Lexically conditioned variation in Harmonic Grammar

Joe Pater, University of Massachusetts, Amherst

with the collaboration of

Paul Boersma, University of Amsterdam
Andries Coetzee, University of Michigan
Isabelle Racine, Université de Genève
Lexically conditioned variation:

(1) A variable process whose probability of application is conditioned by the lexicon

Three extremely well-studied variable phonological processes that display lexical conditioning:

(2) English vowel reduction ([Fidelholtz 1975 et seq.])
   embrace > emporium

   English t/d-deletion ([Labov et al. 1968 et seq.])
   jus(t) > bus(t)

   French ‘schwa’ deletion ([Dell 1973 et seq.])
   s(e)maine > s(e)mestre
Lexically conditioned variation is a particular focus of usage-based and exemplar-based approaches to phonology:

(3) 

*English:* 

*French:* 

Key observation (Dell 1973, Fidelholtz 1975)

(4) Word frequency correlates positively with frequency of application of these variable processes
Argument for usage-based phonology:

(5) Frequency of use ~ frequency of application correlation not captured by other models of phonology

A definition of the domain of phonological theory:

(6) Native speaker knowledge of the sound patterns of language

(7) The acquisition of that knowledge

It remains unclear whether the correlation in (5) is in (6), and therefore whether phonology should account for it (that is, the correlation could well be a product of use, as per Bybee)
However, since speakers do \textit{at least} know that words differ in their probability of application of variable processes, and since they learn this probability, an adequate theory of phonology must provide a way of representing lexically conditioned variation, and its associated learning theory should provide an account of its acquisition.

Current phonological theory provides the needed ingredients (HG = Harmonic Grammar, Smolensky and Legendre 2006)

(8) \textit{Theory of variation} (Boersma 1997 \textit{et seq}.)
  \begin{itemize}
  \item Stochastic OT/Noisy HG
  \item Theory of lexical conditioning (Pater 2000 \textit{et seq}.)
  \item Lexically indexed constraints
  \item Theory of learning (Boersma 1997, Boersma and Pater 2008)
  \item Gradual learning algorithm for OT/HG
  \end{itemize}
Coetzee and Pater (2008) illustrate this approach to lexically conditioned variation with English $t/d$-deletion

**Today:**

§1. Introduction to the theories of grammar (HG), variation (noisy HG) and learning (HG-GLA) I’ll be assuming, with French $e$-deletion as the illustrative example

§2. Presentation of data on lexical conditioning of French $e$-deletion, especially from Racine’s (2007) study of native speaker judgments

§3. Further analysis and learning simulations

§4. Other HG approaches to lexical conditioning of $e$-deletion

§5. Broader implications
1. HG, variation, exceptionality, and learning

In HG (Legendre, Miyata and Smolensky 1990, Smolensky and Legendre 2006) well-formedness, or Harmony \((H)\) of representation \((R)\), is the sum of its weighted constraint violations/scores (constraint score vector \(C\) and weights \(W\))

\[
H(R) = \langle C, W \rangle
\]

Prince and Smolensky (1993/2004:236) note that HG Harmony can be used to choose the optimum in a candidate set:

\[
\begin{array}{c|c|c}
\text{Input-1} & \text{CON-1} & \text{CON-2} \\
\hline
\text{Output-11} & 2 & 1 \\
\text{Output-12} & -1 & -1 \\
\hline
\end{array}
\]
Applied to e-deletion (where [ə] is the vowel represented by orthographic e, which varies from dialect to dialect)

\[
\begin{array}{|c|c|c|}
\hline
\text{səmən} & *[ə] & \text{MAX} \\
\hline
\text{səmən} & -1 & -2 \\
\hline
\text{ʃəmən} & -1 & -1 \\
\hline
\end{array}
\]

Constraint definitions:

\[
\begin{array}{c}
(12) \\
*[ə] \\
\text{Assign a score of } -1 \text{ for each vowel } [ə] \\
\text{MAX} \\
\text{Assign a score of } -1 \text{ for each vowel in the input that lacks an output correspondent}
\end{array}
\]
Boersma’s (1997 et seq.) theory of variation:

(13) **Noisy evaluation**

Each time the grammar is used to evaluate a candidate set, add a Gaussian random variable $N_k$ to each constraint weight $w_k$

