

# **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+bj|

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) MAX-C

Consonants in the UR are not deleted on the surface

A simple, traditional OT tableau for Tibetan “eleven”:

bju+gjig	*#CC	MAX
/bjugjig/	*!	
☞ /jugjig/		*
/jujig/		**!

## Segmentation:

(12) /dʒu+gdʒig/	bdʒu+gdʒig	“eleven”
/dʒu+bʃi/	bdʒu+bʃi	“fourteen”
/ʃib+dʒu/	bʃib+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 |bjū|?

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-/, /-bjju/

“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|, |bjū|, |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)  $M = 0.935$      $F = 0.138$

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


(2)  $M = 1.050$      $F = 0.160$

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

(3)  $M = 2.858$      $F = 1.173$

...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)

(13)

	M	F	
/abd/	2.858	1.173	
[abd]	-1		-2.858
 [abc]		-1	-1.173

Noisy OT/GLA parameters (no  $M \gg F$  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
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  /bshibju/			*!	*	*	

Resulting output distribution (of 10 learners):

jig  /jig/	>0.07	bjubshu  /jubshu/	<1.00
gjig  /jig/	<0.93	bjubnu  /junu/	<1.00
shu  /shu/	<0.01	bjubrgu  /jurgu/	<1.00
bshu  /shu/	>0.99	shu+bjub  /shubjub/	<0.01
nu  /nu/	<1.00	bshu+bjub  /shubjub/	>0.99
gu  /gu/	<0.01	nu+bjub  /nubjub/	<1.00
rgu  /gu/	>0.99	gu+bjub  /gubjub/	<0.01
bjub	<1.00	rgu+bjub  /gubjub/	>0.99
bjub+gjig  /jugjig/	<1.00		

- 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:

gju+bji	*#CC	MAX
/gjubji/	*!	
 /jubji/		*

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)	“ten”  bju	1.153
	*#CC	0.953
	“four”  bshi	0.673
	“nine”  rgu	0.623
	“five”  na	0.403
	“one”  gjig	0.283
	“ten”  ju	0.063
	“nine”  gu	0.003
	“four”  shi	-0.037
	MAX-C	-0.103
	{Rest of M-Phon}	<-0.1

(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			-3.02
>0.53	bshi+bju  /shibju/				-1	-1	-1	-2.11
0.00	bshi+bju  /bshibju/		-1		-1	-1		-4.61

Allomorphy is a valid analysis (cf. e.g. Kager 2009)

## Resulting output distribution:

jig  /jig/	>0.72	ju+bshi  /jubshi/	<0.03
gjig  /jig/	<0.28	bjubshi  /jubshi/	>0.97
shi  /shi/	<0.47	ju+na+ /juna/	<0.03
bshi  /shi/	>0.53	bjuna+ /juna/	>0.97
na  /na/	<1.00	ju+rgu  /jurgu/	<0.03
gu  /gu/	<0.50	bjurgu  /jurgu/	>0.97
rgu  /gu/	>0.50	shi+bju  /shibju/	<0.47
ju  /ju/	<0.03	bshibju  /shibju/	>0.53
bjju  /ju/	<0.97	na+bju  /nabju/	<1.00
ju+gjig  /jugjig/	>0.03	gu+bju  /gubju/	<0.50
bjugjig  /jugjig/	<0.98	rgubju  /gubju/	>0.50

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:

Apoussidou, Diana (2007). *The learnability of metrical phonology*.

Doctoral Dissertation, University of Amsterdam. LOT Dissertation Series No. 148. Available at <http://roa.rutgers.edu/>.

Apoussidou, Diana (2008). Final devoicing, allomorphy, and Freedom of the Lexicon: An online learning approach to underlying representations. Talk presented at CLAY, Yale University, October 30th.

Boersma, Paul (2001). Phonology-semantics interaction in OT, and its acquisition. In Robert Kirchner, Wolf Wikeley & Joe Pater (eds.): *Papers in Experimental and Theoretical Linguistics*, Vol. 6: 24-35. Edmonton: University of Alberta.

Boersma, P. (2006/2008). A programme for bidirectional phonology and phonetics and their acquisition and evolution. Earlier version: Handout LOT Summerschool, June 2006, and Jadertina Summerschool. Rutgers Optimality Archive 868.

- Boersma, Paul, and Joe Pater (2007). Constructing constraints from language data: The case of Canadian English diphthongs. Paper presented to the North Eastern Linguistic Society (handout available at <http://people.umass.edu/pater/boersma-pater-nels.pdf>).
- Boersma, Paul, and Joe Pater (2008). Convergence properties of a gradual learning algorithm for Harmonic Grammar. Ms, University of Amsterdam and UMass Amherst. Available at <http://roa.rutgers.edu/>.
- Halle, M. and G. N. Clements (1983). *Problem Book in Phonology: A Workbook for Introductory Courses in Linguistics and in Modern Phonology*. Cambridge, MA: MIT Press.
- Jesney, Karen and Anne-Michelle Tessier (2008). Gradual learning and faithfulness: consequences of ranked vs. weighted constraints. To appear in Muhammad Abdurrahman, Anisa Schardl & Martin Walkow (eds.), *Proceedings of the 38th Meeting of the North East Linguistics Society (NELS 38)*. Available at <http://roa.rutgers.edu/>.
- Kager, René (2009). Lexical irregularity and the typology of contrast. In Kristin Hanson and Sharon Inkelas (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge, MA: MIT Press. <http://www.let.uu.nl/~Rene.Kager/personal/Papers/contrast.pdf>.

Tesar, Bruce (2009). Learning Phonological Grammars for Output-Driven Maps. Available at <http://roa.rutgers.edu/>.

Wolf, Matthew (2008). Optimal Interleaving: Serial Phonology-Morphology Interaction in a Constraint-Based Model. Ph.D. dissertation, UMass Amherst. Available at <http://roa.rutgers.edu/>.