Variable schwa realization in Laurentian French: A MaxEnt grammar approach

James White (UCL)
Suzanne Robillard (University of Ottawa)
My co-author

Suzanne Robillard
(University of Ottawa)
0. Preliminaries
Variable schwa realization in French

• French ‘schwa’ may be optionally realized in certain contexts:

  Annie le salut. [aniləsaly] ~ [anil_saly]
  ‘Annie greets him.’
  (from Côté 2000)

• Epenthesis or deletion (or a combination)?
  • Differs depending on account.
  • In most cases, it is not crucial for this work, so I will refer to it simply as ‘realized’ or not.
Multiple schwa sites in a row

- In this study, we focus on strings with multiple schwa sites in a row.
  - **Schwa site** = location where a schwa could appear (because it does so in other cases/contexts).
  - Our cases mostly involve **clitics**, which is where the **most variation** is reported.
  - Since each schwa may be realized (or not), the strings have multiple (logically) possible realizations.

- E.g.: *...je te rappelle... ‘...I remind you...’* has four possible realizations:
  
  [ʒɛtʁapɛl], [ʃtɛtʁapɛl], [ʒetʁapɛl], [ʃtʁapɛl]
Variable schwa realization in French

- Realization of schwa is driven in part by social factors, but also very much by phonological factors.
- The literature looking at the phonology of French schwa is immense.
  - But previous work tends to focus on:
    - Accounting for when schwas are obligatory, forbidden, or optional.
    - Single schwa sites.
- Little work has aimed to account for the relative likelihood of variants within a phonological model.
  (but see Côté 2007, Pater et al. 2012)
Why limited account of the variation?

• Some have dismissed the notion that the variation is phonologically principled. (e.g. Pulgram 1961)

• Others have pointed out that the variation is principled, without an explicit model of it:
  • “These segmental factors interact with each other in complex ways. As a general rule, factors... have a cumulative effect on the likelihood of schwa insertion and retention: the more such factors are present, the less probable schwa insertion/retention is.” (Côté 2000, p. 133)

• An explicit model requires:
  1. Extensive data from a spoken corpus.
  2. A probabilistic phonological framework.
Goals and roadmap

• French schwa in clitics is a notoriously tricky issue.

• **Goal**: Bring a new form of evidence to the table by:
  • Extracting cases of multiple schwa sites from a naturalistic speech corpus.
  • Using a probabilistic phonological framework (MaxEnt).
  • Compare syllable-based and licensing-by-cue accounts on their ability to account for variation in the data.

• Contribution to the debate on whether syllables are needed or useful in phonotactics.
  • French schwa.
  • Phonology in general.
1. Corpus details
Corpus details

- *Corpus français de l’Outaouais au nouveau millénaire: milieu scolaire et milieu social*[^1]
  - Housed at the Sociolinguistics Lab, University of Ottawa.

[^1]: Poplack & Bourdages 2005
Corpus details

- We focused on:
  - 44 high school students in the Ottawa-Gatineau region, living in Québec (native speakers of French).
  - Casual speech.
  - ~ 66 hours of recorded speech in our subset.

- Extracted all cases of 2 or more schwa sites in a row.
- 3 native speakers coded for:
  - Schwa realization (1, 0, ambiguous).
  - String identity.
  - Surrounding phonological context (several segments before/after schwa sites).
Corpus details

• Excluded:
  • Cases that were too ambiguous to code.
  • Cases with pauses/major breaks between schwa sites.
  • Words with a schwa-like vowel that was always realized (e.g. *devoir*, *besoin*).
  • Left us with 2694 strings.

• Breakdown:

![Pie chart showing breakdown of schwa sites: 2 schwa sites, 3 schwa sites, and 4 schwa sites.](image-url)
"…ça fait quoi genre que je l’aie?"
"…what does it matter if I have it?"

"…puis la seule année que je suis t allée...
"...and then the only year that I went...

"…français, ça fait quatre ans que je l’ai aussi.”
"…French, I’ve had it for four years now too.”

"peut-être m’a me changer d’idée mais que je travaille là.”
“maybe I’ll change my mind once I start working, eh.”

Note: Codes refer to speaker number and line number in the Corpus du français parlé au nouveau millénaire: milieu scolaire et milieu social (Poplack & Bourdages 2005). Examples are reproduced verbatim from speaker utterances.
2. To syllabify, or not to syllabify...
The syllable

See Bridget Samuels’s talk yesterday!
The syllable

- Research on the syllable in phonology goes back to at least the 19\textsuperscript{th} century. (see Goldsmith 2011 for a review)
  - Tossed out in SPE.
  - Resurgence in the 1970s, 80s, and beyond.

