Segmenting with M-Phon constraints

Tibetan numerals

Diana Apoussidou
Universiteit van Amsterdam/
University of Massachusetts Amherst

Sound Circle, February 19, 2009
1. The segmentation problem: Tibetan numerals

(Halle and Clements 1983)

(1) /dʒig/ “one” /dʒugdʒig/ “eleven”
/ʃi/ “four” /dʒubʃi/ “fourteen”
/ʃibdʒu/ “fourty”
/ŋa/ “five” /dʒuŋa/ “fifteen”
/ŋabdʒu/ “fifty”
/gu/ “nine” /dʒurgu/ “nineteen”
/gubdʒu/ “ninety”
/dʒu/ “ten”
(2)  /ju/  +  /shi/  =  /jubshi/
      “ten”  “four”  “fourteen”

(3)  /shi/  +  /ju/  =  /shibju/
      “four”  “ten”  “fourty”
Segmentation possibilities:

(4) \( |jub+shi|, \ |shib+ju| \)

(5) \( |ju+bshi|, \ |shi+bju| \)
Hypothesis 1:

Morpheme boundary is between word-internal consonant cluster

(6) /jubshi/: |jub+shib|

/shibju/: |shib+jub|

Grammatical consequence: word-final consonants delete on the surface.

(7) |jub| → /ju_#/
Problems:

Inconsistent with other forms of the language:

• /jig/ “one” ends in a consonant
• /jur-/ and /-bju/ are also observed (in /jurgu/ “nineteen” and /gubju/ “ninety”)
Hypothesis 2:
Morpheme boundary is before word-internal consonant cluster

(8) /jubshi/: |bju+bshi|
    /shibju/: |bshi+bju|

Grammatical consequence: word-initial consonant clusters are forbidden and resolved.

(9) |bju| → /#_ju/
A simple, traditional OT analysis:

Markedness constraint:

(10) *#CC
   No consonant clusters word-initially

Faithfulness constraint:

(11) $\text{MAX-C}$
   Consonants in the UR are not deleted on the surface
A simple, traditional OT tableau for Tibetan “eleven”:

<table>
<thead>
<tr>
<th>/bju+gjig/</th>
<th>*#CC</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>/bjugjig/</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>/jugjig/</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>/ujujig/</td>
<td>**!</td>
<td></td>
</tr>
</tbody>
</table>
Segmentation:

(12) /dʒu+gdʒig/ |bdʒu+gdʒig| “eleven”
/dʒu+bʒi/ |bdʒu+bʒi| “fourteen”
/jib+dʒu/ |bʒib+dʒu| “fourty”
/dʒu+ŋa/ |bdʒu+ŋa| “fifteen”
/ŋa+bdʒu/ |ŋa+bdʒu| “fifty”
/dʒu+rgu/ |bdʒu+rgu| “nineteen”
/gu+bdʒu/ |rgu+bdʒu| “ninety”
Learnability questions:

• How does a learner figure out that the segmentation is |ju+bshi| and not |jub+shi|?
• How does a learner find out that /ju/ is underlyingly |bju|?
Assumption:
Learner hypothesizes only observed forms as underlying, and uses $M$-$Phon$ constraints to establish the lexicon of the language

“one” /jig/, /-gjig/
“ten” /ju/, /jub-/, /jur-/, /jug-/, /-bju/
“four” /shi/, /shib-/, /-bshi/
“five” /na/, /nab-/
“nine” /gu/, /gub-/, /-rgu/
15 resulting M-Phon (a.k.a. lexical) constraints (see Boersma 1999/2001, Apoussidou 2007, 2008):

\{|jig|, |gjig|\} “one”
\{|shi|, |bshi|, |shib|\} “four”
\{|na|, |nab|\} “five”
\{|gu|, |rgu|, |gub|\} “nine”
\{|ju|, |bju|, |jub|, |jug|, |jur|\} “ten”
Modeling segmentation of Tibetan numerals with:

- Noisy OT/GLA learners (Boersma 1997) vs.
- Noisy HG/GLA learners (Boersma & Pater 2008)

Ten virtual learners each.

