From Harmonic Grammar to Optimality Theory:

Production and maturation in q-HG

Tamás Biró*

Department of Assyriology and Hebrew Studies

ELTE Eötvös Loránd University, Budapest, Hungary

tamas.biro@btk.elte.hu, ORCID: 0000-0002-8943-9276

http://birot.web.elte.hu/, http://www.birot.hu/

Generative Linguistics in the Old World (GLOW 41) — 11 April 2018, Budapest

1. Optimality Theory (OT) and Harmonic Grammar (HG)

"*[C]ognition is computation.* This hypothesis permits the rigorous analysis of cognition – even at its most abstract – through a formal characterization of cognitive calculation. But computation is a rich notion that can be formalized in many ways. So the fundamental hypothesis of cognitive science – cognition is computation – immediately gives rise to the fundamental question of human **cognitive architecture**: just what *type* of computation is cognition?" (Smolensky & Legendre, 2006, vol. I, p. 5, emphases are original).

Basic building blocks:

- $\mathcal{U}$ — Set of **underlying forms**, a non-empty set (universal, cf. the *Richness of the Base Principle*).
- $\mathcal{X}$ — Set of potential **candidates/surface forms**, a non-empty set (universal).
- Gen — the **Generator function**, a one-to-many mapping $\mathcal{U} \rightarrow \mathcal{X}$ (postulated to be universal).
- $C_k(x)$ — elementary functions ("constraints" – a misnomer?), $\mathcal{X} \rightarrow \mathbb{N}_0$ (universal?), where $k = 1 \dots n$. NB: we suppose that the range of the constraints are the non-negative integers ("number of stars") although there are some exceptions to it in the linguistic literature.

Harmony function:

Optimality Theory:

For *hierarchy* $C_n \gg C_{n-1} \gg \ldots \gg C_k \gg \ldots \gg C_1$, use $H_{\mathrm{OT}}(x) = \Big(-C_n(x), -C_{n-1}(x), \ldots, -C_1(x)\Big)$.

Harmonic Grammar:

For *weight system* $w_n \geq w_{n-1} \geq \ldots \geq w_k \geq \ldots \geq w_1$, use $H_{\mathrm{HG}}(x) = -\sum_{k=1}^{n} w_k \cdot C_k(x)$.

.

Grammatical outputs (surface forms):

The grammatical output corresponding to an input $u \in \mathcal{U}$ **optimizes the target function** H:

$$\mathrm{SF}_{\mathrm{OT}}(u) = \operatorname*{arg\,max}_{x \in \mathrm{Gen}(u)} H_{\mathrm{OT}}(x) \qquad \mathrm{SF}_{\mathrm{HG}}(u) = \operatorname*{arg\,max}_{x \in \mathrm{Gen}(u)} H_{\mathrm{HG}}(x)$$

Questions: *What is the connection between HG and OT?*

2. q-Harmonic Grammar (q-HG)

To answer this question, **a formalism interpolating between HG and OT** is introduced:

q-Harmonic Grammars: use exponential weights $w_k = q^k$ for some $q > 1$. Hence,

$$H_q(x) = -\sum_{k=1}^{n} q^k \cdot C_k(x) \qquad \mathrm{SF}_q(u) = \operatorname*{arg\,max}_{x \in \mathrm{Gen}(u)} H_q(x)$$

Notes:

1. Without loss of generality, we can assume on this poster that constraint indices reflect constraint ranking.
2. More generally, constraint C_k could be assigned rank r_k, and then postulate weight $w_k = q^{r_k}$. Presently, however, we set $r_k = k$, in order to implement the OT constraint hierarchy $C_n \gg C_{n-1} \gg \ldots \gg C_1$ with the least *ad hoc* decisions. Our results can be applied – *mutatis mutandis* – to the more general case.
3. **Exponential HG** (Boersma & Pater, 2016 [2008]) considers the base q of exponentiation merely as a technical detail, whereas **q-HG** proposes a new *perspective* to view it as an interesting tunable parameter. Exponential HG is used for learning, and tunes the rankings r_k independently, while *q-HG contributes to our understanding of the relation between an HG and an OT grammar.*

Goal: to understand how OT emerges from HG, by observing the behavior of q-HG as q grows large.