Applied to e-deletion (post-noise values in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>*[ə]</th>
<th>MAX</th>
<th></th>
<th>*[ə]</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>səmɛn</td>
<td>1 (1.1)</td>
<td>-1.1</td>
<td>səmɛn</td>
<td>1 (0.9)</td>
<td>-1</td>
</tr>
<tr>
<td>səmɛn</td>
<td>-1</td>
<td>-0.9</td>
<td>səmɛn</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>ɛsɛmɛn</td>
<td>-1</td>
<td>-0.9</td>
<td>ɛsɛmɛn</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

(see Andreassen’s 2004 similar ‘floating constraint’ analysis of French schwa)
The distribution of [ə] is subject to categorical restrictions (Martinet 1969, Dell 1973; cf. Walker 1993):

(15) a. Word-initial deletion never occurs after a tautomorphemic cluster (*gr(e)nouille, *Br(e)ton)

b. [ə] occurs only after consonants, and is productively deleted after vowels (vous joueriez vs. vous parleriez from jouer and parler)

A constraint satisfied by the presence of [ə] (assuming e-deletion results in an empty-headed syllable, as in Rialland 1986, though cf. Steriade 2000 – the constraint may be better defined in articulatory or perceptual terms):

(16) LICENSE-C
    Assign a violation mark to a consonant in syllable without a vocalic nucleus
In HG, the difference between the C_ and CC_ environments can be analyzed as the difference between one or two violations of LICENSE-C
(see Pater, Bhatt and Potts 2007 for critical assessment of extant arguments for strict domination, and discussion of outstanding issues)

<table>
<thead>
<tr>
<th></th>
<th>*[ə]</th>
<th>LICENSE-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>səmɛn</td>
<td>1.5</td>
<td>1</td>
</tr>
<tr>
<td>səmɛn</td>
<td>−1</td>
<td>−1.5</td>
</tr>
<tr>
<td>ʃsɛn</td>
<td>−1</td>
<td>−1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>*[ə]</th>
<th>LICENSE-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>brɛtɛ̃</td>
<td>−1</td>
<td>−1.5</td>
</tr>
<tr>
<td>ʃ brɛtɛ̃</td>
<td>−2</td>
<td>−2</td>
</tr>
</tbody>
</table>

And the restriction of *[ə] to post-consonantal position can be understood as a ‘gang effect’ between LICENSE-C and MAX, as in the following Richness of the Base tableau
(see Dell 1973 for a rule-based analysis of post-V deletion, including a Richness of the Base approach to that phonotactic gap, and an MSC against initial *[ə]*)
This analysis is intermediate between the analyses of ‘schwa’ as epenthetic (Martinet 1969, Côté 2006) and underlying:

(20) Underlying ‘schwa’ can surface only when it serves to satisfy a markedness pressure (e.g. LICENSE-C)
To account for exceptions to e-deletion, I use a lexically specific version of MAX
(lexically specific faithfulness introduced to OT in Pater 2000 to account for lexical conditioning of English vowel reduction; see also esp. Ito and Mester 1995 et seq. for applications to lexical conditioning in Japanese)

(21)  **MAX-L**

Assign a violation mark for every input segment that belongs to lexical item $L$ and that lacks an output correspondent

A word like *belon* (a type of oyster), whose first vowel apparently never undergoes deletion (Angoujard 2006, Eychenne 2007), is distinguished from *semaine* in terms of the relative ranking of the **MAX-L** constraint indexed to each of them.
How does an analyst find a weighting for the constraints that generates the patterns in a given language? How does a learner find such a weighting?

For analysis of categorical patterns in HG, and comparison of typology in HG and standard OT:

(22) HaLP (Potts, Becker, Bhatt, Pater 2007)
     OT-Help (Becker, Pater, Potts 2007)

For patterns of variation, and simulation of human learning:
(see Jäger to appear for an earlier similar proposal)

(23) HG-GLA in Praat (Boersma and Weenink 2007)
The HG-GLA is a very simple learning algorithm, which adapts a machine learning procedure (*perceptron*, *stochastic gradient ascent*) to the learning of HG grammars, and closely resembles Boersma’s (1997 *et seq.*) OT-GLA. Unlike the OT-GLA for stochastic OT, it is guaranteed to converge on correct weightings for categorical data:

\[(24) \quad \text{HG-GLA convergence} \]
\[
\text{Fischer 2005, Boersma and Pater 2008}
\]

