Harmonic Grammar growing into Optimality Theory

Maturation as the strict domination limit (or vice-versa)

including joint work with Klaas Seinhorst (University of Amsterdam)

Tamás Biró

ELTE Eötvös Loránd University

BLINC @ ELTE, June 2, 2017

Learning and maturation

A standard idea in contemporary linguistics:

- Children learning a language:
  setting parameters or re-ranking constraints.
- Hence / because: children speak a typologically different language that resides in the typology of adult languages.

Is it really so?

Goal of this talk: to show how **maturation** (as opposed to *learning*) can be included into OT.

Learning and maturation

A standard idea in contemporary linguistics:

- Children learning a language:
  setting parameters or re-ranking constraints.
- Hence / because: children speak a typologically different language that resides in the typology of adult languages.

Is it really so?

Goal of this talk: to show how **maturation** (as opposed to *learning*) can be included into OT.

Overview

1. Optimality Theory and Harmonic Grammar
2. From HG to OT: the strict domination limit
3. Consonant cluster simplification in Dutch
4. Summary: learning and maturation

Overview

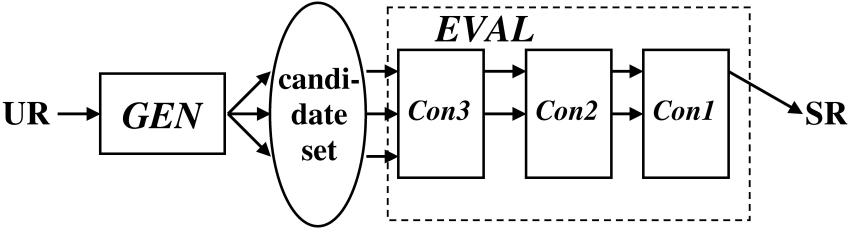
1 Optimality Theory and Harmonic Grammar

2 From HG to OT: the strict domination limit

3 Consonant cluster simplification in Dutch

4 Summary: learning and maturation

Optimality Theory in a broad sense



- Underlying representation $\mapsto$ candidate set.
- Surface representation $=$ optimal element of candidate set.
- Optimality: most harmonic. What is "harmony"?

Optimality Theory in a narrow sense

$\mathrm{Gen}(\sigma\sigma\sigma\sigma) = \{$[suuu], [usuu], [uusu], [uuus]$\}$.

/σσσσ/	EARLY	LATE	NONFINAL
☞ [s u u u]	0	3	0
[u s u u]	1!	2	0
[u u s u]	2!	1	0
[u u u s]	3!	0	1

$\mathrm{SF}(\sigma\sigma\sigma\sigma) =$[suuu].

Optimality Theory in a narrow sense

$\text{Gen}(\sigma\sigma\sigma\sigma) = \{$[suuu], [usuu], [uusu], [uuus]$\}$.

/$\sigma\sigma\sigma\sigma$/	NONFINAL	LATE	EARLY
[s u u u]	0	3!	0
[u s u u]	0	2!	1
☞ [u u s u]	0	1	2
[u u u s]	1!	0	3

$\text{SF}(\sigma\sigma\sigma\sigma) =$[uusu].

Optimality Theory in a narrow sense

$\mathrm{Gen}(\sigma\sigma\sigma\sigma) = \{$[suuu], [usuu], [uusu], [uuus]$\}$.

/$\sigma\sigma\sigma\sigma$/	NONFINAL	LATE	EARLY
☞ [u u s u]	0	1	2
[u s u u]	0	2	1
[s u u u]	0	3	0
[u u u s]	1	0	3

Harmony in terms of **lexicographic order**:
[uusu] $\succ$ [usuu] $\succ$ [suuu] $\succ$ [uuus]. Hence, $\mathrm{SF}(\sigma\sigma\sigma\sigma) =$[uusu].

Harmonic Grammar, symbolic approach

$\text{Gen}(\sigma\sigma\sigma\sigma) = \{$[suuu], [usuu], [uusu], [uuus]$\}$.

