Maturation as changing the base of exponential HG?

Consonant clusters (and pronoun resolution)

– joint work with Klaas Seinhorst (University of Amsterdam) –

Tamás Biró

Yale University

RUMMIT @ MIT, April 26, 2014

Overview

Overview

Learning vs. Maturation?

Learning:

- Knowledge acquired from surrounding linguistic data
- Source of cross-linguistic variation
- Features in the child's language shared by other adult languages

Maturation:

- Skills emerging due to general development
- Universal developmental paths
- Features in child's language not appearing in any adult language

Learning vs. Maturation?

Learning:

- Knowledge acquired from surrounding linguistic data
- Source of cross-linguistic variation
- Features in the child's language shared by other adult languages

Maturation:

- Skills emerging due to general development
- Universal developmental paths
- Features in child's language not appearing in any adult language

Learning vs. Maturation?

Learning from surrounding linguistic data:

- Features in the child's language shared by other adult languages
  - Child learning English produces "Italian-like" pro-drop
    $\rightarrow$ "Pro-drop" parameter not yet switched.
  - Child learning English deleting codas
    $\rightarrow$ *CODA markedness not yet demoted below FAITHFULNESS.

Maturation due to general development:

- Features in child's language not appearing in any adult language
  - Long distance place agreement in consonant harmony?
  - Erroneous pronoun resolution?

Learning vs. Maturation?

Learning from surrounding linguistic data:

- Features in the child's language shared by other adult languages
  - Child learning English produces "Italian-like" pro-drop
    $\rightarrow$ "Pro-drop" parameter not yet switched.
  - Child learning English deleting codas
    $\rightarrow$ *CODA markedness not yet demoted below FAITHFULNESS.

Maturation due to general development:

- Features in child's language not appearing in any adult language
  - Long distance place agreement in consonant harmony?
  - Erroneous pronoun resolution?

Learning vs. Maturation!

Modelling learning and modelling maturation:
shouldn't they be different?

Learning from surrounding linguistic data:

- Setting parameters
- Re-ranking constraints

Maturation due to general development:

- Restrictions on working memory, speed of mental computation...
- Varying q in q-HG?

Learning vs. Maturation!

Modelling learning and modelling maturation:
shouldn't they be different?

Learning from surrounding linguistic data:

- Setting parameters
- Re-ranking constraints

Maturation due to general development:

- Restrictions on working memory, speed of mental computation...
- Varying q in q-HG?

Overview

Exponential Harmonic Grammar, or q-HG

- **Optimality Theory** minimizes a vector of violations:

$$
H(\text{cand}) = \begin{array}{|c|c|c|c|c|c|}
\hline
C_n & C_{n-1} & \dots & C_i & \dots & C_1 \\
r_n (= n) & r_{n-1} & & r_i & & r_1 (= 1) \\
\hline
C_1(\text{cand}) & C_2(\text{cand}) & \dots & C_i(\text{cand}) & \dots & C_n(\text{cand}) \\
\hline
\end{array}
$$

- **Harmonic Grammar** minimizes a weighted sum of violations:

$$H(\text{cand}) = \sum_{i=1}^{n} w_i \cdot C_i(\text{cand})$$

- **Exponential HG:** weights are ranks exponentiated, fixed base

$$w_i = e^{r_i}$$

- ***q*-HG:** weights are ranks exponentiated, with (variable) base q

$$w_i = q^{r_i}$$

Exponential Harmonic Grammar, or q-HG

- **Optimality Theory** minimizes a vector of violations:

$$H(\text{cand}) = \begin{array}{|c|c|c|c|c|c|} \hline C_n & C_{n-1} & \dots & C_i & \dots & C_1 \\ r_n(=n) & r_{n-1} & & r_i & & r_1(=1) \\ \hline C_1(\text{cand}) & C_2(\text{cand}) & \dots & C_i(\text{cand}) & \dots & C_n(\text{cand}) \\ \hline \end{array}$$