*OT-GLA/stochastic OT non-convergence*

\[
\text{Pater 2008}
\]
I provided the HG-GLA with the following distribution of data:

(25) \( \text{semaine}_{L1} \) 50% deletion  \
     \( \text{belon}_{L2} \) 0% deletion  \
     \( \text{breton}_{L3} \) 0% deletion

Initial constraint weights:  
(see Jesney and Tessier 2007 on high Mark and low Faith in HG learning)

(26) LICENSE-C, *[ə] 50  \
     MAX, MAX-L1, MAX-L2, MAX-L3 0

Settings were Praat’s defaults for these parameters:

(27) Noise SD = 2.0  
     Plasticity = 1.0 (decreased by a factor of 0.1 after each of 4 sets of 100,000 learning data)
The grammar model was run in Linear OT mode (Keller 2006), under which it interprets all sub-zero weights as zero.

(28) LICENSE-C 50.3
    *[^e]  49.7
    Max-belon 17.0
    Max-breton 0.0
    Max  -0.3
    Max-semaine -17.3

Probability of e-deletion with grammar in (28), noise 2.0

(29) a. semaine 50% belon 0% breton 0%
    b. eCVC 100% CVeCVC 100%
3. Lexically conditioned e-deletion: data

The simplified data in the previous section could be handled by any standard theory of generative phonology (e.g. Dell 1973)

(30)  *әCVC due to a Morpheme Structure Constraint

Dell (1973) provides an example of the type of lexical conditioning that cannot easily be handled by the theory he was working with, but which falls out from the one in §2.
In general, schwa deletion is blocked by a preceding consonant (judgments from a native speaker in Nantes):

(31)  \textit{la s’maine} ✓ \textit{une s’maine} ??

Dell writes his rule “\textit{VCE}_1” to apply only after vowels. However, he notes that (p. 207 in 1980 English translation):

(32) Contrary to what we said above, in very rapid speech the schwa of a small number of words beginning with $\#C\_\_\_\text{can}$ be dropped even if the preceding word ends in a consonant: \textit{quelle semaine} is sometimes pronounced $[kɛlɛmɛn]$. ...The other words that have this property in our speech are \textit{je}, \textit{semelle}, \textit{cerise}, \textit{chemise}, \textit{fenêtre} and \textit{petit}...there are other words which always behave according to \textit{VCE}_1: \textit{semestre}, \textit{seringue}, \textit{chenille} (glosses and transcriptions omitted)
The generalization:

(33) The probability of schwa deletion is greater for words like *semaine* than for words like *semestre*

Further evidence of finer distinctions than optional application vs. non-application:

(34) a. Dictionaries find the two-way categorization inadequate (Walker 1996)

b. Corpus-based studies note that some words show greater frequency of deletion than others (e.g. Hansen 1994, Eychenne 2007, Eychenne and Pustka 2007)

c. Racine and Grosjean (2002) provide data from a production study showing a wide range of deletion frequencies across words
Dell (1973) notes that exceptions to schwa deletion tend to be rarely used or literary words (as well as proper names). He also claims to have systematically looked for phonological generalizations, and came up empty-handed (see also Walker 1996).

Is it “just” usage frequency that determines deletability of schwa?
Racine and Grosjean (2002) find a correlation between deletion and frequency ratings, but note that the correlation is not perfect.

According to Fougeron, Goldman and Frauenfelder (2003):

Overall the results do not support our hypothesis: lexical frequency and competition do not appear to influence strongly whether liaison and elision are realised or not.

(they also cite Adda-Decker et al. 1999 as having reached the same conclusion)

Therefore, not only do speakers need to learn whether a word has a “stable” or “unstable” e, but they need to learn its degree of stability, or *profil de maintien* (Walter 1977)
Racine’s (2007, in prep.) judgment study provides particularly convincing evidence of speakers’ knowledge of individual words’ *profil de maintien*

**Method:**

12 native-speakers from Nantes, France (and 12 from Neuchâtel, Switzerland)

All 2189 nouns containing schwa in BRULEX database (10 eliminated from analysis for being judged as [œ]/[ø])

Presented orthographically with preceding *la/le*

The list was presented twice, with an interval of several days; once all words were rated with *e* present, and once judged with it absent