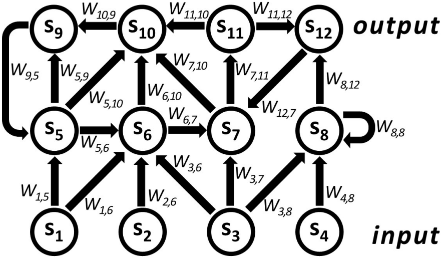
Constraint C_k is assigned weight w_k.

Harmony $H(A)$ of cand. A:

/σσσσ/	NONFINAL	LATE	EARLY	$-H(A) =$
$w_k =$	9	3	1	$\sum_k w_k \cdot C_k[A]$
☞ [u u s u]	0	1	2	$9 \cdot 0 + 3 \cdot 1 + 1 \cdot 2 = 5$
[u s u u]	0	2	1	$9 \cdot 0 + 3 \cdot 2 + 1 \cdot 1 = 7$
[s u u u]	0	3	0	$9 \cdot 0 + 3 \cdot 3 + 1 \cdot 0 = 9$
[u u u s]	1	0	3	$9 \cdot 1 + 3 \cdot 0 + 1 \cdot 3 = 12$

By comparing the integer/real numbers in H:
[uusu] $\succ$ [usuu] $\succ$ [suuu] $\succ$ [uuus]. Hence, $\text{SF}(\sigma\sigma\sigma\sigma) =$[uusu].

Harmonic Grammar, connectionist approach



- s_i: activation of node i.
- $W_{i,j}$: connection strength between nodes i and j.

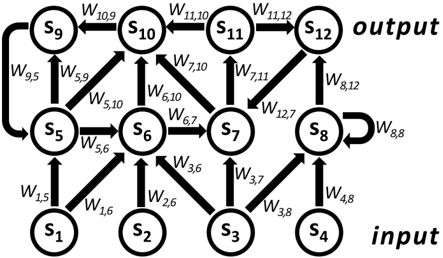
Boltzmann machine:

- The "energy" (negative harmony) of the network:

$$-H(A) = \sum_{i,j=1}^{N} s_i \cdot W_{i,j} \cdot s_j$$

- Input nodes clamped (fixed).
- Output nodes read after optimization of $H(A)$.

Harmonic Grammar, connectionist approach



- Constraint C_k is a set of $W^k_{i,j}$ partial connection strengths.

- Constraint C_k has weight w_k:

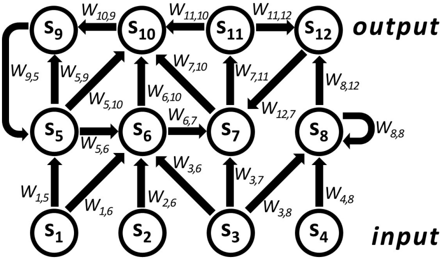
$$W_{i,j} = \sum_{k=1}^{n} w_k \cdot W^k_{i,j}$$

- The “energy” (negative harmony) of the network:

$$-H(A) = \sum_{i,j=1}^{N} s_i \cdot W_{i,j} \cdot s_j =$$

$$= \sum_{i,j=1}^{N} s_i \cdot \sum_{k=1}^{n} w_k \cdot W^k_{i,j} \cdot s_j = \sum_{k=1}^{n} w_k \cdot \sum_{i,j=1}^{N} s_i \cdot W^k_{i,j} \cdot s_j = \sum_{k=1}^{n} w_k \cdot C_k[A].$$

Harmonic Grammar, connectionist approach



- Constraint C_k has weight w_k:

$$W_{i,j} = \sum_{k=1}^{n} w_k \cdot W_{i,j}^k$$

- The "energy" (negative harmony) of the network:

- Constraint C_k is a set of $W_{i,j}^k$ partial connection strengths.

