$_{inf}$	Dep(V)
[kõsjɛʁʒʁi]	1					
[kõsjɛʁʒəʁi]						1

LO_-L$_{derivation}$: *concierge-rie* ‘caretaker’s lodge’

/sufl-ʁi/	*CCC$_{der}$	*OLC$_{inf}$	*LOL$_{inf}$	*CC$_{der}$	*CC$_{inf}$	Dep(V)
[suflʁi]	1					
[sufləʁi]						1

OL_-C$_{derivation}$: *souffle-rie* ‘wind tunnel’

Table 6: Constraint violations for schwa-zero alternations in derived and inflected words, as a function of the nature of surrounding consonants (OL_C vs. LO_), according to the strong lexical-phonology hypothesis (there is a single *CCC$_{der}$ for derived words).

Convergence of the chains on the posterior distribution was assessed using the Gelman-Rubin statistic: it was very close to 1 for all parameters,[8] indicating that the samples were representative of the posterior distribution (Kruschke 2015: 181). The effective sample size for each constraint weight estimated by the model was superior to 10,000, indicating that the MCMC samples were large enough for stable and accurate numerical estimates of the posterior distributions (Kruschke 2015: 184). For model comparison, the deviance information criterion (DIC; Gelman et al. 2013: 172-173) was used.

3.2.2 Description of results

The posterior distributions for the constraint weights are shown in Table 7 and 8 for the grammar that distinguishes three-consonant clusters in derived words and for the grammar that does not, respectively. Note that *CCC and *OLC$_{inf}$ end up having the same weights in Table 8 but this is not a feature of the analysis. The more complex grammar in Table 7 was found to have a smaller deviation information criterion than the more simple one in Table 8 ($\Delta = -19.36$), indicating that it is a better model of the data. The predicted frequencies of schwa deletion under the two grammars

[8] The Gelman-Rubin statistics was calculated individually for each parameter and not globally for all parameters because one of the parameters (the weight of *CC$_{inf}$) was set to a constant value of 1 to help with model convergence and a global Gelman-Rubin statistics cannot be computed in this case. See the following post by Martyn Plummer for more details: https://sourceforge.net/p/mcmc-jags/discussion/610037/thread/28cef6e5/.

are plotted against the attested frequencies in Figures 3 and 4. The predictions of the more complex grammar in Figure 3 better match the attested frequencies. In other words, the data provide evidence for constraints referencing the nature of consonants in CCC clusters even in lexical phonology.

Constraint	Mean	95% CI
$*\text{OLC}_{\text{der}}$	4.17	[3.90, 4.46]
$*\text{OLC}_{\text{inf}}$	3.77	[3.53, 4.03]
$*\text{LOL}_{\text{der}}$	3.46	[3.22, 3.70]
$*\text{LOL}_{\text{inf}}$	2.20	[2.00, 2.40]
Dep(V)	1.89	[1.74, 2.04]
$*\text{CC}_{\text{inf}}$	1.00	
$*\text{CC}_{\text{der}}$	0.65	[0.43, 0.87]

Table 7: Posterior distribution of the constraint weights (mean and 95% CI) in the grammar distinguishing $*\text{OLC}_{\text{der}}$ and $*\text{LOL}_{\text{der}}$.

As expected under the hypothesis that *OLC is more marked than *LOL and *LOL more marked than *CC, *OLC was found to have a greater weight than *LOL and *LOL a greater weight than *CC within each stratum (derivation and inflection), as shown in Table 7. As expected under the hypothesis that three-consonant clusters are more marked in lexical phonology than in post-lexical phonology, *OLC and *LOL were found to have greater weights in lexical phonology than in post-lexical phonology. The unexpected result found in section 3.1.2 was replicated in the analysis with constraints. Indeed, two-consonant clusters were found to be more marked in post-lexical phonology than in lexical phonology: the weight of $*\text{CC}_{\text{inf}}$ is larger than the weight of $*\text{CC}_{\text{der}}$.