- Arguments in favour: (e.g. Selkirk 1982)
  - More elegant/uniform analysis.
  - Elegant account of cross-linguistic markedness tendencies.
  - Accounts for linguistic/meta-linguistic phenomena that seem to affect syllable constituents.
Syllable accounts of French schwa

- **Basic idea**: (e.g. Pulgram 1961, Morin 1974, Anderson 1982, Noske 1982, Tranel 1987)
  - Fully parse the segments into syllables.
  - Restrictions on possible syllable structure.
  - Use [ə] whenever segments cannot be syllabified.

- Côté’s (2000) overview of a syllable-based account:
  1. Max 1 coda C. Complex onsets allowed.
  2. Cs may not resyllabify across boundaries.
  3. Extrasyllabic Cs allowed word-finally.
  4. Extrasyllabic C allowed phrase-initially (or word initially for R + glide sequences).
**Extrasyllabicity**

Je me fais...
“I make myself…”

Ban on resyllabification across a boundary

**garderie**
“daycare”

/gaʁd + ʁi/

[ɡaʁdəʁi] (obligatory [ə])

*[ɡaʁd.ʁi] = illicit coda*

*[ɡaʁ.dʁi] = resyllabification*
Multiple schwas with clitics

- But when it comes to variation in the clitic groups, the syllable approach *undergenerates* (Côté 2000)

  *Il faut que je te le dise* ‘I have to say it to you’

  Attested possibility: [ilfokʃtəldziz]

  * [ilfokʃtəldziz] (* due to illicit coda kʃo*)
  * [ilfokʃtəldziz] (* due to illegal resyllabification and illicit onset *)
  * [ilfokʃtəldziz] (* illegal extrasyllabic C *)

→ Schwa (wrongly) predicted to be *obligatory* after [ʃ] in [ilfokʃətəldziz].
General arguments against syllable-based accounts of phonotactics

• Different parts of the phonology suggest different syllabifications.
  • Blevins (2003) on Klamath and Lithuanian.
• Syllable-based accounts are unable to account for the facts.
  • Côté (2000) on French schwa.
  • Blevins (2003) on Yowulmne Yokuts.
• Syllable accounts almost always require additional string-based phonotactic constraints.
  • Almost all (perhaps all?) syllable-based accounts can be recast with string-based constraints; the reverse is not always true.
• Does this make syllables redundant in phonotactics? (Côté 2000, Blevins 2003)
An alternative: Licensing by cue

- **Basic idea:** Consonants are licensed in contexts where their perceptual cues are robust. (Jun 1995, Steriade 1999, 2000; Côté 2000, Albright (this conference))

- The grammar penalizes consonants that occur in positions where their perceptual cues are weak.

- Formalized using a set of string-based (or sequential) constraints.
String-based constraints necessary

• People seem to agree that sequential constraints are necessary.
  • E.g., Selkirk (1982) expresses the need for additional ‘collocational restrictions’ several times.

• So the real comparison seems to be:
  Syllables + Sequential vs. Sequential only

• The question becomes:
  How much benefit are we getting from the syllables? Are they still adding any explanatory power?
Hypotheses to test

1. Syllables alone provide a better fit to the data than sequential constraints alone.

2. The two accounts are equal in their ability to account for the data; syllables provide a simpler, more elegant account.

3. Sequential constraints + syllables results in a significantly better fit than sequential only.

4. Adding syllables results in no improvement over just sequential constraints.
Strategy

1. Find the best possible sequential-only model.

2. Find the best possible syllable-only model.

3. Put the constraints together and let them fight it out.
3. Analysis
Framework: MaxEnt models

• General statistical classification model used in many fields.

• Implemented here as a probabilistic model of phonological grammar, specifically a type of Harmonic Grammar. (e.g. Goldwater & Johnson 2003, Wilson 2006, Hayes & Wilson 2008, White 2013, Becker)
  • Constraint-based (constraints and violations provided).
  • Constraints associated with weights, which are additive.
  • Has the “ganging” property.

• For a given input, the probability of an output candidate is determined by:
  • Raising $e$ to the summed weighted violations.
  • Taking the candidate’s proportion of the total for all candidates.
MaxEnt learning

- Model is given:
  - a set of input/output training data
  - a set of candidates
  - a set of constraints
  - violations

- Model’s **learning objective**: find the set of constraint weights that maximizes the likelihood of the observed data.

- Also, includes a (weak) smoothing prior.
  - All else being equal, prefers weights to be near 0 and spread out among the constraints.
Simple ‘every other’ account?\textsuperscript{1}

1. **Law of two consonants**: Schwa is pronounced in any possible site that is preceded by CC.

2. **Law of alternating schwas**: Schwa is pronounced in every other possible site.