$$-H(A) = \sum_{i,j=1}^{N} s_i \cdot W_{i,j} \cdot s_j =$$

$$= \sum_{i,j=1}^{N} s_i \cdot \sum_{k=1}^{n} w_k \cdot W_{i,j}^k \cdot s_j = \sum_{k=1}^{n} w_k \cdot \sum_{i,j=1}^{N} s_i \cdot W_{i,j}^k \cdot s_j = \sum_{k=1}^{n} w_k \cdot C_k[A].$$

Summary thus far:

We have three approaches:

- Optimality Theory: symbolic,
  optimization w.r.t. lexicographic order.
- Symbolic Harmonic Grammar:
  optimization w.r.t. $\geq$ relation among real numbers.
- Connectionist Harmonic Grammar:
  optimization w.r.t. $\geq$ relation among real numbers.

Typological predictions? Learnability? ***Cognitive plausibility?***

Question: how to get OT in a "connectionist brain"?

(A major issue for Smolensky's *Integrated Connectionist/Symbolic Architecture.)*

Summary thus far:

We have three approaches:

- Optimality Theory: symbolic, optimization w.r.t. lexicographic order.
- Symbolic Harmonic Grammar: optimization w.r.t. $\geq$ relation among real numbers.
- Connectionist Harmonic Grammar: optimization w.r.t. $\geq$ relation among real numbers.

Typological predictions? Learnability? ***Cognitive plausibility?***

Question: how to get OT in a "connectionist brain"?

(A major issue for Smolensky's *Integrated Connectionist/Symbolic Architecture.)*

Summary thus far:

We have three approaches:

- Optimality Theory: symbolic,
  optimization w.r.t. lexicographic order.
- Symbolic Harmonic Grammar:
  optimization w.r.t. $\geq$ relation among real numbers.
- Connectionist Harmonic Grammar:
  optimization w.r.t. $\geq$ relation among real numbers.

Typological predictions? Learnability? ***Cognitive plausibility?***

Question: how to get OT in a "connectionist brain"?

(A major issue for Smolensky's *Integrated Connectionist/Symbolic Architecture.)*

Overview

1. Optimality Theory and Harmonic Grammar
2. From HG to OT: the strict domination limit
3. Consonant cluster simplification in Dutch
4. Summary: learning and maturation

Exponential Harmonic Grammar, or q-HG

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

	C_n $r_n (= n)$	C_{n-1} r_{n-1}	$\ldots$	C_i r_i	$\ldots$	C_1 $r_1 (= 1)$
$H(\mathrm{A}) =$	$C_1[\mathrm{A}]$	$C_2[\mathrm{A}]$	$\ldots$	$C_i[\mathrm{A}]$	$\ldots$	$C_n[\mathrm{A}]$

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

$$H(\mathrm{A}) = \sum_{i=1}^{n} w_i \cdot C_i[\mathrm{A}].$$

- "Standard" HG: weights $w_i =$ ranks r_i.
- **Exponential HG:** weights are ranks exponentiated, fixed base ($e = 2.7182\ldots$) $w_i = e^{r_i}$.
- 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	C_{n-1}	$\ldots$	C_i	$\ldots$	C_1
	$r_n(=n)$	r_{n-1}		r_i		$r_1(=1)$
$H(\mathrm{A}) =$	$C_1[\mathrm{A}]$	$C_2[\mathrm{A}]$	$\ldots$	$C_i[\mathrm{A}]$	$\ldots$	$C_n[\mathrm{A}]$

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

$$H(\mathrm{A}) = \sum_{i=1}^{n} w_i \cdot C_i[\mathrm{A}].$$

- "Standard" HG: weights $w_i =$ ranks r_i.
- **Exponential HG:** weights are ranks exponentiated, fixed base ($e = 2.7182\ldots$) $$w_i = e^{r_i}.$$
- ***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:

