

Against the Law of Three Consonants in French: Evidence from Judgment Data

Benjamin Storme
Université de Lausanne

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 [dəmãd] or without schwa as [dmãd]. Determining the factors that condition the distribution of schwa-zero alternations has been a central topic in French phonology for more than a century (see Bürki et al. 2011:3982-3985 for an overview). Among these factors, 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 (1-a), but excluded when preceded by a single consonant (C_C), as illustrated in (1-b). 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 (1-a), the schwa form is preferred because it makes it possible to avoid the three-consonant cluster [tdm]. In (1-b), the schwa-less form is preferred in the absence of three-consonant clusters.

- (1) Grammont’s Law of Three Consonants (LTC)
- a. Schwa is obligatory in CC_C
C#C_C [tdm] sept demandes [setdəmãd] ‘seven requests’
 - b. Schwa is excluded in C_C
C_C [dm] la demande [ladmãd] ‘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, the likelihood of schwa presence depends on the consonants involved. For instance, 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 (2).

- (2)
- a. Schwa is more likely in OL_C (O=obstruent, L=liquid)
O#L_C [kls] chaque leçon [ʃak#l(ə)sõ] ‘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): 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_-C_{derivation}), schwa is reported to be categorically pronounced and this regardless of the nature and order of surrounding

* I would like to thank Marie-Hélène Côté and participants in AMP 2020 for helpful feedback and my students at the Université de Lausanne for participating in the experiment. All errors are mine.

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 (3-a) and (3-b), even though OL_C and CO_L are treated differently in general, as illustrated in (2) for word-initial contexts and in (4) before the inflectional future/conditional suffix *-r-*.

- (3) a. Schwa is obligatory in OL_C-derivation (O=obstruent, L=liquid)
 OL_C-derivation [glm] règlement [ʁɛɡlɛ-mã] ‘regulation’
 b. Schwa is obligatory in CO_L-derivation (O=obstruent, L=liquid)
 CO_L-derivation [ʁdʁ] garderie [ɡaʁdɛ-ʁi] ‘Kindergarten’
- (4) a. Schwa is more likely in OL_C-inflection (O=obstruent, L=liquid)
 OL_C-inflection [glm] règlera [ʁɛɡl(ə)-ʁa] ‘adjust-fut.3sg’
 b. Schwa is less likely in CO_L-inflection (O=obstruent, L=liquid)
 CO_L-inflection [ʁdʁ] gardera [ɡaʁd(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, as illustrated in (2) and (4). 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 (3).

The hypothesis of clearly distinct lexical and post-lexical phonologies will be referred to as the strong version of the lexical-phonology hypothesis. The predictions of the strong version of the lexical-phonology hypothesis are summarized in (5) for the two relevant consonant sequences (OL_C and LO_L) and the two morphological contexts this paper will focus on (derivation and inflection). The predictions are summarized in (5-a) at the level of the grammar (constraint set in each stratum) and in (5-b) at the level of the data (probability distribution of schwa-zero alternations).

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

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, 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. Indeed, the corpus used in this study contains 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).

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

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.

stratum. This hypothesis will be referred to as the weak version of the lexical-phonology hypothesis. Its predictions are summarized in (6), with (6-a) focusing on the predictions at the level of the grammar and (6-b) on the predictions at the level of the data.

- (6) 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)

$$P(\emptyset|OL_C_{\text{derivation}}) \neq P(\emptyset|LO_L_{\text{derivation}})$$

$$P(\emptyset|OL_C_{\text{inflection}}) \neq P(\emptyset|LO_L_{\text{inflection}})$$

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 (6).

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). 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. Following Racine (2007), the judgments were elicited 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.

la biscuiterie								
	1	2	3	4	5	6	7	
la biscuiterie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	la biscuit'rie

Figure 1: Judgment task

2.2 Participants 21 Swiss French speakers (8 females, 13 males; recruited among students at the Université de Lausanne) 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 1 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 can be found in a longer version of this paper.⁴

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

Table 1: Experimental items

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 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). 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

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.

⁴ Available on LingBuzz: <https://ling.auf.net/lingbuzz/005465>.

frameworks (e.g. Smith & Pater 2020).

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).⁵ 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 4,000 iterations with a warm-up period of 2,000 iterations for each chain were run, resulting in a total of 8,000 samples. To avoid initialization at too small or too large values, initial values for the MCMC sampler were set to zero.⁶

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 ($E(\mu_{\text{French, LO_L, der}} - \mu_{\text{French, OL_C, der}}) = 0.54$, CI= [0.15, 0.93], $P(\Delta > 0) = 1$) and Switzerland ($E(\mu_{\text{Swiss, LO_L, der}} - \mu_{\text{Swiss, OL_C, der}}) = 0.65$, CI= [0.25, 1.08], $P(\Delta > 0) = 1$).⁷ 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 ($E(\mu_{\text{French, C_C, der}} - \mu_{\text{French, LO_L, der}}) = 2.49$, CI= [2.02, 2.99], $P(\Delta > 0) = 1$) and Switzerland ($E(\mu_{\text{Swiss, C_C, der}} - \mu_{\text{Swiss, LO_L, der}}) = 2.28$, CI= [1.74, 2.85], $P(\Delta > 0) = 1$).