Expectation:

- Resulting correct grammars (so that the observed surface forms are reproduced as heard)
HG in a tiny, tiny nutshell: Given the constraint weights...

(1) $M = 0.935 \quad F = 0.138$

...noise is added... (sampled from a Gaussian distribution, $\mu=0$, SD=0.2)

(2) $M = 1.050 \quad F = 0.160$

...the exponential is taken... (to keep the weights positive)

(3) $M = 2.858 \quad F = 1.173$

...and these weights are used to evaluate candidate mappings (constraint violations are negative integers, a candidate’s score is the weighted sum of violations, and highest score wins)

<table>
<thead>
<tr>
<th>/abd/</th>
<th>M</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>[abd]</td>
<td>-1</td>
<td>-2.858</td>
</tr>
<tr>
<td>✈ [abc]</td>
<td>-1</td>
<td>-1.173</td>
</tr>
</tbody>
</table>
Noisy OT/GLA parameters (no M >> F bias):

(14) OTMulti in Praat

*Initial ranking values*
  All equal (at 0)

*Learning rate*
  1.0 (with a plasticity decrement of 0.1)

*Number of learning data*
  400,000

*Noise*
  2.0
A resulting grammar with ranking values:

(15) “ten” |bju| 10.181
    “nine” |rgu| 8.868
    “four” |bshi| 8.309
    “five” |na| 5.498
    “one” |gjig| 3.686
    *#CC  1.975
    “ten” |ju| -0.095
    :      :
    :      :
    “nine” |gu| -6.179
\text{Max-C} -18.749
The tableau shows pairs of UR and SF pairs as candidates:

| “fourty” | |bju| |bshi| *#CC | |ju| |shi| MAX-C |
|---|---|---|---|---|---|---|---|---|
| 0.00 | |shib+ju| *! | * | | | | |
|       | /shibju/ | | | | | | | |
| >0.00 | |shi+bju| *! | * | | | | |
|       | /shibju/ | | | | | | | |
| ≤1.00 | |bshi+bju| | * | * | * | | |
|       | /shibju/ | | | | | | | |
| 0.00  | |bshi+bju| *! | * | * | | | |
Resulting output distribution (of 10 learners):

|jig| /jig/   >0.07  |bju+bshi| /jubshi/   <1.00  
|gjig| /jig/  <0.93  |bju+na| /juna/  <1.00  
|shi| /shi/  <0.01  |bshi+bju| /shibju/  <0.01  
|bshi| /shi/  >0.99  |bju+rgu| /jurgu/   <1.00  
|na| /na/  <1.00  |shi+bju| /shibju/   <0.01  
|bshi+bju| /shibju/  >0.99  
|gu| /gu/  <0.01  |na+bju| /nabju/   <1.00  
|rgu| /gu/  >0.99  |gu+bju| /gubju/   <0.01  
|bju| <1.00  |gu+bju| /gubju/   <0.01  
|bju+gjig| /jugjig/ <1.00  |rgu+bju| /gubju/   >0.99  

• Segmentation of morphemes and grammar are successfully learned: on the surface, no word-initial clusters occur.

• The lexicon is fairly restrictive.

• The learners pass the OT “richness of the base” test: the final state grammar filters out structures that are disallowed in the target language, i.e. *#CC-structures.
(16) Never-before heard form:

<table>
<thead>
<tr>
<th>gju+bji</th>
<th>*#CC</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>/gjubji/</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>/jubji/</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>
Noisy HG/GLA parameters (M > F bias imposed as in Jesney and Tessier 2008):

(17) Initial weights
   1 for Markedness constraint, 0 for others

Learning rate
   0.01 for Markedness constraint, 0.001 for others

Number of learning data
   100,000
A resulting grammar, shown with exponential (weight) values:

(18)  