• Predictions for doubles:
  • $C\_C\_CV \rightarrow CC\_C\_CV$ (10)
  • $\text{V} \ C\_C\_CV \rightarrow \text{V}C\_C\_CV$ (01)

\textsuperscript{1} See Côté 2000; see also Leray 1930, Fouché 1959, Grammont 1961, Selkirk 1978, Anderson 1982, Morin 1987
Is it just every other?

Doubles by preceding context

- **B** = Preceding pause
- **C** = Preceding C
- **V** = Preceding V

- **01** and **10** more common than **11** or **00**.
- More **11** and **10** after **CC**.
- More **01** and **00** after **VC**.

**Conclusion:** These ‘laws’ reflect accurate tendencies, but the details are far from that simple.
Licensing-by-cue account (based on Côté 2000)
(See also Jun 1995; Steriade 1999, 2001; Wright 2004)

- **Basic idea**: Realize schwa whenever it’s needed to enhance the cues of a C.

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Violated by</th>
</tr>
</thead>
<tbody>
<tr>
<td>C ↔ V</td>
<td>C not adj. to V.</td>
</tr>
<tr>
<td>C → V</td>
<td>C not followed by V. (no effect)</td>
</tr>
<tr>
<td><strong>Stop</strong> ↔ V</td>
<td>Stop not adj. to V.</td>
</tr>
<tr>
<td><strong>Stop</strong> → V</td>
<td>Stop not followed by V.</td>
</tr>
<tr>
<td><strong>Stop[−cont]</strong> ↔ V</td>
<td>Stop not adj. to V &amp; not followed by a cont.</td>
</tr>
<tr>
<td><strong>Stop[−cont]</strong> → V</td>
<td>Stop not followed by a continuant.</td>
</tr>
</tbody>
</table>

(Consonants cues are strongest adjacent to vowels, esp. before vowels.)

(Stops, especially, are best cued in vowel transitions and stop releases.)

(Stop bursts are poorly cued before non-continuants, in particular.)
Licensing-by-cue account (based on Côté 2000)

Constraints: Violated by:

\[ \text{NONSIB} \leftrightarrow \text{V} \quad \text{Non-sibilant obstruents not adj. to V.} \]

\[ \text{NONSIB} \rightarrow \text{V} \quad \text{Non-sibilant obstruents not adj. to V.} \]

(Sibilants have strong internal cues.)

cf. the unexpected acceptability of sT clusters in many languages.

\[ \text{V} \leftarrow \text{L/R} \quad /l/ \text{ or } /R/ \text{ not preceded by V.} \]

(Approximants strongly coarticulated on preceding V.)

Note: Metathesis in Québec French – L/R are preferred targets (Côté 2012).

regarder ‘look’ \[ \text{[ʁəɡəʁdɛ]} \rightarrow \text{[əɾɡəɾdɛ]} \]

le guide ‘the guide’ \[ \text{[ɛɡɪd]} \rightarrow \text{[əɾɡɪd]} \]

*NO Nasal followed by an obstruent.

(Nasal place cues are weak before obstruents.)
Syllable-independent SSP

• Violation for any local sonority peak that is not a vowel. (Côté 2000)

   "Bien, je le fais pas …"

Sonority coded following Clements 1990 (i.e. stops = fricatives).
Syllable-based account

Tried two general approaches

1. Syllabify: Max 1 coda C. Otherwise, onset.
2. Penalize sonority-violating syllabifications.

(*FALL, *PLATEAU, SSP, NOCODA, SYLLCONTACT, LIC-L/R(coda))