Rated on a scale of 1 (infrequent) to 7 (very frequent); data here are the means over the 12 subjects for each word
Acceptability of deletion - retention

n=30

n=338
Frequency vs. deletability

\[ belon \quad -5.4 \quad semestre \quad -4.6 \quad semaine \quad 0.0 \]
4. Analysis of probabilistic lexical conditioning

Constraint for e-deletion with preceding consonant-final word:

(35)  *C.C.
       An unlicensed consonant must be preceded by a vowel

Probability of deletion in learning data:

(36)  \textit{la semaine}_L1 50\% \quad \textit{quelle semaine}_L1 10\%
       \textit{le semestre}_L2 10\% \quad \textit{quel semestre}_L2 0\%
       \textit{le breton}_L3 0\% \quad \textit{quel breton}_L3 0\%
       \textit{la belon}_L4 0\% \quad \textit{quelle belon}_L4 0\%

The markedness constraints again started out at 50, and the five faithfulness constraints (MAX, MAXL1...4) at zero
The constraint weights found by the learner:

(37) \*[^ə] & 50.3  
License-C & 49.7  
Max-belon & 15.0  
*C.C. & 4.9  
Max-semestre & 4.8  
Max-breton & 0.0  
Max & -0.3  
Max-semaine & -20.1

Deletion rates:

(38) la semaine & 52%  
le semestre & 11%  
le breton & 0%  
la belon & 0%  
\*CVC & 100%  
quelle semaine & 10%  
quel semestre & 1%  
quel breton & 0%  
quelle belon & 0%  
CV\*CVC & 100%
4. Alternative accounts in HG
4.1 Lexically scaled weights (≈ connectionist lexicon)

The numerically valued constraints of HG allow for views of lexical representation that diverge much further from “standard” generative phonology than lexically specific constraints.

Coetzee and Pater (2008) sketch one such alternative: that individual lexical items (or segments) bear numerical indices that rescale the basic weight of faithfulness constraints (see also Goldrick 2008).

For the word *semaine*, the value of the faithfulness constant would be relatively low, thus leading to variable deletion.
Lexical constants are more restrictive than lexically specific constraints, and faithfulness constraint scaling can be extended to register/style differences (Coetzee and Pater 2008; see van Oostendorp 1998 on register-based reranking of faithfulness).
4.2 Probabilistic allomophy (= exemplarist lexicon)

In OT, “allomorphy” is used to describe a case when the phonology chooses between two lexically supplied forms of a morpheme (e.g. Kager, Mascaró, Tranel).

If the lexicon provides in addition a default probability distribution over the allomorphs, then this could interact with the contextual probabilities provided by the grammar to yield the observed conditioning by lexicon and context.

One way of formalizing this, similar in spirit to lexically specific constraints, is with a set of “Use-/X/” constraints (Kager 1996, Boersma 1999, Apoussidou 2006).
5. Broader consequences of lexically conditioned variation

Lexically specific constraints, and the related approaches in HG discussed in §4, are highly interactive theories of the lexicon-grammar relationship.

This high degree of interaction seems to be required to account for the fine-grained consequences of the lexicon on probability of e-deletion.

Coetzee and Pater (2008) point out that lexically conditioned variation is problematic for some widespread assumptions about the relationship between the lexicon and the phonological grammar.
The hard question: what type of process is e-deletion?

(41) Lexicon $\rightarrow$ Early Phonology $\rightarrow$ Late Phonology $\rightarrow$ Phonetic Implementation

As Walker (1996) points out, e-deletion has contradictory properties under the view that processes at each of these levels have a set of fixed characteristics (a view most closely associated with classical Lexical Phonology; e.g. Kiparsky 1985)

(42) i. Has exceptions (Lex/EP)

ii. Sensitive to morphology (Lex/EP) (Dell 1973)

iii. Variable (LP/PI)

iv. Non-structure preserving, and may even lead to between category outcomes (LP/PI) (Fougeron and Steriade 1997, Barnes and Kavitskaya 2003; cf. Côté and Morrison 2007)
The fine-grained lexical conditioning of rate of variation in e-deletion (and the English processes mentioned in 2) seems to put the final nail in the coffin of the idea that processes at each of the levels in (41) have non-overlapping characteristics (see already Kiparsky 1993 on variation in OT).

The fact that this variable process is sensitive to morphology and the lexicon places it firmly within the domain of phonology, and makes it hard to imagine how one could maintain a “categorical vs. probabilistic” distinction for distinguishing phonology from phonetics.
References


