Learning phonetically and phonologically natural classes through constraint indexation

Aleksei Nazarov, Utrecht University

Phonological processes tend to be defined over natural classes (Chomsky & Halle 1968, Mielke 2004), but there are some arbitrary and language-specific aspects to class behaviour (e.g., Mielke 2004). This has led to a literature on modelling phonological class induction (Dresher 2014, Sandstedt 2018, Mayer 2020). Contrast detection may be used to induce features (Dresher 2014, Sandstedt 2018). However, these approaches lie outside of OT, and, in addition, these approaches use (at least somewhat) domain-specific methods for detecting contrast rather than domain-general ones. Is it possible to implement such an approach in standard OT with domain-general methods?

One way to detect contrast is through the domain-general tool of inconsistency detection (Tesar 1995) in OT: different outputs require mutually incompatible constraint rankings (see Mackenzie 2016 for a different approach to contrast in OT). This was first used for inducing lexically indexed constraints (Becker 2009, Pater 2010). Round (2017) proposes that constraints can be indexed to individual segments in specific morphemes (see also Temkin-Martinez 2010), which localizes contrast to individual segments. The ability of constraint indexation induction to find phonetically defined phonological classes was tested in a setup where the grammar starts out with surface forms and Markedness constraints on sequences of phonetic properties; underlying forms do not yet have any feature content: Faithfulness does not yet play a role. This resembles a phonotactic learning approach (Hayes 2004, Prince & Tesar 2004, Jarosz 2006), where URs are not yet fixed.

The phonetically based Markedness constraints, in combination with segmentally local indexation, can represent phonetically and phonologically defined classes. Context-free Markedness constraints define phonetically natural classes (see also Van ‘t Veer 2015). For instance, a ranking \{\{eo\}, >> \{iu\} >> \{eo\}\} ensures that all [+i] vowels are high (i, u), while [-i] vowels are mid (e, o). Context-sensitive Markedness constraints help define what I call phonologically natural classes: classes of sounds that are defined not only by sound, but also by participating in a phonological pattern (cf. Mielke 2004). E.g., a ranking \{\{iu\}\}...\{eo\} >> \{si\}, \{e\} >> \{f\}, \{iu\}\} means that [+j] vowels are high (i, u) and trigger palatalization (*[si]), while [-j] sounds are either mid (e, o) OR are high vowels (i, u) triggered by harmony.

This setup was tested on set of toy languages adapted from Prickett & Jarosz (2021). These feature a dominant/recessive vowel harmony process (e → i / \\{\{iu\}\}, \{iu\}\}) and a palatalization process (s → f / \\{\{iu\}\}), which can either feed each other (transparent palatalization) or counterfeed each other (opaque palatalization). I also included a version where palatalization was lexically specific (applying to /es-, us-/ but not /is-/). These languages, exemplified in (1), were presented on stems /ef, -if, uf, es-, is-, us-/ with suffixes /e-, i-, u/.

\[ *\{eo\}...\{iu\}, \{iu\}\} \]

Surface candidates included all VCV combinations of [f,s\{e,i,o,u\}], and the constraint set included \{\{iu\}, \{eo\}, \{iu\}...\{eo\}, \{eo\}...\{iu\}, \{ie\}...\{uo\}, \{uo\}...\{ie\}, \{f\}, \{f\}, \{s\}, \{si\}\}. The learner had access to morphological analysis (e.g., [isi] and [usi] share the morpheme /-i/). An implemented version of Round’s (2017) segmentally local constraint indexation learner was used for these tests. This learner starts out with unindexed constraints, and adds a segmentally local indexed constraint with an arbitrary index i every time ranking these constraints leads to inconsistency. In this process, segments in the lexicon for which this indexed constraint prefers the winning option are marked as [+i] or [-i]. This learner was run 10 times for each of the languages. The results of the simulations were evaluated based on the segments that were contrasted for each indexed constraint (classes were conflated based on subset relations). The transparent language is predicted to lead to phonetically natural classes: [+/- back] (u vs. i) and [+/- coronal] (s/f vs. f) ([+/- high] (i/u vs. e) is

---

1 In the simulations, the phonetic content of URs was not taken into account: these URs are shown to explain the data.
partially neutralized), while the opaque and lexical languages will lead to a mixture of phonetically and phonologically natural classes, the latter being [+/- high front palatalizing] (i_inpal vs. i_lnonpal) for the opaque case and [+/- coronal palatalizing] (i/s/ in i/es-,us/- vs. /s/ in /is-/ ) for the lexical case.

Table 2 summarizes the number of times (out of 10) that such classes were found for each language. Phonetically natural classes were found for all cases, and appropriate phonologically natural classes were found for the opaque and lexical cases. Interestingly, the high/mid contrast, which is partially neutralized, is not reflected in the classes. This means that the inconsistency-based algorithm finds phonetically natural classes based on non-neutralized contrast, and finds phonologically natural classes when phonetically natural ones are not sufficient.

<table>
<thead>
<tr>
<th>(2) Language</th>
<th>i/u vs. e</th>
<th>u vs. i</th>
<th>s/f vs. f</th>
<th>i_inpal vs. i_lnonpal</th>
<th>i/es-,us/- vs. /i/s-/</th>
<th>Other phonetic</th>
<th>Other phonological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transparent</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Opaque</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Lexical</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

References