Learning Alternations without Bias

**keywords:** Korean morphophonology, output-output faithfulness constraint demotion, MaxEnt Grammar Tool, unbiased learner

Korean nouns show various alternations, including intervocalic voicing (1a), allophonic [l] ~[ɾ] alternations (1b), and laryngeal neutralization of obstruents in coda position (1c). In addition, stem-final coronal stops of nouns have a wide set of coronal obstruents as variants (1d)(Jun 2010, Oh 2006 and many others).

(1) a. Voicing: cin ‘house’ vs. cib-i ‘house- NOM’
   b. [l] ~ [ɾ] alternations: tal ‘moon’ vs. tar-i ‘moon-NOM’
   c. Laryngeal: ip ‘leaf’ vs. ih-i ‘leaf-NOM’
   d. Variation: pat ‘field’ vs. pas-i ~ pac^h^-i ~ pac-i ‘field- NOM’

As documented in Do (to appear), children acquiring Korean go through a stage in which they produce some alternations (1a-1c) but not others (1d). Surprisingly, child inflections of (1d) types are overwhelmingly asymmetrical: they are nearly all based on an isolated form such as pat ‘field’. Children delete an inflectional suffix or insert an extra noun or preposition to an adult form, as in (2).

(2) Adult forms Child forms
   a. pas-i ~ pac^h^-i ~ pac-i ‘field- NOM’ pat ‘field-Ø’, pat sek’ar-i ‘field color- NOM’
   b. pat^h^-il ~ pac^h^-il~pas-il ‘field- ACC’ pat ‘field-Ø’, pat k^h^igi-rii ‘field size-ACC’
   c. pat^h^-e ~ pas-e ~ pac^h^-e ‘field- LOC’ pad an-e ‘field in-LOC’, pad wi-e ‘field on-LOC’

The current learning simulation using the MaxEnt Grammar Tool (Hayes 2009) explores (a) why some alternations are acquired later and (b) why late-acquired alternations are produced based on an isolated form in the intermediate learning stage. The claim of this paper is that child outputs are motivated by a grammatical preference among children to inflect forms faithful to a base form, which is assumed to be an isolated form (Kang 2003, Ko 2006). I model this with highly promoted output-output faithfulness (OO-F) constraints in child grammar, following Hayes (2004) and McCarthy (1999). The simulation shows that the demotion of OO-F constraints is successful purely by the statistics of Korean alternations: OO-F constraints relevant to late-acquired alternations are demoted later. Since late-acquired alternations are also perceptually more salient (e.g. (1d) Δ(t-s) < (1a) Δ(p-b)) (Steriade 2001), the child grammar could reflect a prior bias for non-salient alternations. However, this study shows that a statistical model works for Korean without assuming intrinsic bias.

**The learner:** The learner is assumed to have a set of constraints, and weights them according to the frequency of violations in the data. Following Hayes (2004), learning is simulated in two stages: phonotactic and morphological learning. The phonotactic learning stage models the fact that prior to learning morphology, learners may master some phonotactic distributions.

**Phonotactic learning:** From Sejong Corpus of spoken Korean, 680 most frequent Korean nouns were categorized into 12 groups, according to the stem-final segments. Data was constructed by pairing each stem type with 10 most frequent affixes, yielding 120 noun forms. Training data
was given in the form of OT tableaux. The UR is assumed to be same as the output. A winner candidate gets its frequency based on the corpus. Violations of Markedness and Input-output faithfulness constraints were assessed by Ruby script as in (3).

(3) Training data for learning intervocalic voicing

<table>
<thead>
<tr>
<th>UR</th>
<th>Output</th>
<th>Frequency</th>
<th>*[+son,+voi]</th>
<th>*[+voi][-voi][+voi]</th>
<th>IDENT-IO (voi)</th>
<th>IDENT-IO (voi)/_+(syl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>...ibi</td>
<td>...ibi</td>
<td>12924</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...ipi</td>
<td>...ipi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After phonotactic learning, the learner has a set of weights that serve as the starting point for morphological learning.

**Morphological learning**: Assuming that a learner initially has high-ranking OO-F constraints, a set of OO-F constraints was given initial weights higher than trained weights of Markedness constraints. OO-F constraints demand forms to be faithful to the base, which was assumed to be an isolated form. The training for the morphological learning employed representative frequencies of different inflected forms according to the corpus. For forms showing variants (1d), multiple outputs were fed to the learner with their relative frequency, adapted from a survey by Choi (2004). The expectation was that the more often a given alternation occurs in the data, the more the relevant OO-F constraints will be demoted.

**Results**: In the process of learning, an intermediate stage was found when some OO-F constraints are demoted (3a-3c), while some remain over Markedness constraints (3d-3e). That is, the learner mastered some alternations, but not others. Numbers in parenthesis are trained weights.

(3) a. *[+voi][-voi][+voi](13.54) > OO-F[(±voice)](0)
b. *[+voi][+lateral][+voi](9.79) > OO-F[(±lateral)](5.93)
c. *[+aspiration](3.45) > OO-F[(±aspiration)](4.18)
d. *[+fricative](1.13E-11) < OO-F[(±fricative)](1.73)
e. *[+affricate](1.02E-13) < OO-F[(±affricate)](3.85)

The grammar trained purely by the frequency of different alternations perfectly predicts the attested intermediate learning stage, in which children cannot produce variants of stem-final coronal stops of nouns. Therefore, the results demonstrate that even without intrinsic bias, the statistics of Korean can give rise to the attested intermediate stage.

Word count: 742

**Selected References**