Differences between clusters in inflected words. Participants were also found to rate schwa absence

⁵ There was a random intercept by word and words were allowed to be treated differently depending on the speaker's origin (France vs. Switzerland).

⁶ 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>).

⁷ Notation $E()$ is a shorthand for the expectation (mean) of the posterior distribution of interest.

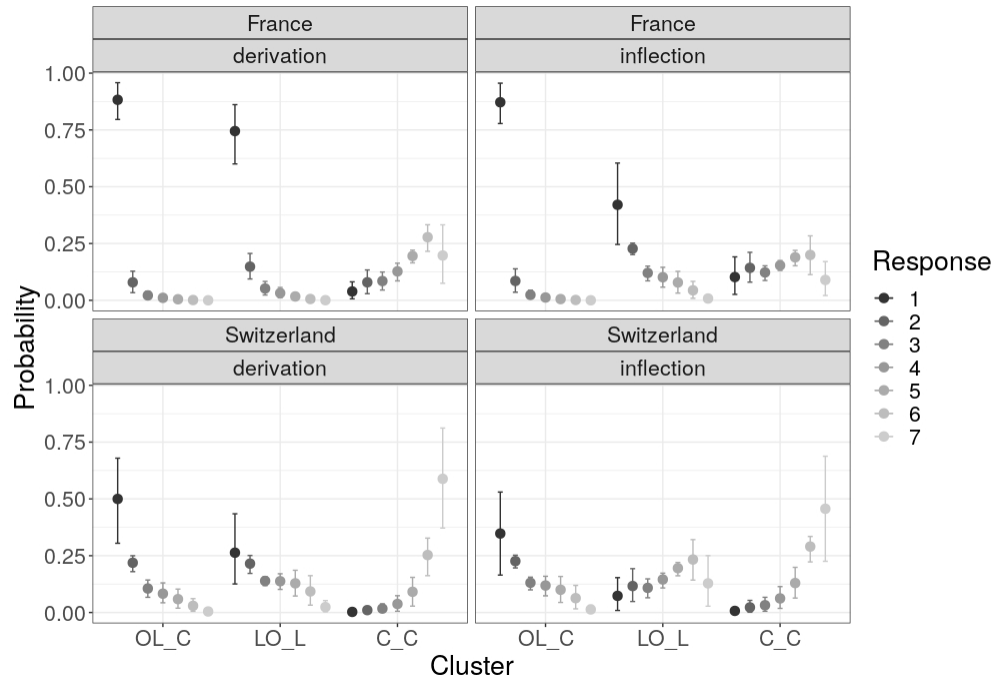


Figure 2: Posterior distribution (mean and 95% CI) of each of the seven response categories as a function of Morphology, Cluster, and Origin

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 ($E(\mu_{\text{French, LO_L, inf}} - \mu_{\text{French, OL_C, inf}}) = 1.37$, $\text{CI} = [0.92, 1.78]$, $P(\Delta > 0) = 1$) and Switzerland ($E(\mu_{\text{Swiss, LO_L, inf}} - \mu_{\text{Swiss, CO_L, inf}}) = 1.11$, $\text{CI} = [0.63, 1.57]$, $P(\Delta > 0) = 1$). 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 ($E(\mu_{\text{French, C_C, inf}} - \mu_{\text{French, LO_L, inf}}) = 1.10$, $\text{CI} = [0.74, 1.45]$, $P(\Delta > 0) = 1$) and Switzerland ($E(\mu_{\text{Swiss, C_C, inf}} - \mu_{\text{Swiss, LO_L, inf}}) = 1.07$, $\text{CI} = [0.67, 1.48]$, $P(\Delta > 0) = 1$).

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.

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 shown in Figure 2. 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_{inf}, *LOL_{inf}, *CC_{inf}, *OLC_{der}, *LOL_{der}, *CC_{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_{inf}, *LOL_{inf}, *CC_{inf}, *CCC_{der}, *CC_{der}).⁸ Table 2 shows how the first grammar assigns different constraint violations for OL_C and LO_L in derived words. Table 3 shows how the two clusters are treated identically in derived words in the second grammar.