Constraint	Mean	95% CI
$*\text{CCC}_{\text{der}}$	3.78	[3.57, 3.99]
$*\text{OLC}_{\text{inf}}$	3.78	[3.53, 4.03]
$*\text{LOL}_{\text{inf}}$	2.20	[1.99, 2.40]
Dep(V)	1.89	[1.74, 2.04]
$*\text{CC}_{\text{inf}}$	1.00	
$*\text{CC}_{\text{der}}$	0.65	[0.43, 0.87]

Table 8: Posterior distribution of the constraint weights (mean and 95% CI) in the grammar with $*\text{CCC}_{\text{der}}$.

4 Discussion

The study's results support the weak version of the lexical-phonology hypothesis: in derived words, not only the number but also the nature of surrounding consonants matters for schwa-zero alternations. This means that Grammont's Law of Three Consonants should be relaxed not only for post-lexical but also for lexical phonology. Furthermore, relative markedness of clusters was found to be the same in both derived and inflected words, with OLC being more strongly avoided than LOL and LOL being more strongly avoided than CC. This is consistent with substance-based theories of phonotactic constraints that hold that asymmetries between phonotactic constraints ultimately derive from perceptual/articulatory or sonority asymmetries and therefore predict that these phonotactic asymmetries should be consistent across domains. Finally, three-consonant clusters were found to be more strongly avoided in derived than in inflected words, in line with the hypothesis that phonotactic restrictions are stronger in lower than in higher morphosyntactic domains. However, this result did not extend to two-consonant clusters: two-consonant clusters were unexpectedly

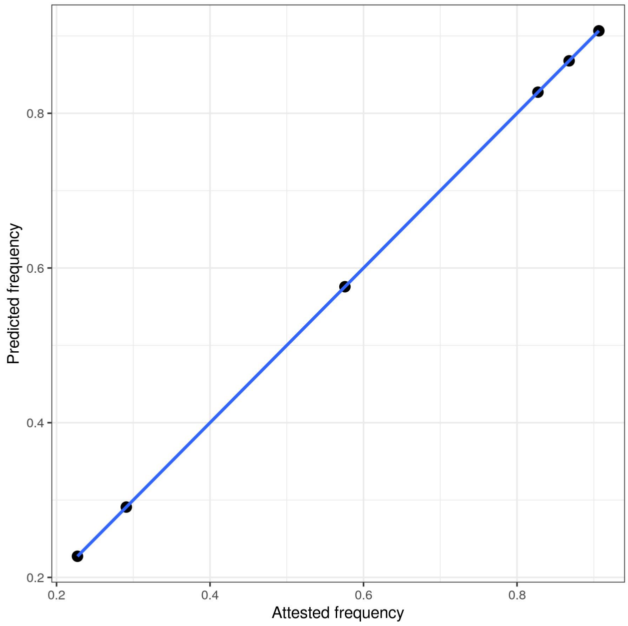

Figure 3: Grammar with $*OL\_C_{der}$ and $*LO\_L_{der}$: model predictions vs. data.

found to be more strongly avoided in inflected than in derived words. Furthermore, this result cannot be explained away as an effect of sonority. Section 4.1 briefly discusses this unexpected result. Section 4.2 discusses a lexical context that was not included in the present study and should be investigated in future work.

4.1 Effect of morphosyntactic domain

The judgment data collected in this study support the hypothesis that markedness constraints are relativized to domains (post-lexical and lexical phonology). However they do not strongly support Côté (2001: 162)'s hypothesis that clusters are more marked at the boundary of lower morphosyntactic/prosodic domains. The analysis of the judgment data suggests that participants from France do not weigh *OLC more in lexical phonology than in post-lexical phonology. Furthermore, both participants from France and Switzerland weigh *CC more in post-lexical phonology than in lexical phonology, contrary to Côté's predictions. The effect of morphosyntactic/prosodic domains on cluster markedness should be tested more thoroughly than what has been done. In particular, one should test whether two-consonant clusters behave differently from three-consonant clusters systematically across domains or whether this is limited to the opposition between derived words and inflected words.