The *strict domination limit:* $q \rightarrow +\infty$

With larger values of q, less cases of cumulativity (Jäger & Rosenbach, 2006) are encountered:

counting cumulativity

/u/	C_2	C_1	3-HG	5-HG	OT
	$r_2 = 2$	$r_1 = 1$			
$q = 3$	$w_2 = 3^2 = 9$	$w_1 = 3^1 = 3$			
$q = 5$	$w_2 = 5^2 = 25$	$w_1 = 5^1 = 5$			
[x]		****	-12	☞ -20	☞
[y]	*		☞ -9	-25	

ganging-up cumulativity

/u/	C_3	C_2	C_1	3-HG	5-HG	OT
	$r_3 = 3$	$r_2 = 2$	$r_1 = 1$			
$q = 3$	$w_3 = 27$	$w_2 = 9$	$w_1 = 3$			
$q = 5$	$w_3 = 125$	$w_2 = 25$	$w_1 = 5$			
[x]		**	****	-30	☞ -70	☞
[y]	*			☞ -27	-125	

3. Competence

For an OT grammar $C_n \gg C_{n-1} \gg \ldots \gg C_1$, a corresponding q-HG grammar can be constructed, for any $q > 1$. For which q **would they generate the same language,** i.e., map any u.f. to the same s.f.?

Theorem 1. *Given are non-negative integer constraints $C_n, C_{n-1}, \ldots, C_1$ (ordered by their indices) and a Generator function* Gen*. Then, for any underlying form $u \in \mathcal{U}$ there exists some threshold $q_0 \geq 1$ such that for all $q > q_0$, $SF_{OT}(u) = SF_q(u)$.*

Proof. Refer to Biró (2017). □

Corollary 2. *The language generated by q-HG* ***converges*** *to the language generated by OT in the strict domination limit:*

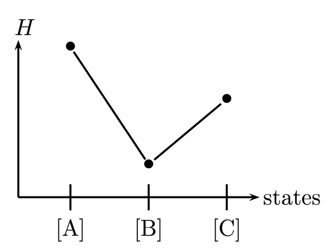
$$\lim_{q \to +\infty} SF_q = SF_{OT} \quad \text{pointwise.}$$

Notes:

1. Known since Prince & Smolensky 1993: *OT* and q-HG are equivalent, if $q \geq C_k(x) + 1$ for all k and $x \in \mathcal{X}$.
2. Does not necessarily hold if constraints are not integer-valued.

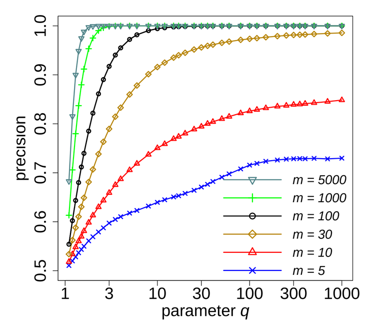
4. Performance

Implementation of a grammar with **simulated annealing** as a model of linguistic performance. Experiments with a 3-candidate landscape and different tableaux (Biró, 2017):



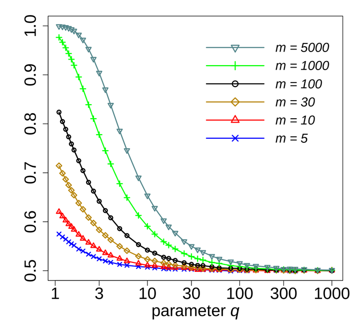
Three candidates, two of which are locally optimal.

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



$H(B) - H(A) = q^2$ and $H(B) - H(C) = q$ different magnitudes.

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



$H(B) - H(A) = q^2 + q$ and $H(B) - H(C) = q^2$ same magnitude.

***Precision of simulated annealing** with different cooling schedules, as a function of q*

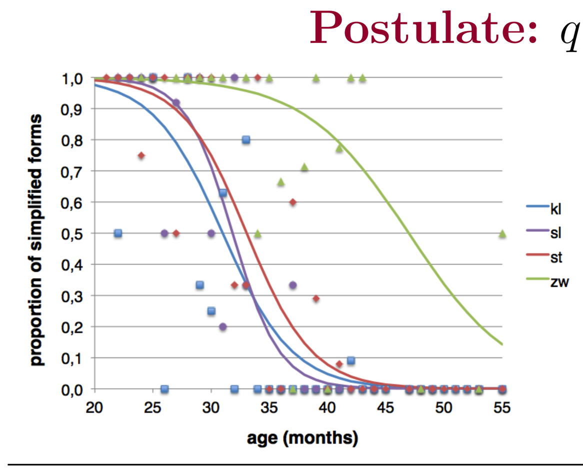
5. Language acquisition

Word initial consonant cluster simplification in Dutch child speech (collected from CHILDES by Klaas Seinhorst): [kl] → [k], [sl] → [l], [st] → [t], [zw] → [z], with significant production differences.