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

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

$$H(\mathrm{A}) = \sum_{i=1}^{n} w_i \cdot C_i[\mathrm{A}].$$

- "Standard" HG: weights w_i = ranks r_i.
- **Exponential HG:** weights are ranks exponentiated, fixed base ($e = 2.7182\dots$) $$w_i = e^{r_i}.$$
- *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:

$$H(\mathrm{A}) = \begin{array}{|c|c|c|c|c|c|} \hline C_n & C_{n-1} & \ldots & C_i & \ldots & C_1 \\ r_n(=n) & r_{n-1} & & r_i & & r_1(=1) \\ \hline C_1[\mathrm{A}] & C_2[\mathrm{A}] & \ldots & C_i[\mathrm{A}] & \ldots & C_n[\mathrm{A}] \\ \hline \end{array}$$

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

$$H(\mathrm{A}) = \sum_{i=1}^{n} w_i \cdot C_i[\mathrm{A}].$$

- “Standard” HG: weights $w_i = $ ranks r_i.
- **Exponential HG:** weights are ranks exponentiated, fixed base ($e = 2.7182\ldots$) $$w_i = e^{r_i}.$$
- 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

- In q-HG:

$$H(\mathrm{A}) = q^{r_n} \cdot C_n[\mathrm{A}] + \ldots + q^{r_i} \cdot C_i[\mathrm{A}] + \ldots + q^{r_1} \cdot C_1[\mathrm{A}]$$

- Or simply (if $r_i = i - 1$):

$$\begin{aligned}
H(\mathrm{A}) &= q^{n-1} \cdot C_n[\mathrm{A}] + \ldots + q^2 \cdot C_i[\mathrm{A}] + q^1 \cdot C_i[\mathrm{A}] + q^0 \cdot C_1[\mathrm{A}] \\
H(\mathrm{A}) &= 2^{n-1} \cdot C_n[\mathrm{A}] + \ldots + 4 \cdot C_i[\mathrm{A}] + 2 \cdot C_i[\mathrm{A}] + 1 \cdot C_1[\mathrm{A}] \\
H(\mathrm{A}) &= 3^{n-1} \cdot C_n[\mathrm{A}] + \ldots + 9 \cdot C_i[\mathrm{A}] + 3 \cdot C_i[\mathrm{A}] + 1 \cdot C_1[\mathrm{A}] \\
H(\mathrm{A}) &= 10^{n-1} \cdot C_n[\mathrm{A}] + \ldots + 100 \cdot C_i[\mathrm{A}] + 10 \cdot C_i[\mathrm{A}] + 1 \cdot C_1[\mathrm{A}]
\end{aligned}$$

- Main difference between OT and HG is **strict domination**.

- If q grows large, q-HG turns into OT, because…

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

- In q-HG:

$$H(\mathrm{A}) = q^{r_n} \cdot C_n[\mathrm{A}] + \ldots + q^{r_i} \cdot C_i[\mathrm{A}] + \ldots + q^{r_1} \cdot C_1[\mathrm{A}]$$

- Or simply (if $r_i = i - 1$):

$$\begin{aligned}
H(\mathrm{A}) &= q^{n-1} \cdot C_n[\mathrm{A}] + \ldots + q^2 \cdot C_i[\mathrm{A}] + q^1 \cdot C_i[\mathrm{A}] + q^0 \cdot C_1[\mathrm{A}] \\
H(\mathrm{A}) &= 2^{n-1} \cdot C_n[\mathrm{A}] + \ldots + 4 \cdot C_i[\mathrm{A}] + 2 \cdot C_i[\mathrm{A}] + 1 \cdot C_1[\mathrm{A}] \\
H(\mathrm{A}) &= 3^{n-1} \cdot C_n[\mathrm{A}] + \ldots + 9 \cdot C_i[\mathrm{A}] + 3 \cdot C_i[\mathrm{A}] + 1 \cdot C_1[\mathrm{A}] \\
H(\mathrm{A}) &= 10^{n-1} \cdot C_n[\mathrm{A}] + \ldots + 100 \cdot C_i[\mathrm{A}] + 10 \cdot C_i[\mathrm{A}] + 1 \cdot C_1[\mathrm{A}]
\end{aligned}$$

- Main difference between OT and HG is **strict domination**.
- If q grows large, q-HG turns into OT, because…