4.2 Morpheme-internal consonant clusters

There is another context that would require further examination in lexical phonology, namely CC_C sequences that are entirely contained within a morpheme. This context was not considered here because schwa is usually considered to be banned altogether in CC_C within a morpheme. For instance, in the monomorphemic word *breton* 'Breton', the vowel is not described as a schwa [ə] but as a full mid vowel [ø]/[œ]. The reason is that this vowel is reported to not alternate with zero (it is undeletable) and alternating with zero is widely considered as a defining property for schwa. However, in theoretical frameworks that assume Richness of the Base, the unavailability of schwa-zero alternations in this specific context has to be explained through constraint interaction.

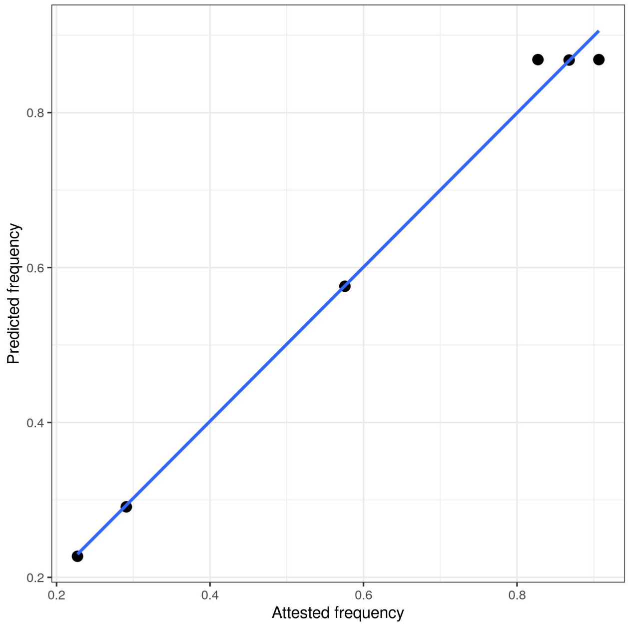

Figure 4: Grammar with $*CCC_{der}$: model predictions vs. data.

In this particular case, the absence of alternation with zero could be due to a *OLC constraint with a very large weight. Under this view, the vowel occurring in *breton* could be a schwa but this schwa would not alternate with zero due to some phonotactic constraint. Note that schwa status cannot be established phonetically because schwa is not clearly distinct from [ø]/[œ] phonetically (Malécot & Chollet 1977).

There is another reason to doubt that schwa should be defined entirely based on its pattern of alternations with zero. Schwa is involved in another vowel alternation, namely [ə]-[ɛ] (e.g. *nous appelons* [ap(ə)lɔ̃] 'we call' vs. *il appelle* [apɛl] 'he calls'). However, for some words, [ə] alternates with [ɛ] but not with zero, for instance in *crever* 'burst' (e.g. *nous crevons* [kʁøvɔ̃] 'we burst' - *il crève* [kʁɛv] 'he bursts'). The absence of alternation with zero is not arbitrary: the vowel does not alternate with zero because the variant with zero would involve an OLC cluster, as in *breton*, and this cluster is arguably highly marked within a morpheme.

If schwa can occur in CC_C within a single morpheme, then one should test whether the nature of surrounding consonants also plays a role in schwa-zero alternations in this context as it does in derived and inflected words. Future work should compare whether the effect of the nature of surrounding consonants within derived words (e.g. *souffle-rie* vs. *garde-rie*) extends to lexical stems (e.g. *breton* vs. *squelette* 'skeleton'). Intuitively, schwa seems to be more likely to delete in derived words (*gard(e)-rie*) than within lexical stems (*squ(e)lette*). If correct, this means that one should distinguigh markedness constraints that apply within lexical stems and those that apply at the boundary of stems and derivational suffixes. Testing this is left to future work.