- **Harmonic Grammar** minimizes a weighted sum of violations:

$$H(\text{cand}) = \sum_{i=1}^{n} w_i \cdot C_i(\text{cand})$$

- **Exponential HG:** weights are ranks exponentiated, fixed base

$$w_i = e^{r_i}$$

- **q-HG:** weights are ranks exponentiated, with (variable) base q

$$w_i = q^{r_i}$$

Exponential Harmonic Grammar, or *q*-HG

- **Optimality Theory** minimizes a vector of violations:

$$H(\text{cand}) = \begin{array}{|c|c|c|c|c|c|} \hline C_n & C_{n-1} & \dots & C_i & \dots & C_1 \\ r_n (= n) & r_{n-1} & & r_i & & r_1 (= 1) \\ \hline C_1(\text{cand}) & C_2(\text{cand}) & \dots & C_i(\text{cand}) & \dots & C_n(\text{cand}) \\ \hline \end{array}$$

- **Harmonic Grammar** minimizes a weighted sum of violations:

$$H(\text{cand}) = \sum_{i=1}^{n} w_i \cdot C_i(\text{cand})$$

- **Exponential HG:** weights are ranks exponentiated, fixed base

$$w_i = e^{r_i}$$

- ***q*-HG:** weights are ranks exponentiated, with (variable) base *q*

$$w_i = q^{r_i}$$

Exponential Harmonic Grammar, or *q*-HG

- **Optimality Theory** minimizes a vector of violations:

	C_n $r_n (= n)$	C_{n-1} r_{n-1}	...	C_i r_i	...	C_1 $r_1 (= 1)$
$H(\text{cand}) =$	$C_1(\text{cand})$	$C_2(\text{cand})$	...	$C_i(\text{cand})$	...	$C_n(\text{cand})$

- **Harmonic Grammar** minimizes a weighted sum of violations:

$$H(\text{cand}) = \sum_{i=1}^{n} w_i \cdot C_i(\text{cand})$$

- **Exponential HG:** weights are ranks exponentiated, fixed base

$$w_i = e^{r_i}$$

- ***q*-HG:** weights are ranks exponentiated, with (variable) base q

$$w_i = q^{r_i}$$

Strict domination in OT is q-HG in the $q \to +\infty$ limit

- 1.5-HG has *ganging-up cumulativity*:

$w =$	C_3 2.25	C_2 1.5	C_1 1	H
☞ cand1	1			2.25
cand2		1	1	2.5

- 1.5-HG also has *counting cumulativity*:

$w_i =$	C_3 2.25	C_2 1.5	C_1 1	H
☞ cand1	1			2.25
cand3		2		3

(Cf. Jäger and Rosenbach 2006)

Strict domination in OT is q-HG in the $q \to +\infty$ limit

- But OT does not have *ganging-up cumulativity*:

	C_3	C_2	C_1
cand1	*		
☞ cand2		*	*

- OT does not have *counting cumulativity* either:

	C_3	C_2	C_1
cand1	*		
☞ cand3		**	

(Regarding Stochastic OT, cf. Jäger and Rosenbach 2006)

Strict domination in OT is q-HG in the $q \to +\infty$ limit

- 3-HG does not have *ganging-up cumulativity*:

	C_3 9	C_2 3	C_1 1	H
cand1	1			9
☞ cand2		1	1	4

- 3-HG does not have *counting cumulativity*, either:

	C_3 9	C_2 3	C_1 1	H
cand1	1			9
☞ cand3		2		6

(Cf. Jäger and Rosenbach 2006)

Strict domination in OT is q-HG in the $q \to +\infty$ limit

As we have known it since Prince and Smolensky 1993,

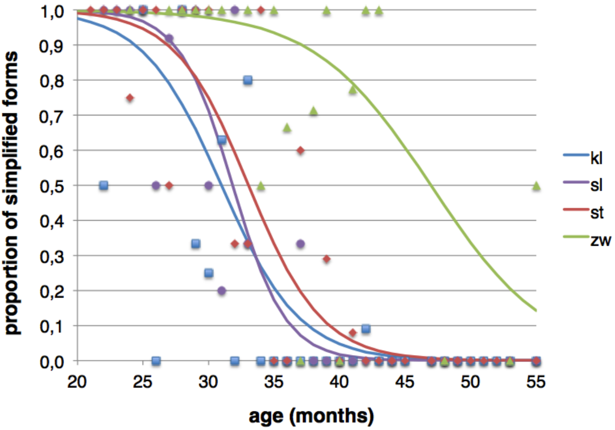
strict domination in OT can be reproduced

using q-HG with sufficiently large q.

