From Harmonic Grammar to Optimality Theory

The strict domination limit

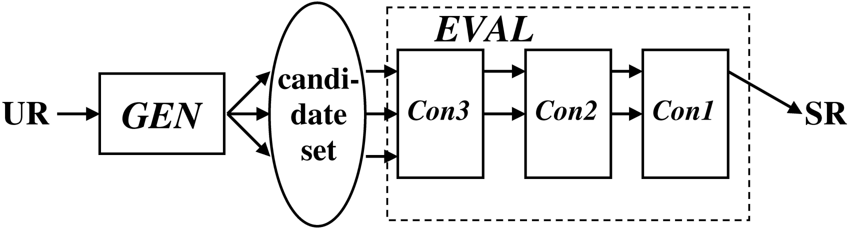
Tamás Biró

ELTE Eötvös Loránd University

'K + K = 120' Workshop
Dedicated to László Kálmán and András Kornai
on the occasion of their 60th birthdays,

RIL HAS, Budapest, 18 December 2017

Optimality Theory in a broad sense



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

Optimization as a linguistic architecture

The output/surface form optimizes an objective function:

$$\mathrm{SF}(u) = \underset{x \in \mathrm{Gen}(u)}{\arg\max}\; H(x)$$

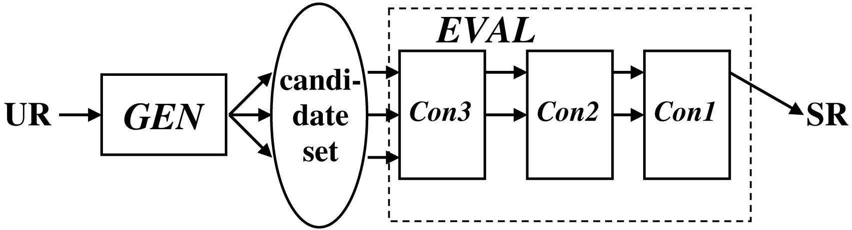
- **Harmony Grammar (HG):** $H_{\mathrm{HG}}(x) = -\sum_{k=1}^{n} w_k \cdot C_k(x)$.

  $\arg\max$ with respect to the arithmetic *greater than* relation $\geq$.

- **Optimality Theory (OT):**
  $H_{\mathrm{OT}}(x) = \Big(-C_n(x), -C_{n-1}(x), \ldots, -C_1(x) \Big)$.

  $\arg\max$ with respect to the *lexicographic order* relation $\succeq_{\mathrm{lex}}$.

Optimality Theory — strict domination



- The *filter series* view of OT **vs.** optimization.
- **Strict domination**: candidate once filtered out never comes back.
- Inherent in OT (lexicographic order), not necessary in HG.

Exponential Harmonic Grammar, or q-HG

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

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

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

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

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

$$w_i = q^{r_i}.$$

Exponential Harmonic Grammar, or *q*-HG

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

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

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

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

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

$$w_i = q^{r_i}.$$

Exponential Harmonic Grammar, or q-HG

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

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

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

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

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

$$w_i = q^{r_i}.$$

Exponential Harmonic Grammar, or q-HG

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

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

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

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

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

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

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

$$\begin{aligned}
-H(x) &= q^n \cdot C_n[x] + \ldots &+ q^2 \cdot C_2[x] + q^1 \cdot C_1[x] \\
-H(x) &= 2^n \cdot C_n[x] + \ldots &+ 4 \cdot C_2[x] + 2 \cdot C_1[x] \\
-H(x) &= 3^n \cdot C_n[x] + \ldots &+ 9 \cdot C_2[x] + 3 \cdot C_1[x] \\
-H(x) &= 10^n \cdot C_n[x] + \ldots &+ 100 \cdot C_2[x] + 10 \cdot C_1[x]
\end{aligned}$$

- 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:

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

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

$$\begin{aligned}
-H(x) &= q^n \cdot C_n[x] + \ldots & + q^2 \cdot C_2[x] + q^1 \cdot C_1[x] \\
-H(x) &= 2^n \cdot C_n[x] + \ldots & + 4 \cdot C_2[x] + 2 \cdot C_1[x] \\
-H(x) &= 3^n \cdot C_n[x] + \ldots & + 9 \cdot C_2[x] + 3 \cdot C_1[x] \\
-H(x) &= 10^n \cdot C_n[x] + \ldots & + 100 \cdot C_2[x] + 10 \cdot C_1[x]
\end{aligned}$$

- 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
☞ $x1$	-1			-2.25
$x2$		-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
☞ $x1$	-1			-2.25
$x3$		-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
$x1$	*		
☞ $x2$		*	*

- OT does not have *counting cumulativity* either:

	C_3	C_2	C_1
$x1$	*		
☞ $x3$		**	

(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
$x1$	-1			-9
☞ $x2$		-1	-1	-4

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

	C_3 9	C_2 3	C_1 1	H
$x1$	-1			-9
☞ $x3$		-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:

if $q \geq 1 + C_k[x]$ for all k and x.

New result: *upper bound not needed!*

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:

if $q \geq 1 + C_k[x]$ for all k and x.

***New result**: upper bound not needed!*

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

$$C_i : \bigcup_{u \in \mathcal{U}} \mathrm{Gen}(u) \to \mathbb{N}_0$$

$$H_q(x) = -\sum_{k=1}^{n} q^k \cdot C_k(x)$$

$$H_{\mathrm{OT}}(x) = \Big(-C_n(x), \ldots, -C_1(x) \Big)$$

$$\underset{\mathrm{OT}}{\mathrm{SF}}(u) = \underset{x \in \mathrm{Gen}(u)}{\arg\max}\, H_{\mathrm{OT}}(x)$$

$$\underset{q}{\mathrm{SF}}(u) = \underset{x \in \mathrm{Gen}(u)}{\arg\max}\, H_q(x)$$

Theorem

Given are non-negative integer constraints $C_n \gg C_{n-1} \gg \ldots \gg C_1$ and a Generator function Gen. Then, for any underlying form $u \in \mathcal{U}$ $\exists$ some threshold $q_0 \geq 1$ such that $\forall$ $q > q_0$, $\mathrm{SF}_{OT}(u) = \mathrm{SF}_q(u)$.

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

Theorem

Given are non-negative integer constraints $C_n \gg C_{n-1} \gg \ldots \gg C_1$ and a Generator function Gen. *Then, for any underlying form $u \in \mathcal{U}$ $\exists$ some threshold $q_0 \geq 1$ such that $\forall$ $q > q_0$, $\mathrm{SF}_{OT}(u) = \mathrm{SF}_q(u)$.*

Proof.

(Sketch.) Fix $u \in \mathcal{U}$. For $s \in \mathrm{SF}_{\mathrm{OT}}(u)$, let

$$q_0 = 1 + \max\left\{C_k(s), C_{k-1}(s), \ldots, C_1(s)\right\}$$

Then, for all $q > q_0$,

1. if $s_1 \in \mathrm{SF}_{\mathrm{OT}}(u)$ and $s_2 \in \mathrm{SF}_{\mathrm{OT}}(u)$, then $H_q(s_1) = H_q(s_2)$, but
2. if $s \in \mathrm{SF}_{\mathrm{OT}}(u)$ and $x \notin \mathrm{SF}_{\mathrm{OT}}(u)$, then $H_q(s) > H_q(x)$.

$\square$

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

Theorem

Given are non-negative integer constraints $C_n \gg C_{n-1} \gg \ldots \gg C_1$ and a Generator function Gen*. Then, for any underlying form $u \in \mathcal{U}$ $\exists$ some threshold $q_0 \geq 1$ such that $\forall\ q > q_0$, $\mathrm{SF}_{OT}(u) = \mathrm{SF}_q(u)$.*

Corrolary:

$$\lim_{q \to +\infty} \mathrm{SF}_q = \mathrm{SF}_{\mathrm{OT}} \qquad \text{pointwise.}$$

As q grows, more and more $u \in \mathcal{U}$ are mapped by q-HG onto $\mathrm{SF}_{\mathrm{OT}}(q)$.