- Child has acquired Faithf ≫ NoComplOns earlier, probably already at pre-linguistic age.
- Relative ranks *[w] ≫ *[s] ≫ *[l] ≫ *[z] ≫ *[k] ≫ *[t], maybe motivated by *natural phonology.*
- $H_q([\mathrm{zw}]) - H_q([\mathrm{z}]) = q^8 - q^7 - q^6$, $H_q([\mathrm{kl}]) - H_q([\mathrm{k}]) = q^8 - q^7 - q^4$, etc.

C_i	Fthf	NoCmpl 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
$(1.5)^{r_i}$	25.6	17.1	11.4	7.59	5.06	3.38	2.25	1.5
2^{r_i}	256	128	64	32	16	8	4	2

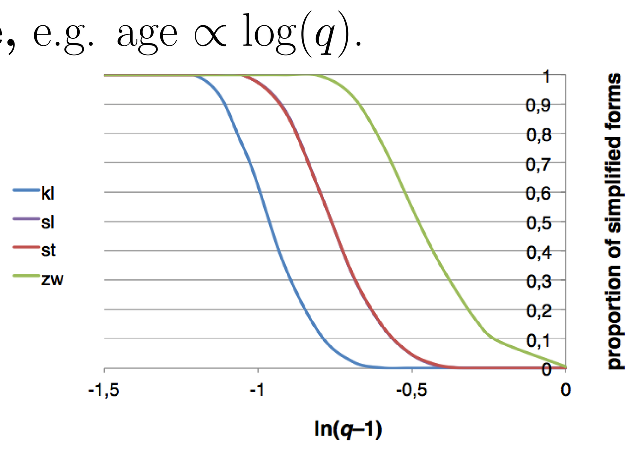
Postulate: q is a function of age, e.g. age $\propto \log(q)$.



Before maturation: small q, e.g., $q = 1.1$.

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

+ *Stochastic OT* →



6. Summary and "concluding hypotheses"

OT or HG? Biró (2017): a q-HG with a higher q – an HG closer to OT – is more prone to errors, but is faster to compute. Hence, in certain domains (in certain domains of certain languages?), grammars prefer a higher q (removing cumulativity effects); but in other domains they prefer a lower q (hence, some cumulativity).

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).

> Smolensky and Legendre (2006, vol. 1, p. 87) lists "the emergence of OT's strict domination constraint interaction (…) from network-level principles" as one of the major open problems in ICS. While it is unclear yet what mechanisms cause the emergence of strict domination in the brain, we now have a hypothesis for what motivates it to happen during maturation.

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.
- (Much of) **phonology** goes from HG to OT (q from $1 + \epsilon$ to large): speed ≫ precision.
- (Much of) **syntax-semantics** goes from OT to HG (q from large to $1 + \epsilon$): precision ≫ speed.

References

Biró, T. (2017). OT grammars dont count, but make errors: The consequences of strict domination for simulated annealing. In B. Gyuris, K. Mády, & G. Recski (Eds.), $K + K = 120$: *Papers dedicated to László Kálmán and András Kornai on the occasion of their 60th birthdays.* Budapest: Research Institute for Linguistics, Hung. Academy of Sciences.

Boersma, P., & Pater, J. (2016 [2008]). Convergence properties of a gradual learning algorithm for Harmonic Grammar. In J. J. McCarthy & J. Pater (Eds.), *Harmonic Grammar and Harmonic Serialism.* Equinox. (First published on ROA in 2008.)

Jäger, G., & Rosenbach, A. (2006). The winner takes it all – almost: Cumulativity in grammatical variation. *Linguistics*, *44*(5), 937–971.

Smolensky, P., & Legendre, G. (Eds.). (2006). *The Harmonic Mind: From neural computation to Optimality-Theoretic grammar.* Cambridge, MA – London, UK: MIT Press.

* *This research was supported by a Marie Curie FP7 Career Integration Grant (grant no. 631599, "MeMoLI") within the 7th European Union Framework Programme.*

#Istand withCEU

MARIE CURIE ACTIONS