**Lexical propensities in variable phonology: corpus and experimental evidence from Slovenian and French**

Jesse Zymet, UCLA, jzymet@ucla.edu

April 21st, 2018

---

**1. Overview**

Theories disagree about how phonological variation is represented:
- Some encode morpheme-by-morpheme variation on binary scale (e.g. [+/- Rule]; e.g. Walker & Wiese 1999; Antilla 1997, Pater 2000, Becker 2009, Jurgec 2016). Others predict morphemes can pattern on propensities spectrum ([0.7 Rule]; Moore-Cantwell & Pater 2016, Smolensky & Goldrick 2016, Zaraw & Hayes 2017).

My claim: morphemes can have differing propensities to trigger/undergo variable process —what I call **LEXICAL PROPENSITIES**. Favors theories coding morphemes on a spectrum.

**This poster:**
- Corpus investigations into Slovenian palatalization and French liaison shows morphemes can have different propensities to trigger/undergo these processes;
- Experimental investigation suggests speakers track these propensities.

**2. Lexical propensities to trigger/undergo palatalization in Slovenian**


Like Jurgec, I extracted velar-final stems & suffixes from Dictionary of Standard Slovenian (Bajec et al. 2000), online corpus with over 110,000 different word types.
- 9 common suffixes show propensity to trigger palatalization; around 200 do not.

Concatenated stems with 9 suffixes, extracted token counts of each word from Gigafida (Logar-Berginc et al. 2012), corpus containing ~1 billion word tokens.
- Resulted in list of ~4 million word tokens.

Suffxes pattern on a propensity spectrum to trigger palatalization:

```
Suffix-specific rates of triggering palatalization across stems

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>-veze</td>
<td>0.1</td>
</tr>
<tr>
<td>-ifg</td>
<td>0.2</td>
</tr>
<tr>
<td>-je</td>
<td>0.3</td>
</tr>
<tr>
<td>-nat</td>
<td>0.4</td>
</tr>
<tr>
<td>-ina</td>
<td>0.5</td>
</tr>
<tr>
<td>-n</td>
<td>0.6</td>
</tr>
<tr>
<td>-k</td>
<td>0.7</td>
</tr>
<tr>
<td>-ina</td>
<td>0.8</td>
</tr>
<tr>
<td>-ina</td>
<td>0.9</td>
</tr>
</tbody>
</table>
```

Jurgec (2016): process is phonologically conditioned, and applies in restricted set of suffixes.
- Gives account mostly of phonologically conditioning: whether suffix is [+init], target consonant identity, whether stem contains another velar or postalveolar, etc.
- Jurgec finds that these suffixes have different rates, but encodes them with [+ Pal’n]. Would model be improved if suffixes were encoded as idiosyncratic, gradient palatalizers (0.7 Pal’n), (0.3 Pal’n) rather than all as categorical palatalizers [+ Pal’n]?
- Compared logit model with only Jurgec’s phonological factors + log(wordfreq)...
- against mixed model with same main effects + suffix identity as random effect.
- **AIC(categorical)=3167.4 > AIC(suffix propensities) = 3167.4**

Strikingly suggests that these suffixes have lexical propensities.

Beyond Jurgec: Of 259 common stems, 127 are undergoers, 30 are nonundergoers, 102 have medial rates:

```
# stems with particular palatalization rate

<table>
<thead>
<tr>
<th>Rate</th>
<th># Stems</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>(127)</td>
</tr>
<tr>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>
```

Would model be improved if we also encoded stems with differing propensities?
- Compared mixed model containing Jurgec’s factors, log(wordfreq), and suffix identity...
- against mixed model with the same factors + stem identity as random effect.
- **AIC(suffixes) = 3167.4 > AIC(stems+suffixes) = 3091.0**

**Suggests that triggering and undergoing morphemes have lexical propensities.**

---

**3. Lexical propensities to undergo liaison in French**

**French liaison:** word-final silent consonants are pronounced overtly if following word starts with vowel (Selkirk 1974, Tranel 1981, a.o.).

- Word1: [tres] ‘very’
- Word2: [trez aktiv] ‘very active’

Liaison applies to some Word1’s more regularly than to others.
- **tres énergétique**: [trez energetik] ‘very energetic’
- plus énergétique: [plzy energetik] ‘more energetic’

Phonologie du Français Contemporain (PFC; Durand & Lyche 2008); annotated spoken corpus of liaison forms; I extracted entire corpus, ~54,000 Word1-Word2 pairs.

Common Word1’s show patterning across a spectrum of propensity to liaise:

```
Number of Word1’s with liaison rate

<table>
<thead>
<tr>
<th>Rate</th>
<th># Stems</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>(85)</td>
</tr>
<tr>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>
```

**A good deal of factors have been found to condition liaison:**
- Part of speech of Word1, Word2 (Durand & Lyche 2008)
- Identity of liaison consonant (Mallet 2008, Ranson 2008)
- Syllable count of Word1 (Mallet 2008, Laks 2009)
- Frequency of Word1 (de Jong 1994, Fougner et al. 2001)

**Current study:** Word1 identity matters, even after considering the above factors.
- Compared logit model containing above factors...
- against mixed-effects model, W1 identity as random effect.
- **AIC(categorical)=4119.3 > AIC(Word1 propensity)=3566.7**

Suggests different Word1s have different lexical propensities.

**4. Knowledge of lexical propensities in French liaison**

Do speakers track morpheme-specific rates? Consider corpus rates for following adverbs:
- **tres (‘very’) plus (‘more’) bien (‘very’) moins (‘less’) pas (‘not’)***
- de Jong (1994) 99% 96% 82% — 7%
- Mallet (2008) 97% 64% 43% — 1%
- My study, PFC 95% 79% 85% 33% 0%

**Average:** 97% 80% 70% 33% 3%

**Nonsense probe task:** Expose speakers to très X, plus X, etc., X ∈ 16 nonce V-init. Adj.s.
- très émouvé plus émouvé ...
- très ingratuable plus ingratable ...

Speakers read a sentence containing adverb + nonce, hear two recordings of the word pair—one with liaison, one without — and are asked which recording they prefer.
- Also asked to give a liaison score for each word pair (1-5).

**Forced choice results:** Speakers showing variability (27/35) track these propensities!

```
Word1    Word1 + V-init Word 2    Word1 + V-init Word 2
          Word1 + V-init Word 2

<table>
<thead>
<tr>
<th>Word1</th>
<th>Word2</th>
<th>Word2</th>
<th>Word2</th>
</tr>
</thead>
<tbody>
<tr>
<td>très</td>
<td>beau</td>
<td>très actif</td>
<td>très aktiv</td>
</tr>
<tr>
<td>[trez bo] ‘very pretty’</td>
<td>[trez aktiv] ‘very active’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**High rates probably due to reading the frame, and normative pressures (Armstrong 2001).**
- Scale scores give similar story — très, plus > bien, moins > pas.
- Propensities associated with individual Word1’s—it cannot be that learners store propensities specific only to word pairs (cf. Bybee 2001, 2002).

**5. Conclusion**

- Slovenian corpus: triggering/undergoing morphemes show different palatalization rates spanning entire spectrum;
- French corpus: Word1’s show different liaison rates spanning entire spectrum;
- French nonce-probe task: suggests speakers acquire different Word1-specific liaison rates.

**BOTTOM LINE:** Morphemes are coded on a **propensity spectrum** to trigger/undergo variable process ([0.7 Rule]), not a binary scale ([+/- Rule]).

---

**Selected references**