/kõsjɛʁʒi/	*OLC _{der}	*OLC _{inf}	*LOL _{der}	*LOL _{inf}	*CC _{der}	*CC _{inf}	Dep(V)
[kõsjɛʁʒi]			1				
[kõsjɛʁʒəi]							1

LO_-L_{derivation} : *concierge-rie* ‘caretaker’s lodge’

/sufli-ʁi/	*OLC _{der}	*OLC _{inf}	*LOL _{der}	*LOL _{inf}	*CC _{der}	*CC _{inf}	Dep(V)
[sufliʁi]	1						
[sufliʁəi]							1

OL_-C_{derivation} : *souffle-rie* ‘wind tunnel’

Table 2: Constraint violations for schwa-zero alternations in derived and inflected words, as a function of the nature of surrounding consonants (OL_C vs. LO_L), according to the weak lexical-phonology hypothesis (constraints *OLC_{der} and *LOL_{der} are distinct)

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_{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

⁸ 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_{der}) and a general markedness constraint that does not depend on morphological domains (*OLC) instead of two morphologically indexed markedness constraints (*OLC_{der}, *OLC_{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/	*CCC _{der}	*OLC _{inf}	*LOL _{inf}	*CC _{der}	*CC _{inf}	Dep(V)
[kõsjɛʁʒɪ]	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)
[suflɪ]	1					
[sufləʁi]						1

OL₋C_{derivation} : *souffle-rie* ‘wind tunnel’

Table 3: Constraint violations for schwa-zero alternations in derived and inflected words, as a function of the nature of surrounding consonants (OL_C vs. LO_L), according to the strong lexical-phonology hypothesis (there is a single *CCC_{der} for derived words)

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). Convergence of the chains on the posterior distribution was assessed using the Gelman-Rubin statistic: it was very close to 1 for all parameters,⁹ 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 4 and 5 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 5 but this is not a feature of the analysis. The more complex grammar in Table 4 was found to have a smaller deviation information criterion than the more simple one in Table 5 ($\Delta = -19.36$), indicating that it is a better model of the data. The predicted frequencies of schwa presence under the two grammars 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
*OLC _{der}	4.17	[3.90, 4.46]
*OLC _{inf}	3.77	[3.53, 4.03]
*LOL _{der}	3.46	[3.22, 3.70]
*LOL _{inf}	2.20	[2.00, 2.40]
Dep(V)	1.89	[1.74, 2.04]
*CC _{inf}	1.00	
*CC _{der}	0.65	[0.43, 0.87]

Table 4: Posterior distribution of the constraint weights (mean and 95% CI) in the grammar distinguishing *OLC_{der} and *LOL_{der}

⁹ 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/>.

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 4.

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

Table 5: Posterior distribution of the constraint weights (mean and 95% CI) in the grammar with *CCC_{der}

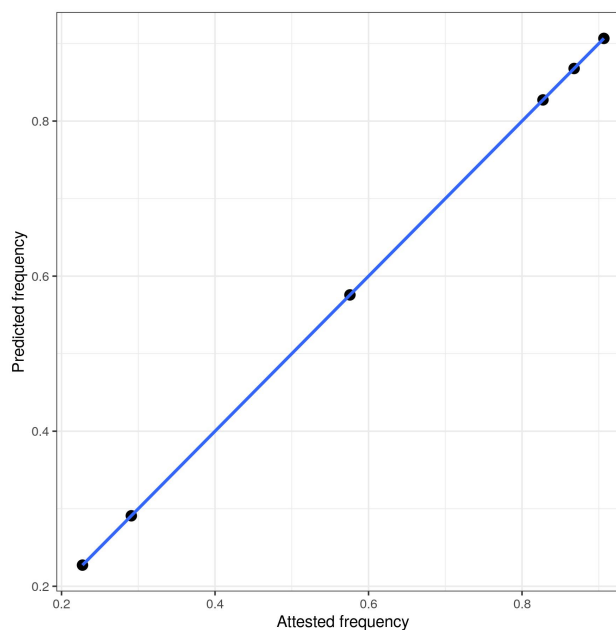


Figure 3: Grammar with *OL_C_{der} and *LO_L_{der}: model predictions vs. data

4 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.