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
☞ A1	1			2.25
A2		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
☞ A1	1			2.25
A3		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
A1	*		
☞ A2		*	*

- OT does not have *counting cumulativity* either:

	C_3	C_2	C_1
A1	*		
☞ A3		**	

(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
A1	1			9
☞ A2		1	1	4

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

	C_3 9	C_2 3	C_1 1	H
A1	1			9
☞ A3		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

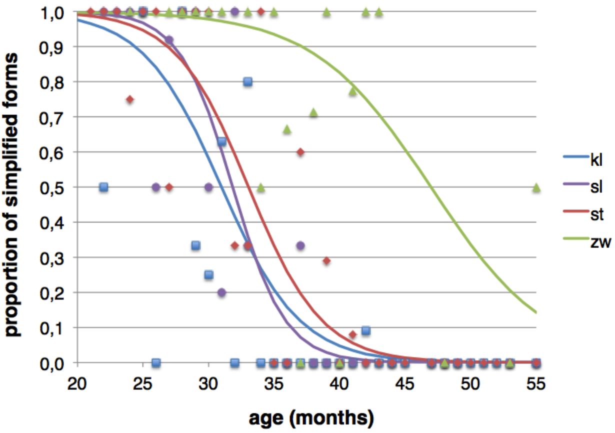
using q-HG with sufficiently large q:

$$q - 1 \geq C_k[\mathrm{A}] \qquad \text{for all } k \text{ and } \mathrm{A}.$$

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 in OT: learning = constraint re-ranking.

- Before learning: Markedness ≫ Faithfulness

/klɛin/	NoComplexOnset	FaithF	*[l]	*[k]
[klɛin]	*!		*	*
☞ [kɛin]		*		*
[lɛin]		*	*!	

- After learning: Faithfulness ≫ Markedness

/klɛin/	FaithF	NoComplexOnset	*[l]	*[k]
☞ [klɛin]		*	*	*
[kɛin]	*!			*
[lɛin]	*!		*	

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 ≫ 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

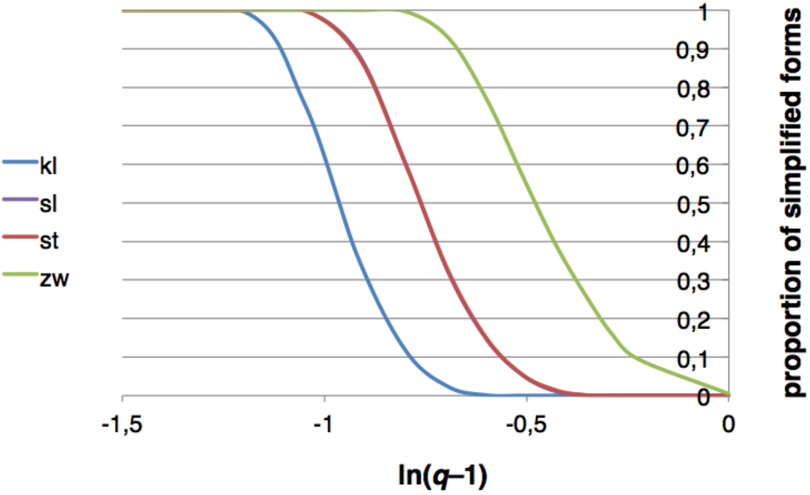
Word initial consonant cluster simplification: *q*-HG

- 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

1. Optimality Theory and Harmonic Grammar
2. From HG to OT: the strict domination limit
3. Consonant cluster simplification in Dutch
4. Summary: learning and maturation

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?

Thank you for your attention!

Tamás Biró:
tamas[dot]biro[at]btk[dot]elte[dot]hu
http://birot.web.elte.hu/

This research was supported by a *Veni Grant* (project number 275-89-004) offered by the Netherlands Organisation for Scientific Research (NWO), as well as by a *Marie Curie FP7 Integration Grant* (no. PCIG14-GA-2013-631599, “MeMoLI”), *7th EU Framework Programme*.