Overview

Word initial consonant cluster simplification in Dutch

Klaas Seinhorst collecting data from CHILDES (*Laura*):



Cf. Becker and Tessier (2011)

Word initial consonant cluster simplification in Dutch

Using logistic regression or probit regression:

cluster	simplifies to	lower boundary (age in days)	upper boundary (age in days)
kl-	*k*-	894.32	1010.55
sl-	*l*-	943.82	1028.68
st-	*t*-	962.60	1076.23
zw-	*z*-	1344.24	1551.39

Table: 95% confidence intervals of the locations of the inflection points.

Differences among *kl*, *sl* and *st*: statistically not significant.
But differences between *zw* and each of the three others: $p < 10^{-11}$!

Word initial consonant cluster simplification: OT

The traditional account:

- Before learning: Markedness ≫ Faithfulness

/klεin/	NoComplexOnset	FaithF	*[l]	*[k]
[klεin]	*!		*	*
☞ [kεin]		*		*
[lεin]		*	*!	

- After learning: Faithfulness ≫ Markedness

/klεin/	FaithF	NoComplexOnset	*[l]	*[k]
☞ [klεin]		*	*	*
[kεin]	*!			*
[lεin]	*!		*	

Word initial consonant cluster simplification: OT

Questions to the traditional account:

- Child is exposed to huge amount of evidence way before correct production. Why no learning?
- If only *NoComplexOnset* and *Faithf* are involved, why significant difference for *zw* onset?
- If cluster-specific constraints: factorial typology predicted.

Word initial consonant cluster simplification: *q*-HG

An alternative approach:

- Child has acquired FAITHF $\gg$ NOCOMPLEXONSET much earlier, probably already at pre-linguistic age.
- Relative ranks *[w] $\gg$ *[s] $\gg$ *[l] $\gg$ *[z] $\gg$ *[k] $\gg$ *[t] motivated by *natural phonology* (? feedback appreciated!).
- No more ranking needed. For instance,

C_i	FAITHF	NOCOMPL ONSET	*[w]	*[s]	*[l]	*[z]	*[k]	*[t]
r_i	8	7	6	5	4	3	2	1
$(1.1)^{r_i}$	2.14	1.95	1.77	1.61	1.46	1.33	1.21	1.1
2^{r_i}	256	128	64	32	16	8	4	2

Word initial consonant cluster simplification: *q*-HG

- Before maturation: small q, e.g., $q = 1.1$ (NB: Faithfulness $\gg$ Markedness!)

/klɛin/ $w_i =$	FAITHF 2.14	NOCOMPLONS 1.95	*[l] 1.46	*[k] 1.21	H
[klɛin]		*	*	*	4.62
☞ [kɛin]	*!			*	3.35
[lɛin]	*!		*		3.60

- After maturation: large q, e.g., $q = 2$

/klɛin/ $w_i =$	FAITHF 256	NOCOMPLONS 128	*[l] 16	*[k] 4	H
☞ [klɛin]		*	*	*	148
[kɛin]	*!			*	260
[lɛin]	*!		*		272

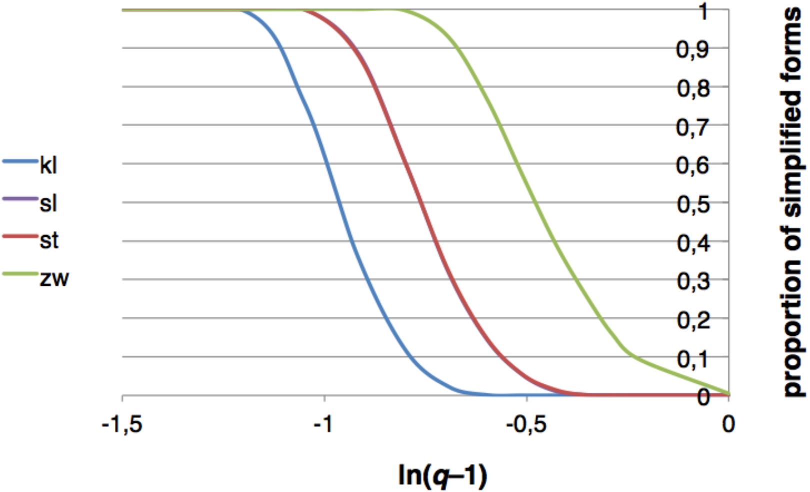
Word initial consonant cluster simplification: q-HG

- q is a function of age, e.g. age $\propto \log(q)$.
- [xy] produced by q-HG, if q is s.t. $q^c + q^x + q^y = q^f + q^y$ or larger:

/xy/ $r_i =$ $w_i =$	FAITHF f q^f	*COMPLONS c q^c	*[x] x q^x	*[y] y q^y	H for given q
[xy]	0	1	1	1	$q^c + q^x + q^y$
[y]	1	0	0	1	$q^f + q^y$
[x]	1	0	1	0	$q^f + q^x$

- Critical age function of deleted segment [x], but not remaining [y].
- If $f > c$, $x > y$, then: step function predicted.
- To get S-shaped curve, use Stochastic OT.