5 Conclusion

Grammont's Law of Three Consonants (LTC) states that schwa is obligatorily pronounced in CC_C sequences in French to avoid three-consonant clusters. Although the LTC has been shown to depend on the nature and order of consonants in CC_C in post-lexical phonology, Grammont's categorical formulation is still considered as accurate to describe schwa-zero alternations in lexical phonology. The judgment data collected in this study support the hypothesis that not only the number but also the nature of surrounding consonants matters for schwa-zero alternations in derived words. This means that Grammont's Law of Three Consonants should be relaxed not only for post-lexical but also for lexical phonology. The results suggest that the same phonotactic constraints are relevant in lexical and post-lexical phonology, but with potentially different weights in the two strata. Furthermore, the same phonotactic asymmetries were found in both strata. This is compatible with theories of phonotactics that hold that phonotactic asymmetries are not arbitrary but rooted in extragrammatical factors such as perception, articulatory effort or sonority.

6 Appendix

6.1 Experimental stimuli

Word	Morphology	Cluster
adéquatement	derivation	C_C
biscuiterie	derivation	C_C
branchement	derivation	C_C
brutalement	derivation	C_C
cochonnerie	derivation	C_C
coquelet	derivation	C_C
entièreté	derivation	C_C
fumerie	derivation	C_C
immédiateté	derivation	C_C
osselet	derivation	C_C
rondelet	derivation	C_C
soulagement	derivation	C_C
tacheté	derivation	C_C
tartufferie	derivation	C_C
âcreté	derivation	OL_C
cidrerie	derivation	OL_C
déchiffrement	derivation	OL_C
dérèglement	derivation	OL_C
doublement	derivation	OL_C
effondrement	derivation	OL_C
espièglerie	derivation	OL_C
goinfrerie	derivation	OL_C
opiniâtreté	derivation	OL_C
pauvreté	derivation	OL_C
propreté	derivation	OL_C
ronflement	derivation	OL_C
sensiblerie	derivation	OL_C
soufflerie	derivation	OL_C
souffleté	derivation	OL_C
conciergerie	derivation	LO_L
conserverie	derivation	LO_L
corderie	derivation	LO_L
écorcherie	derivation	LO_L
étourderie	derivation	LO_L
fourberie	derivation	LO_L
garderie	derivation	LO_L
gendarmerie	derivation	LO_L
infirmerie	derivation	LO_L
porcherie	derivation	LO_L
porterie	derivation	LO_L
solderie	derivation	LO_L
tarderie	derivation	LO_L
turquerie	derivation	LO_L
elles dérouleront	inflection	C_C
il baillera	inflection	C_C
il cuisinera	inflection	C_C
il traquera	inflection	C_C
ils vengeront	inflection	C_C
je chanterai	inflection	C_C
je condamnerai	inflection	C_C
je désirerai	inflection	C_C
nous découperons	inflection	C_C
nous trouverons	inflection	C_C
tu aideras	inflection	C_C
tu consumeras	inflection	C_C
tu entoureras	inflection	C_C
vous annoncerez	inflection	C_C
vous coulerez	inflection	C_C
elle doublera	inflection	OL_C
elles déchiffreront	inflection	OL_C
il règlera	inflection	OL_C
il soufflera	inflection	OL_C
ils gifleront	inflection	OL_C
ils intègreront	inflection	OL_C
je délivrerai	inflection	OL_C
je montrerai	inflection	OL_C
nous contemplerons	inflection	OL_C
nous engouffrerons	inflection	OL_C
nous sombrerons	inflection	OL_C
tu consacreras	inflection	OL_C
tu rassembleras	inflection	OL_C
vous encadrerez	inflection	OL_C
vous encerclerez	inflection	OL_C
elle concernera	inflection	LO_L
elle occultera	inflection	LO_L
il regardera	inflection	LO_L
je calmerai	inflection	LO_L
je confirmerai	inflection	LO_L
je conserverai	inflection	LO_L
nous garderons	inflection	LO_L
tu consulteras	inflection	LO_L
tu forceras	inflection	LO_L
tu perceras	inflection	LO_L
tu porteras	inflection	LO_L
tu tourneras	inflection	LO_L
vous armerez	inflection	LO_L
vous discernerez	inflection	LO_L
vous marcherez	inflection	LO_L

6.2 Fillers

Word
cerisier
chemin
cheminée
chemise
fenêtre
fenouil
gobelet
levure
melon
menu
neveu
peluche
regret
remontant
remords
repassage
retard
revanche
secours
secret
semaine
seringue
breloque
grenade
breton
frelon
grelot