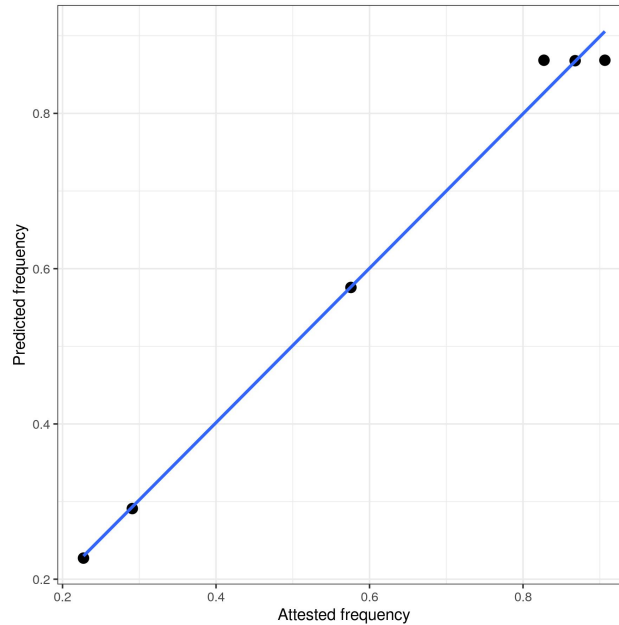


Figure 4: Grammar with *CCC_{der}: model predictions vs. data

References

- 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.
- Bayles, Andrew, Aaron Kaplan & Abby Kaplan (2016). Inter-and intra-speaker variation in French schwa. *Glossa: a journal of general linguistics* 1:1.
- Boersma, Paul & Bruce Hayes (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32, 45–86.
- 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.
- Bürkner, Paul-Christian (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80:1, 1–28.
- 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.
- 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.
- Clements, George N (1990). The role of the sonority cycle in core syllabification. *Papers in laboratory phonology* 1, 283–333.
- Côté, Marie-Hélène (2001). *Consonant cluster phonotactics: a perceptual approach*. Ph.D. thesis, MIT, Cambridge, MA.
- Côté, Marie-Hélène (2012). Laurentian French (Quebec). Extra vowels, missing schwas and surprising liaison consonants. Gess, Randall, Chantal Lyche & Trudel Meisenburg (eds.), *Phonological Variation in French: Illustrations from three continents*, John Benjamins, Amsterdam/Philadelphia, 235–274.
- Dell, François (1976). Schwa précédé d'un groupe obstruante-liquide. *Recherches linguistiques Saint-Denis* :4, 75–111.
- 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*. Hermann, Paris, 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.
- Flemming, Edward (2002). *Auditory Representations in Phonology*. Routledge, New York.
- Franke, Michael & Timo B Roettger (2019). Bayesian regression modeling (for factorial designs): A tutorial.

- Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari & Donald B Rubin (2013). *Bayesian data analysis*. Chapman and Hall/CRC, New York.
- Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a maximum entropy model. Spenader, Jennifer, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, Stockholm University, Department of Linguistics, Stockholm, 111–120.
- Grammont, Maurice (1914). *Traité pratique de prononciation française*. Delagrave, Paris.
- Hambye, Philippe & Anne Catherine Simon (2012). The variation of pronunciation in Belgian French. Gess, Randall, Chantal Lyche & Trudel Meisenburg (eds.), *Phonological Variation in French: Illustrations from three continents*, John Benjamins, Amsterdam/Philadelphia, 129–149.
- Hansen, Anita Berit (2012). A study of young Parisian speech: Some trends in pronunciation. Gess, Randall, Chantal Lyche & Trudel Meisenburg (eds.), *Phonological Variation in French: Illustrations from three continents*, John Benjamins, Amsterdam/Philadelphia, 151–172.
- Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379–440.
- Kawasaki-Fukumori, Haruko (1992). An acoustical basis for universal phonotactic constraints. *Language and Speech* 35, 73–86.
- Kruschke, John K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press, 2nd edn.
- 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.
- LimeSurvey (2012). LimeSurvey: An open source survey tool. URL <http://www.limesurvey.org>.
- Myers, James (2017). Acceptability judgments. *Oxford Research Encyclopedia of Linguistics*, Oxford University Press.
- Plummer, Martyn (2016). *rjags: Bayesian Graphical Models using MCMC*. URL <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, URL <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*. Ph.D. thesis, University of Geneva.
- Racine, Isabelle & Helene Andreassen (2012). A phonological study of a Swiss French variety: data from the canton of Neuchâtel. Gess, Randall, Chantal Lyche & Trudel Meisenburg (eds.), *Phonological Variation in French: Illustrations from three continents*, John Benjamins, Amsterdam/Philadelphia, 211–233.
- 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.
- Schütze, Carson T (2016). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Language Science Press, Berlin.
- Schütze, Carson T. & Jon Sprouse (2013). Judgment data. Podesva, Robert J. & Devyani Sharma (eds.), *Research methods in linguistics*, Cambridge University Press, Cambridge, 27–50.
- Smith, Brian W & Joe Pater (2020). French schwa and gradient cumulativity. *Glossa: a journal of general linguistics* 5:1.
- Smith, Jennifer L & Elliott Moreton (2012). Sonority variation in Stochastic Optimality Theory: Implications for markedness hierarchies. Parker, Steve (ed.), *The Sonority Controversy*, De Gruyter Mouton, 167–194.