From Performance Errors to Optimal Competence
Learnability of OT and HG with Simulated Annealing

Tamás Biró
ACLC, University of Amsterdam, t.s.biro@uva.nl

A self-evident, and yet too often ignored fact about (child) language acquisition is that the learner acquiring her *linguistic competence* is exposed to the teacher's *linguistic performance* – hence, also to performance errors, fast speech forms, or other variations. The *performance pattern*, which may be more complex than "simple random noise", could in theory render the learning problem extremely difficult, but a clever learning algorithm could also make use of the errors, thereby enriching the allegedly poor stimulus.

The computational approach employed in this paper has a threefold structure. Linguistic competence (both of the teacher, and of the learner) is modelled either by standard Optimality Theory (Prince and Smolensky 1993), or by a symbolic Harmonic Grammar with exponential weights (as discussed, for instance, in Biró 2009a). Performance patterns are produced either by always taking the most harmonic form, or by symbolic simulated annealing (Bíró 2006), an algorithm introducing performance errors as a function of the "speech rate". Finally, online learning employs either Paul Boersma's update rule (Boersma 1997), or Giorgio Magri's (2009) one.

The grammar ("phenomenon") studied is the abstract string grammar proposed by Biró (2007), arguably mimicking a simple but typical phonological grammar. As several constraint rankings or weight families correspond to the same language, the learner is not expected to converge to the teacher's competence (grammar), but to his performance (distribution of forms). In particular, the learner's distance from the teacher is measured by the *Jensen-Shannon divergence* between a sample of the teacher's performance pattern and a sample of the learner's performance pattern. The learner is said to have learnt the target language if this distance is smaller than the divergence of two random samples of the same size produced by the teacher.

Table 1 summarizes the results of an initial experiment (Biró 2009b). Magri's approach is significantly faster than Boersma's. If performance errors are present, then learning OT is faster than learning HG. Yet, we do not want to draw far-reaching conclusions from this toy grammar; so the talk will focus more on methodological issues of this novel approach, such as the "stability" of the learning process, its dependence on the initial conditions and the order of the learning data, etc.

		OT	10-HG	4-HG	1.5-HG
always gramm.	M	*13* ; **27** ; *45*	*13* ; **28** ; *46*	*12* ; **27** ; *48*	*15* ; **30** ; *47*
	B	*23* ; **43** ; *65*	*22* ; **41** ; *64*	*22* ; **42** ; *64*	*23* ; **40** ; *60*
sim. annealing,	M	*53* ; **109** ; *233*	*63* ; **140** ; *328*	*60* ; **148** ; *366*	*83* ; **199** ; *508*
$t_{\text{step}} = 0.1$	B	*80* ; **171** ; *462*	*92* ; **240** ; *772*	*92* ; **239** ; *785*	*117* ; **290** ; *694*
sim. annealing,	M	*64* ; **131** ; *305*	*62* ; **134** ; *304*	*63* ; **137** ; *329*	*72* ; **163** ; *437*
$t_{\text{step}} = 1$	B	*90* ; **212** ; *560*	*92* ; **233** ; *572*	*84* ; **212** ; *646*	*101* ; **242** ; *616*

Table 1: Comparing four competence models (OT vs. exponential HG with different bases), three performance algorithms (always the grammatical form vs. simulated annealing with different production speeds) and two learning methods (Boersma's update rule vs. Magri's update rule). For each combination, 2000 learning experiments were conducted, measuring the number of learning steps until convergence. A cell contains the 1st quartile, the median and the 3rd quartile of the distribution of these learning steps.

References

Tamás Bíró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.* PhD thesis, University of Groningen. Also as ROA-896.

Tamás Biró (2007). 'The benefits of errors: Learning an OT grammar with a structured candidate set'. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*, pages 81–88, ACL Prague, June 2007. ROA-929.

Tamás Biró (2009a). 'Elephants and Optimality Again. SA-OT accounts for pronoun resolution in child language'. To appear in: *Selected papers from CLIN 19.* ROA-1038.

Tamás Biró (2009b). 'Learning Competence from Performance Data'. Poster presented at *KNAW Academy Colloquium on Language Acquisition and Optimality Theory* July 2-3, 2009, Amsterdam. `http://www.birot.hu/publications/Biro-KNAW-poster-2009.pdf`.

Paul Boersma (1997). 'How we learn variation, optionality, and probability'. IFA Proceedings 21: 43-58.

Giorgio Magri (2009). 'New update rules for on-line algorithms for the Ranking problem in Optimality Theory'. Handout. LMA workshop, DGfS 31, Osnabrück, March 2009.

Alan Prince and Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar.* Rutgers U Center for Cognitive Science Technical Report 2. ROA-537. Malden, MA & Oxford: Blackwell.