The two languages converge: if q is large enough, then any u is expressed by the same $\mathrm{SF}(u)$.

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

Theorem

Given are non-negative integer constraints $C_n \gg C_{n-1} \gg \ldots \gg C_1$ and a Generator function Gen. Then, for any underlying form $u \in \mathcal{U}$ $\exists$ some threshold $q_0 \geq 1$ such that $\forall\ q > q_0$, $\mathrm{SF}_{OT}(u) = \mathrm{SF}_q(u)$.

Corrolary:

$$\lim_{q \to +\infty} \mathrm{SF}_q = \mathrm{SF}_{\mathrm{OT}} \qquad \text{pointwise.}$$

As q grows, more and more $u \in \mathcal{U}$ are **mapped** by q-HG onto $\mathrm{SF}_{\mathrm{OT}}(q)$.

The two languages converge: if q is large enough, then any u is **expressed** by the same $\mathrm{SF}(u)$. **Theory vs. implementation?**

Implementation: how to find the global optimum?

	C_2	C_1
[A]		*
[B]	*	*
[C]	*	

$$H(B) - H(A) = q^2$$
$$H(B) - H(C) = q$$

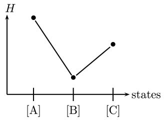
different magnitude

Candidate [A] is most harmonic.

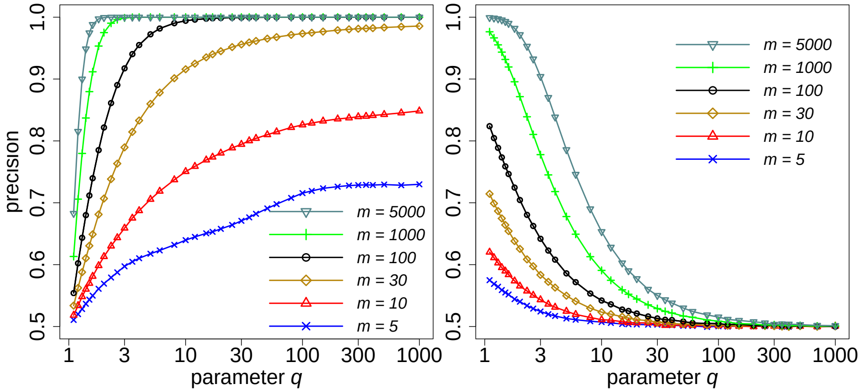
	C_2	C_1
[A]		
[B]	*	*
[C]		*

$$H(B) - H(A) = q^2 + q$$
$$H(B) - H(C) = q^2$$

same magnitude



Implementation: how to find the global optimum?



Precision of simulated annealing with different cooling schedules, for two different grammars, as a function of *q* (Biró, in prep.).

Summary: q-HG in the strict domination limit

- q-HG converges to OT as $q \to +\infty$:
- more and more inputs are mapped to the same outputs,
- and no cumulativity effects.

- And implementation may be prone to errors,
  both q-HG in the strict domination limit, and OT.

Summary: q-HG in the strict domination limit

- q-HG converges to OT as $q \to +\infty$:
- more and more inputs are mapped to the same outputs,
- and no cumulativity effects.

- And implementation may be prone to errors, both q-HG in the strict domination limit, and OT.

Thank you for your attention!

$60 \cdot K_{\mathcal{L}} + 60 \cdot K_{\mathcal{A}} = 120$ *Happy birthday!*

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

Acknowledgement: this research was supported by a *Marie Curie FP7 Integration Grant* (no. PCIG14-GA-2013-631599, "MeMoLI"), *7th EU Framework Programme*.