<table>
<thead>
<tr>
<th>Word</th>
<th>Symbol</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>“ten”</td>
<td></td>
<td>1.153</td>
</tr>
<tr>
<td>*#CC</td>
<td></td>
<td>0.953</td>
</tr>
<tr>
<td>“four”</td>
<td></td>
<td>0.673</td>
</tr>
<tr>
<td>“nine”</td>
<td></td>
<td>0.623</td>
</tr>
<tr>
<td>“five”</td>
<td></td>
<td>0.403</td>
</tr>
<tr>
<td>“one”</td>
<td></td>
<td>0.283</td>
</tr>
<tr>
<td>“ten”</td>
<td></td>
<td>0.063</td>
</tr>
<tr>
<td>“nine”</td>
<td></td>
<td>0.003</td>
</tr>
<tr>
<td>“four”</td>
<td></td>
<td>-0.037</td>
</tr>
<tr>
<td>MAX-C</td>
<td></td>
<td>-0.103</td>
</tr>
<tr>
<td>{Rest of M-Phon}</td>
<td></td>
<td>&lt;-0.1</td>
</tr>
</tbody>
</table>
(19) A HG-learner displaying allomorphy:

| “fourty” | |bju| | #CC | |bshi| |ju| |shi| |MAX-C |
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|           | 3.17| 2.59| 1.96| 1.06| 0.96| 0.9  |

| 0.00      | |shib+ju| /shibju/ | −1 | −1 | −1 | −5.13 |
| <0.47     | |shi+bju| /shibju/ | −1 | −1 | −5.13 |
| >0.53     | |bshi+bju| /shibju/ | −1 | −1 | −1  |
| 0.00      | |bshi+bju| /bshibju/ | −1 | −1 | −1  |

Allomorphy is a valid analysis (cf. e.g. Kager 2009)
Resulting output distribution:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>jig</td>
<td>/jig/</td>
<td>&gt;0.72</td>
<td>j+sh</td>
<td>/jubshi/</td>
</tr>
<tr>
<td>gjig</td>
<td>/jig/</td>
<td>&lt;0.28</td>
<td>b+sh</td>
<td>/jubshi/</td>
</tr>
<tr>
<td>shi</td>
<td>/shi/</td>
<td>&lt;0.47</td>
<td>jn+</td>
<td>/juna/</td>
</tr>
<tr>
<td>bshi</td>
<td>/shi/</td>
<td>&gt;0.53</td>
<td>j+n</td>
<td>/juna/</td>
</tr>
<tr>
<td>na</td>
<td>/na/</td>
<td>&lt;1.00</td>
<td>j+r</td>
<td>/jurgu/</td>
</tr>
<tr>
<td>gu</td>
<td>/gu/</td>
<td>&lt;0.50</td>
<td>sh+b</td>
<td>/shibju/</td>
</tr>
<tr>
<td>rgu</td>
<td>/gu/</td>
<td>&gt;0.50</td>
<td>b+sh</td>
<td>/shibju/</td>
</tr>
<tr>
<td>ju</td>
<td>/ju/</td>
<td>&lt;0.03</td>
<td>n+b</td>
<td>/nabju/</td>
</tr>
<tr>
<td>bju</td>
<td>/ju/</td>
<td>&lt;0.97</td>
<td>g+b</td>
<td>/gubju/</td>
</tr>
<tr>
<td>j+gjig</td>
<td>/jugjig/</td>
<td>&gt;0.03</td>
<td>r+bu</td>
<td>/gubju/</td>
</tr>
<tr>
<td>bju+gjig</td>
<td>/jugjig/</td>
<td>&lt;0.98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Still, crucial weightings of constraints are maintained: all HG-learners have M > F, and rule out ungrammatical inputs (i.e. generalize correctly).

Tiny percentage of “errors” (<0.05):

“four” |shib| /.shib./

“five” |nab| /.nab./

“nine” |gub| /.gub./

“ten” |jub| /.jub./

“ten” |jug| /.jug./

“ten” |jur| /.jur./

“fourteen” |ju+shi| /.ju.shi./

“nineteen” |ju+gu| /.ju.gu./
Discussion:

- A bias of Markedness over Faithfulness is not always necessary.
- Restrictive lexicons are possible (discussion point: allomorphy).
- HG-learners might need more time (parameters).
- Size of hypothesized underlying forms is decreased by use of positively formulated constraints (cf. Tesar’s 2009 criticism)
Issues:

• The candidate set
• The size of the M-Phon constraint set
• The learnability of M-Phon constraints
References:


