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

1. Learning and Maturation
2. Exponential Harmonic Grammar, or $q$-HG
3. Consonant cluster simplification in Dutch
4. Summary
Overview

1. Learning and Maturation

2. Exponential Harmonic Grammar, or $q$-HG

3. Consonant cluster simplification in Dutch

4. Summary
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
    → “Pro-drop” parameter not yet switched.
  - Child learning English deleting codas
    → *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
    → “Pro-drop” parameter not yet switched.
  - Child learning English deleting codas
    → *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?
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

1. Learning and Maturation

2. Exponential Harmonic Grammar, or $q$-HG

3. Consonant cluster simplification in Dutch

4. Summary
Exponential Harmonic Grammar, or \( q \)-HG

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

\[
H(\text{cand}) = \begin{array}{cccccc}
C_n & C_{n-1} & \ldots & C_i & \ldots & C_1 \\
r_n(=n) & r_{n-1} & \ldots & r_i & \ldots & r_1(=1) \\
C_1(\text{cand}) & C_2(\text{cand}) & \ldots & C_i(\text{cand}) & \ldots & C_n(\text{cand})
\end{array}
\]

- **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:

  \[
  H(\text{cand}) = \begin{pmatrix}
  C_n \\
  r_n (= n) \\
  C_{n-1} \\
  r_{n-1} \\
  \vdots \\
  C_i \\
  r_i \\
  \vdots \\
  C_1 \\
  r_1 (= 1)
  \end{pmatrix}
  \]

- **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:

\[
H(\text{cand}) = \begin{bmatrix}
C_n \\
r_n (= n)
\end{bmatrix}
\begin{bmatrix}
C_{n-1} \\
r_{n-1}
\end{bmatrix}
\ldots
\begin{bmatrix}
C_i \\
r_i
\end{bmatrix}
\ldots
\begin{bmatrix}
C_1 \\
r_1 (= 1)
\end{bmatrix}
\]

- **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:

  \[
  H(\text{cand}) = \begin{bmatrix}
  C_n & C_{n-1} & \ldots & C_i & \ldots & C_1 \\
  r_n (= n) & r_{n-1} & \ldots & r_i & \ldots & r_1 (= 1)
  \end{bmatrix}
  \]

- **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 \( q \)

  \[
  w_i = q^{r_i}
  \]

Tamás Biró

Maturation as changing the base of exponential HG?
Strict domination in OT is \( q\text{-HG} \) in the \( q \to +\infty \) limit

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

\[
\begin{array}{|c|c|c|c|c|}
\hline
w = & C_3 & C_2 & C_1 & H \\
\hline
\downarrow & 2.25 & 1.5 & 1 & \\
\hline
cand1 & 1 & & & 2.25 \\
\hline
cand2 & 1 & 1 & & 2.5 \\
\hline
\end{array}
\]

- 1.5-HG also has *counting cumulativity*:

\[
\begin{array}{|c|c|c|c|c|}
\hline
w_i = & C_3 & C_2 & C_1 & H \\
\hline
\downarrow & 2.25 & 1.5 & 1 & \\
\hline
cand1 & 1 & & & 2.25 \\
\hline
cand3 & 2 & & & 3 \\
\hline
\end{array}
\]

(Cf. Jäger and Rosenbach 2006)
Strict domination in OT is $q$-HG in the $q \rightarrow +\infty$ limit

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

<table>
<thead>
<tr>
<th></th>
<th>$C_3$</th>
<th>$C_2$</th>
<th>$C_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cand1</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cand2</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

- OT does not have *counting cumulativity* either:

<table>
<thead>
<tr>
<th></th>
<th>$C_3$</th>
<th>$C_2$</th>
<th>$C_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cand1</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cand3</td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

(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 \textit{ganging-up cumulativity}:

\[
\begin{array}{c|c|c|c|c}
 & C_3 & C_2 & C_1 & H \\
\hline
\text{cand1} & 1 & & & 9 \\
\text{cand2} & 1 & 1 & & 4 \\
\end{array}
\]

- 3-HG does not have \textit{counting cumulativity}, either:

\[
\begin{array}{c|c|c|c|c}
 & C_3 & C_2 & C_1 & H \\
\hline
\text{cand1} & 1 & & & 9 \\
\text{cand3} & 2 & & & 6 \\
\end{array}
\]

(Cf. Jäger and Rosenbach 2006)
Strict domination in OT is \( q \)-HG in the \( q \rightarrow +\infty \) 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

1. Learning and Maturation
2. Exponential Harmonic Grammar, or $q$-HG
3. Consonant cluster simplification in Dutch
4. Summary
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:

<table>
<thead>
<tr>
<th>cluster</th>
<th>simplifies to</th>
<th>lower boundary (age in days)</th>
<th>upper boundary (age in days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>kl-</td>
<td>k-</td>
<td>894.32</td>
<td>1010.55</td>
</tr>
<tr>
<td>sl-</td>
<td>l-</td>
<td>943.82</td>
<td>1028.68</td>
</tr>
<tr>
<td>st-</td>
<td>t-</td>
<td>962.60</td>
<td>1076.23</td>
</tr>
<tr>
<td>zw-</td>
<td>z-</td>
<td>1344.24</td>
<td>1551.39</td>
</tr>
</tbody>
</table>

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: $p < 10^{-11}$!
Learning and Maturation

**Word initial consonant cluster simplification: OT**

The traditional account:

- **Before learning: Markedness \(\gg\) Faithfulness**

<table>
<thead>
<tr>
<th>/klɛin/</th>
<th>NO\text{COMPLEX} \text{ONSET}</th>
<th>FAITHF</th>
<th>*[l]</th>
<th>*[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[klɛin]</td>
<td>![]</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>![kɛin]</td>
<td>![]</td>
<td></td>
<td>![]</td>
<td>![]</td>
</tr>
<tr>
<td>![lɛin]</td>
<td>![]</td>
<td>![]</td>
<td>![]</td>
<td>![]</td>
</tr>
</tbody>
</table>

- **After learning: Faithfulness \(\gg\) Markedness**

<table>
<thead>
<tr>
<th>/klɛin/</th>
<th>FAITHF</th>
<th>NO\text{COMPLEX} \text{ONSET}</th>
<th>*[l]</th>
<th>*[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>![kɛin]</td>
<td>![]</td>
<td>![]</td>
<td>![]</td>
<td>![]</td>
</tr>
<tr>
<td>![lɛin]</td>
<td>![]</td>
<td>![]</td>
<td>![]</td>
<td>![]</td>
</tr>
</tbody>
</table>

Tamás Biró

Maturation as changing the base of exponential HG?
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 F A I T H F \gg N O C O M P L E X O N S E T 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,

<table>
<thead>
<tr>
<th>$C_i$</th>
<th>FaithF</th>
<th>NoCompL Onset</th>
<th><em>[w]</em></th>
<th><em>[s]</em></th>
<th><em>[l]</em></th>
<th><em>[z]</em></th>
<th><em>[k]</em></th>
<th><em>[t]</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_i$</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>$(1.1)^{r_i}$</td>
<td>2.14</td>
<td>1.95</td>
<td>1.77</td>
<td>1.61</td>
<td>1.46</td>
<td>1.33</td>
<td>1.21</td>
<td>1.1</td>
</tr>
<tr>
<td>$2^{r_i}$</td>
<td>256</td>
<td>128</td>
<td>64</td>
<td>32</td>
<td>16</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
# Word initial consonant cluster simplification: $q$-HG

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

<table>
<thead>
<tr>
<th>/klɛin/</th>
<th>FAITHF</th>
<th>NoCOMPLONS</th>
<th>*[l]</th>
<th>*[k]</th>
<th>$H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_i =$</td>
<td>2.14</td>
<td>1.95</td>
<td>1.46</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>[klɛin]</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>4.62</td>
</tr>
<tr>
<td>[kɛin]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td>3.35</td>
</tr>
<tr>
<td>[lɛin]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td>3.60</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>/klɛin/</th>
<th>FAITHF</th>
<th>NoCOMPLONS</th>
<th>*[l]</th>
<th>*[k]</th>
<th>$H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_i =$</td>
<td>256</td>
<td>128</td>
<td>16</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>[klɛin]</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>148</td>
</tr>
<tr>
<td>[kɛin]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td>260</td>
</tr>
<tr>
<td>[lɛin]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td>272</td>
</tr>
</tbody>
</table>
Word initial consonant cluster simplification: \( q \)-HG

- \( q \) is a function of age, e.g. \( \text{age} \propto \log(q) \).

- \([xy]\) produced by \( q \)-HG, if \( q \) is s.t. \( q^c + q^x + q^y = q^f + q^y \) or larger:

<table>
<thead>
<tr>
<th>( /xy/ )</th>
<th>( \text{F\text{AITHF}} )</th>
<th>( *\text{C\text{OMPLONS}} )</th>
<th>( *[x] )</th>
<th>( *[y] )</th>
<th>( H ) for given ( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_i = )</td>
<td>( f )</td>
<td>( c )</td>
<td>( x )</td>
<td>( y )</td>
<td>( q^c + q^x + q^y )</td>
</tr>
<tr>
<td>( w_i = )</td>
<td>( q^f )</td>
<td>( q^c )</td>
<td>( q^x )</td>
<td>( q^y )</td>
<td>( q^f + q^y )</td>
</tr>
<tr>
<td>[xy]</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( q^c + q^x + q^y )</td>
</tr>
<tr>
<td>[y]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>( q^f + q^y )</td>
</tr>
<tr>
<td>[x]</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>( q^f + q^x )</td>
</tr>
</tbody>
</table>

- 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. Learning and Maturation
2. Exponential Harmonic Grammar, or \(q\)-HG
3. Consonant cluster simplification in Dutch
4. Summary
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] ≫ *[s] ≫ *[l] ≫ *[z] ≫ *[k] ≫ *[t]*
Thank you for your attention!

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

Many thanks to:
Pronoun resolution problem: data

The elephant is hitting him. The elephant is hitting himself.


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

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

<table>
<thead>
<tr>
<th>$q$</th>
<th>precision</th>
<th>$q$</th>
<th>precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>OT</td>
<td>0.500</td>
<td>1.4</td>
<td>0.790 ± 0.004</td>
</tr>
<tr>
<td>30</td>
<td>0.499 ± 0.008</td>
<td>1.3</td>
<td>0.847 ± 0.001</td>
</tr>
<tr>
<td>20</td>
<td>0.500 ± 0.012</td>
<td>1.2</td>
<td>0.911 ± 0.002</td>
</tr>
<tr>
<td>10</td>
<td>0.499 ± 0.003</td>
<td>1.15</td>
<td>0.945 ± 0.003</td>
</tr>
<tr>
<td>5</td>
<td>0.511 ± 0.001</td>
<td>1.10</td>
<td>0.978 ± 0.001</td>
</tr>
<tr>
<td>3</td>
<td>0.550 ± 0.005</td>
<td>1.08</td>
<td>0.986 ± 0.001</td>
</tr>
<tr>
<td>2.5</td>
<td>0.580 ± 0.003</td>
<td>1.06</td>
<td>0.994 ± 0.001</td>
</tr>
<tr>
<td>2.0</td>
<td>0.633 ± 0.003</td>
<td>1.05</td>
<td>0.997 ± 0.001</td>
</tr>
<tr>
<td>1.8</td>
<td>0.666 ± 0.003</td>
<td>1.04</td>
<td>0.9985 ± 0.0003</td>
</tr>
<tr>
<td>1.7</td>
<td>0.687 ± 0.007</td>
<td>1.03</td>
<td>0.9991 ± 0.0005</td>
</tr>
<tr>
<td>1.6</td>
<td>0.716 ± 0.006</td>
<td>1.02</td>
<td>0.99977 ± 0.00015</td>
</tr>
<tr>
<td>1.5</td>
<td>0.749 ± 0.008</td>
<td>1.01</td>
<td>0.99997 ± 0.00006</td>
</tr>
</tbody>
</table>