Optimality Theory Syntax Learnability:
An Empirical Exploration of the Perceptron and GLA

Ann Irvine, Mark Dredze, Geraldine Legendre, and Paul Smolensky
Johns Hopkins University

This work brings together several threads of research on Optimality Theory (OT) and Harmonic Grammar (HG) learnability. As noted in previous work, including Pater (2008) and Magri (2010), the perceptron learning algorithm is well-established in the Machine Learning field and is a natural choice for modeling human grammar acquisition. The algorithm learns from one observation at a time, and it is capable of learning from a noisy corpus of observed natural language. In this work, we use the perceptron algorithm to learn a model that specifies a set of constraint weights relevant to one syntax phenomenon, Czech word order. We extract training data (sentences annotated with grammatical and information structure and their surface word orders) from the Prague Dependency Treebank (Hajic et al., 2001) and use basic alignment (edge-most) constraints on grammatical and information structure to predict the surface order of the subject, verb, and object. The perceptron algorithm learns a set of numeric, weighted constraints (a Harmonic Grammar). Ordering the constraints by the magnitude of their weights may specify a hierarchical constraint ranking (an OT Grammar), which is the essence of the classic Gradual Learning Algorithm (GLA) (Boersma, 1997). We describe and compare the two learning algorithms in detail and use a held out set of empirical data to quantitatively evaluate each. We show that by allowing for so-called *ganging-up-effects*, the more expressive Harmonic Grammar models Czech Word Order more accurately than the GLA OT grammar. Finally, crucially, it is also capable of modeling variation in production.

References

Boersma, P. (1997). How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Studies*, Volume 21, pp. 43–58.

Hajic, J., B. Hladka, and P. Pajas (2001). The prague dependency treebank: Annotation structure and support. In *In Proc. of the IRCS Workshop on Linguistic Databases*, pp. 105–114.

Magri, G. (2010). Hg has no computational advantages over ot: consequences for the theory of ot online algorithms. *Manuscript, IJN, ENS. Available as ROA-1104*.

Pater, J. (2008). Gradual learning and convergence. *Linguistic Inquiry 39*(2), 334–345.