References

Anttila, Arto & Giorgio Magri. 2018. Does MaxEnt overgenerate? Implicational universals in Maximum Entropy grammar. In Gillian Gallagher, Maria Gouskova & Yin Sora (eds.), *AMP 2017: Proceedings of the 2017 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America. https://doi.org/10.3765/amp.v5i0.4260.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bayles, Andrew, Aaron Kaplan & Abby Kaplan. 2016. Inter-and intra-speaker variation in French schwa. *Glossa: a journal of general linguistics* 1(1). https://doi.org/10.5334/gjgl.54.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32. 45–86. https://doi.org/10.1162/002438901554586.

Bürki, Audrey, Mirjam Ernestus, Cédric Gendrot, Cécile Fougeron & Ulrich Hans Frauenfelder. 2011. What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *The Journal of the Acoustical Society of America* 130(6). 3980–3991. https://doi.org/10.1121/1.3658386.

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.

Bürkner, Paul-Christian & Matti Vuorre. 2019. Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science* 2(1). 77–101. https://doi.org/10.1177/2515245918823199.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li & Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software* 76(1). https://doi.org/10.3102/1076998615606113.

Clements, George N. 1990. The role of the sonority cycle in core syllabification. *Papers in laboratory phonology* 1. 283–333. https://doi.org/10.1017/CBO9780511627736.017.

Côté, Marie-Hélène. 2001. *Consonant cluster phonotactics: a perceptual approach*. Cambridge, MA: MIT dissertation. https://doi.org/10.7282/T3HD7TGR.

Côté, Marie-Hélène. 2007. Rhythmic constraints on the distribution of schwa in French. In J. Camacho, N. Flores-Ferrán, L. Sánchez, V. Déprez & J. Cabrera (eds.), *Romance Linguistics 2006*, 79–92. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.287.07cot.

Côté, Marie-Hélène. 2009. Contraintes segmentales et variation dans la perte et la stabilisation du schwa en syllabe initial (segmental constraints and variation in the loss and stabilization of schwa in initial syllables). In Luc Baronian & France Martineau (eds.), *Le français d'un continent à l'autre*, 93–121. Quebec: Presses de l'Université de Laval.

Côté, Marie-Hélène. 2012. Laurentian French (Quebec). Extra vowels, missing schwas and surprising liaison consonants. In Randall Gess, Chantal Lyche & Trudel Meisenburg (eds.), *Phonological variation in French: Illustrations from three continents*, 235–274. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/silv.11.13cot.

Dell, François. 1976. Schwa précédé d'un groupe obstruante-liquide. *Recherches linguistiques Saint-Denis* (4). 75–111.

Dell, François. 1977. Paramètres syntaxiques et phonologiques qui favorisent l'épenthèse de schwa en français moderne. In Christian Rohrer (ed.), *Actes du colloque franco-allemand de linguistique théorique*. 141–153. Tübingen: Niemeyer.

Dell, François. 1978. Certains corrélats de la distinction entre morphologie dérivationnelle et morphologie flexionnelle dans la phonologie du français. *Etudes linguistiques sur les langues romanes. Montreal working papers in Linguistics* 10. 1–10.

Dell, François. 1985. *Les règles et les sons*. Paris: Hermann 2nd edn.

Durand, Jacques & Bernard Laks. 2000. Relire les phonologues du français: Maurice Grammont et la loi des trois consonnes. *Langue française* (126). 29–38. https://doi.org/10.3406/lfr.2000.4670.

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

Flemming, Edward. 2002. *Auditory Representations in Phonology*. New York: Routledge. https://doi.org/10.4324/9781315054803.

Franke, Michael & Timo B Roettger. 2019. Bayesian regression modeling (for factorial designs): A tutorial. https://doi.org/10.31234/osf.io/cdxv3.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari & Donald B Rubin. 2013. *Bayesian data analysis*. New York: Chapman and Hall/CRC. https://doi.org/10.1201/b16018.