Word initial consonant clusters: stochastic *q*-HG



Overview

Five levels of cognitive modeling

1. **General cognitive principles:** e.g., optimize a target function.
2. **Cognitive architecture:** e.g., OT, bi-OT, Stoch OT, or *q*-HG.
3. **Cognitive infrastructure:** e.g., value of *q* in *q*-HG.
4. **Knowledge:** e.g., constraint ranking.
5. **Implementation,** which might be prone to error (performance).

Maturation vs. learning

- **Learning:** acquiring knowledge based on observations possibly already in the pre-linguistic stage.
- **Maturation:** fine-tuning the infrastructure possibly due to physical and general cognitive development.
- **Phonology** goes from HG to OT (q from $1 + \epsilon$ to large): speed $\gg$ precision.
- **Syntax-semantics** goes from OT to HG (q from large to $1 + \epsilon$): precision $\gg$ speed.

Points of discussion?

- Would you buy *architecture* vs. *infrastructure* distinction?
- Would you buy a *q*-HG model of maturation?
- *[w] $\gg$ *[s] $\gg$ *[l] $\gg$ *[z] $\gg$ *[k] $\gg$ *[t]

Thank you for your attention!

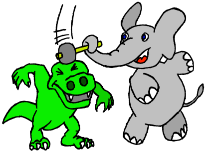
Tamás Biró:
tamas[dot]biro[at]yale[dot]edu

Many thanks to:

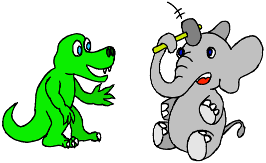
Amsterdam Center for Language and Communication ACLC

NWO Netherlands Organisation for Scientific Research

Pronoun resolution problem: data



The elephant is hitting him.



The elephant is hitting himself.

Source: P. Hendriks, `http://www.let.rug.nl/hendriks/vici.htm`. Drawings by R. Prins.

- *Here is an elephant and an alligator. The elephant hits him*—true?
- *What does the elephant do?*
- Children of age 4-6 are better at producing pronouns (and reflexives) than interpreting them. Interpretation performance: 50-80 %.

Pronoun interpretation problem: possible explanations

Government and Binding (GB):

- Principle A: anaphors must be bound within their domain.
- Principle B: pronouns must not be bound within their domain.
- Principle C: R-expressions must not be bound.

- Chien and Wexler: children do not have Principle B yet, due to apparent violations (*He$_i$ looks like him$_j$).*
- Reinhart: insufficient working memory for mental computations.
- Petra Hendriks and Jacolien van Rij: too slow mental computation.
- Hendriks and Spenader: Principle A + bidirectional OT (Principle B not necessary). Children do not have bi-OT before fully developed Theory of Mind.
- Biró: implementation of OT (performance model) prone to errors, but not so much in Harmonic Grammar (HG).

From Harmonic Grammar to Optimality Theory

Candidate set 1 (no insertion), $K_{max} = 5$, $T_{step} = 0.1$.

Precision: probability of correctly interpreting *The elephant hits him*.

q	precision
OT	0.500
30	0.499 ± 0.008
20	0.500 ± 0.012
10	0.499 ± 0.003
5	0.511 ± 0.001
3	0.550 ± 0.005
2.5	0.580 ± 0.003
2.0	0.633 ± 0.003
1.8	0.666 ± 0.003
1.7	0.687 ± 0.007
1.6	0.716 ± 0.006
1.5	0.749 ± 0.008

q	precision
1.4	0.790 ± 0.004
1.3	0.847 ± 0.001
1.2	0.911 ± 0.002
1.15	0.945 ± 0.003
1.10	0.978 ± 0.001
1.08	0.986 ± 0.001
1.06	0.994 ± 0.001
1.05	0.997 ± 0.001
1.04	0.9985 ± 0.0003
1.03	0.9991 ± 0.0005
1.02	0.99977 ± 0.00015
1.01	0.99997 ± 0.00006