1. Syllabify: Max 1 coda C. Only licit onsets. Other Cs extrasyllabic.
2. Penalize non-phrase-initial extrasyllabic Cs.

(EXHAUSTIVE, NOCODA, SYLLCONTACT, LIC-L/R(coda))
Both accounts

- *ə
- *əC₀ə
## Learned grammars

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Licensing-by-cue only</th>
<th>Combined</th>
<th>Syllables only</th>
</tr>
</thead>
<tbody>
<tr>
<td>C $\leftrightarrow$ V</td>
<td>0.48</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>STOP $\leftrightarrow$ V</td>
<td>1.17</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>STOP $\rightarrow$ V</td>
<td>0.64</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>STOP[$\text{--cont}$] $\leftrightarrow$ V</td>
<td>0.64</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>STOP[$\text{--cont}$] $\rightarrow$ V</td>
<td>0.75</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>NONSIB $\leftrightarrow$ V</td>
<td>0.31</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>NONSIB $\rightarrow$ V</td>
<td>2.23</td>
<td>2.19</td>
<td></td>
</tr>
<tr>
<td>V $\leftrightarrow$ LIQ</td>
<td>1.27</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td>*NO</td>
<td>2.39</td>
<td>2.30</td>
<td></td>
</tr>
<tr>
<td>NON-syllabic SSP</td>
<td>2.90</td>
<td>2.58</td>
<td></td>
</tr>
<tr>
<td>*ə</td>
<td>1.22</td>
<td>1.64</td>
<td>0.49</td>
</tr>
<tr>
<td>*əC₀ə</td>
<td>1.55</td>
<td>1.27</td>
<td>1.58</td>
</tr>
<tr>
<td>EXHAUSTIVE</td>
<td>0.88</td>
<td></td>
<td>1.77</td>
</tr>
<tr>
<td>SYLLCONTACT</td>
<td>0.05</td>
<td></td>
<td>0.65</td>
</tr>
<tr>
<td>NOCODA</td>
<td>0.17</td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td>LIC(LIQ, CODA)</td>
<td>0</td>
<td></td>
<td>0.98</td>
</tr>
</tbody>
</table>
Model comparison

- An advantage of using an explicit, probabilistic model is that we can compare models making different assumptions on their fit to the data.

- **(Log) Likelihood:**
  - Measure of fit to the data.
  - Which model results in the greatest likelihood of the observed data (therefore minimizing the likelihood of unobserved data)?

- **AIC/BIC:**
  - Measure of likelihood, with a penalty for model complexity (= number of constraints).
Model comparison

Higher (less negative) LogLikelihood = better fit.
Lower AIC/BIC suggests a better model.

**Licensing-by-cue only**
LogLikelihood = −2422
AIC = 4869
BIC = 4940

**Syllable-based only**
LogLikelihood = −2934
AIC = 5880
BIC = 5915

**Combined**
LogLikelihood = −2382
AIC = 4794
BIC = 4883

p < .01
How are the models doing?

• Classify the segments into broad types based on manner classes (stop, nasal...).
  • Je me fais... $\rightarrow /S_N_FV$

• Take cases with at least 8 observations (including all possible schwa patterns, e.g. 11, 10, 01, 00).

Licensing-by-cue only
\[ r^2 = .61 \]

Syllable-based only
\[ r^2 = .40 \]

Combined
\[ r^2 = .62 \]
How are the models doing?

- Others sources of variance not accounted for by the model:
  - Morphosyntactic effects.
  - Lexical/type effects.
  - Prosodic effects.
  - Missing phonological generalizations.
  - Social factors.
  - Speech rate.
  - Individual speaker variation.
  - Noise.
4. Conclusions
Hypotheses to test

1. Syllables alone provide a better fit to the data than sequential constraints alone.

2. The two accounts are equal in their ability to account for the data; syllables provide a simpler, more elegant account.

3. Sequential constraints + syllables results in a significantly better fit than sequential only.

4. Adding syllables results in no improvement over just sequential constraints.
Concluding thoughts

• Syllables add a ‘significant’ benefit to the model, but the magnitude is small. **Weak support** for the syllable.
  • Supports a model in which syllables only have a weak influence on phonotactics.
  • But see Albright’s talk for evidence that we might need both.

• One consideration: the ‘complexity’ component of the model comparison only involved # of constraints.
  • But syllabification requires additional model complexity not taken into account here.

• We can also use this approach to test other assumptions in the model as well.
  • E.g. theories about sonority scales and sonority sequencing.
‘Poverty of the stimulus’ experiment in the L1?

• Perhaps...
  • The types in our data follow a Zipfian distribution.
    • Few types with high frequency.
    • Many types with very low frequency (and occurring only once in the corpus).
  • In ~66 hours of speech: Only 249 cases of 3 schwas, 23 cases of 4 schwas, 0 cases of 5 or more.
    • These are possible: ...que je ne te le redemanderais pas. ‘...that I wouldn’t ask you it again.’ (7 schwa sites, $2^7 = 128$ logically possible realizations)

• We have a predictive model of speaker behaviour, but would they follow the input available? (e.g. Hayes et al. 2009; Becker et al. 2011, 2012; Hayes & White 2013)
Acknowledgments

• Many thanks to:
  • Shana Poplack for access to the corpus.
  • Nathalie Dion for assistance with the corpus.
  • Myriam Lapierre and Sarah Fitzpatrick for assistance with coding.
  • Audience at the London Phonology Seminar for discussion.

• Funded by a seed funding grant from the Faculty of Arts, University of Ottawa.
Differential behaviour of ‘me’ and ‘le’

Doubles according to second word of string

- le
  - 00: [Blue]
  - 01: [Teal]
  - 10: [Red]
  - 11: [Brown]

- me
  - 00: [Blue]
  - 01: [Teal]
  - 10: [Red]
  - 11: [Brown]