Gess, Randall, Chantal Lyche & Trudel Meisenburg. 2012. *Introduction to phonological variation in French: Illustrations from three continents* 1–19. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/silv.11.01ges.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University, Department of Linguistics.

Grammont, Maurice. 1914. *Traité pratique de prononciation française*. Paris: Delagrave.

Hambye, Philippe & Anne Catherine Simon. 2012. The variation of pronunciation in Belgian French. In Randall Gess, Chantal Lyche & Trudel Meisenburg (eds.), *Phonological variation in French: Illustrations from three continents*, 129–149. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/silv.11.08ham.

Hansen, Anita Berit. 2012. A study of young Parisian speech: Some trends in pronunciation. In Randall Gess, Chantal Lyche & Trudel Meisenburg (eds.), *Phonological variation in French: Illustrations from three continents*, 151–172. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/silv.11.09han.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440. https://doi.org/10.1162/ling.2008.39.3.379.

Kawasaki-Fukumori, H. 1992. An acoustical basis for universal phonotactic constraints. *Language and Speech* 35. 73–86. https://doi.org/10.1177/002383099203500207.

Kruschke, John K. 2015. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press 2nd edn. https://doi.org/10.1016/B978-0-12-405888-0.00001-5.

Liddell, Torrin M & John K Kruschke. 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 79. 328–348. https://doi.org/10.1016/j.jesp.2018.08.009.

LimeSurvey. 2012. LimeSurvey: An open source survey tool. http://www.limesurvey.org.

Malécot, André. 1976. The effect of linguistic and paralinguistic variables on the elision of the French mute-e. *Phonetica* 33(2). 93–112. https://doi.org/10.1159/000259716.

Malécot, André & Gérard Chollet. 1977. The Acoustic Status of Mute-e in French. *Phonetica* 34. 19–30.

Myers, James. 2017. Acceptability judgments. In *Oxford Research Encyclopedia of Linguistics*, Oxford University Press. https://doi.org/0.1093/acrefore/9780199384655.013.333.

Plummer, Martyn. 2016. *rjags: Bayesian graphical models using MCMC*. https://CRAN.R-project.org/package=rjags. R package version 4-5.

R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Racine, Isabelle. 2007. Effacement du schwa dans des mots lexicaux: constitution d'une base de données et analyse comparative. *Proceedings of JEL* 125–130.

Racine, Isabelle. 2008. *Les effets de l'effacement du schwa sur la production et la perception de la parole en français*: University of Geneva dissertation. https://doi.org/10.13097/archive-ouverte/unige:602.

Racine, Isabelle & Helene Andreassen. 2012. A phonological study of a Swiss French variety: data from the canton of Neuchâtel. In Randall Gess, Chantal Lyche & Trudel Meisenburg (eds.), *Phonological variation in French: Illustrations from three continents*, 211–233. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/silv.11.10rac.

Racine, Isabelle, Jacques Durand & Helene Andreassen. 2016. PFC, codages et représentations: la question du schwa. *Corpus* (15). https://doi.org/10.4000/corpus.3014.

Racine, Isabelle & François Grosjean. 2002. La production du e caduc facultatif est-elle prévisible? Un début de réponse. *Journal of French Language Studies* 12(3). 307–326. https://doi.org/https://doi.org/10.1017/S0959269502000340.

Schütze, Carson T. 2016. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Berlin: Language Science Press. https://doi.org/https://doi.org/10.17169/langsci.b89.100.

Schütze, Carson T. & Jon Sprouse. 2013. Judgment data. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 27–50. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139013734.004.

Smith, Brian W & Joe Pater. 2020. French schwa and gradient cumulativity. *Glossa: a journal of general linguistics* 5(1). https://doi.org/doi.org/10.5334/gjgl.583.

Smith, Jennifer L & Elliott Moreton. 2012. Sonority variation in Stochastic Optimality Theory: Implications for markedness hierarchies. In Steve Parker (ed.), *The sonority controversy*, 167–194. De Gruyter Mouton. https://doi.org/10.1515/9783110261523.

Vehtari, Aki, Andrew Gelman & Jonah Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing* 27(5). 1413–1432. https://doi.org/10.1007/s11222-016-9696-4.