

January 2001

Review of Bruce Tesar and Paul Smolensky (2000) Learnability in Optimality Theory

John J. McCarthy

University of Massachusetts, Amherst, jmccarthy@linguist.umass.edu

Follow this and additional works at: https://scholarworks.umass.edu/linguist_faculty_pubs

 Part of the [Morphology Commons](#), [Near Eastern Languages and Societies Commons](#), and the [Phonetics and Phonology Commons](#)

Recommended Citation

McCarthy, John J., "Review of Bruce Tesar and Paul Smolensky (2000) Learnability in Optimality Theory" (2001). *Trends in Cognitive Sciences*. 69.

Retrieved from https://scholarworks.umass.edu/linguist_faculty_pubs/69

This Article is brought to you for free and open access by the Linguistics at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Linguistics Department Faculty Publication Series by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Learnability in Optimality Theory

by Bruce Tesar and Paul Smolensky, MIT Press, 2000. \$25.00 (vii + 140 pages) ISBN 0 262 20126 7

Theories of language are often evaluated for how successfully they address the problem of language *learnability*: how can a learner equipped with this theory acquire the ambient language in a reasonable length of time with exposure to impoverished data? For example, in Chomsky's Principles and Parameters model, languages differ along certain dimensions called parameters. Learners discover how to set these parameters by attempting to analyze data that are incompatible with the current parameter settings¹.

Optimality Theory^{2,3} is a new approach to linguistic theory that has significant implications for how languages differ and how they can be learned. In OT, the grammar of a language is a hierarchy of defeasible constraints. If the constraints C1 and C2 are ranked in the order C1 \gg C2 ("C1 dominates C2"), then C2 will be violated when necessary to achieve better performance on C1. To cite a social rather than linguistic example, suppose I have a primary career goal of making lots of money (C1) and a secondary goal of living in an exciting city (C2). These are ranked constraints, and if I respect their ranking, then I will take a job in Paris, Texas over one that pays less but is based in Paris, France — even if the salary difference is only a few hundred dollars. OT constraints, then, are ordered in strict domination hierarchies, with higher-ranking constraints taking absolute priority over lower-ranking ones.

The null hypothesis in OT is that all constraints are universal. Accordingly, the grammars of languages differ only in the ranking of these universal constraints, just as

somebody else might prioritize salary and job location differently. The task of the learner is to acquire this ranking, and *Learnability in Optimality Theory* shows how that can be done.

The nucleus of LOT is the principle of *constraint demotion* and the learning algorithms based on it. Think of a grammar G as a function from inputs (entries in the mental lexicon, syntactic d-structures) to outputs (instructions to speech articulators, syntactic s-structures). In English, for instance, $G(\textit{John not eat}) \rightarrow \textit{John doesn't eat}$ and not $\textit{*John not eats}$. This means that $\textit{John doesn't eat}$ is better than $\textit{*John not eats}$ on the constraint hierarchy G — specifically, $\textit{John doesn't eat}$ beats $\textit{*John not eats}$ if and only if the highest ranking constraint in G on which they differ is one that favors $\textit{John doesn't eat}$ ⁴. Now suppose the learner's current grammar G_0 wrongly maps $\textit{John not eat}$ to $\textit{*John not eats}$. To correct this situation, it is sufficient to locate all the high-ranking constraints that favor $\textit{*John not eats}$ over $\textit{John doesn't eat}$ and demote them below the highest-ranking constraint that makes the opposite (and correct) judgment. In other words, for A to beat B , every constraint favoring B over A must be dominated by some constraint favoring A over B . Learning a constraint hierarchy consists of bringing this state of affairs into existence: every B -favoring constraint is demoted below some A -favoring constraint. By proceeding in this way, the learner is guaranteed to find some hierarchy consistent with the data, if there is any hierarchy to find. The most efficient algorithm, Error-Driven Constraint Demotion, will find a consistent hierarchy with no more than $N(N-1)$ informative examples, where N is the number of universal constraints.

Compare this to a random search. There are $N!$ ways to rank N constraints. This means that the search space expands rapidly as N grows larger — for example, $27!$ is about

10^{28} . (For comparison, the weight of the Earth is $6 \cdot 10^{27}$ g.) Random search in a space this large is a poor strategy for language learning. The key idea in constraint demotion, and the source of its efficiency, is that it starts with some ranking and gradually moves toward the right one, instead of starting with the whole ranking space and hunting the elusive snark.

LOT also addresses another closely-related issue. Most theories of language posit a significant amount of hidden structure: the full structural description of an utterance includes inaudible constituent structure; and, especially in phonology, the input has various non-obvious properties. For learners to use the principle of constraint demotion effectively, access to this hidden structure is essential: many constraints can only be evaluated with reference to the inaudible constituent structure; and access to the input is necessary for the class of *faithfulness* constraints, which favor $G(input) \rightarrow output$ identity mappings. This is the classic Catch-22 of language learning: you can't learn the grammar without knowing the hidden structure, but the hidden structure itself depends upon the grammar. This might be the hardest problem in studies of language learnability.

LOT's resolution of this paradox is to do everything at once, using the developing grammar to improve the analysis of the hidden structure, and using the imperfect but improving analysis of the hidden structure to aid the process of constraint demotion. The results so far are encouraging: when this approach is applied to learning the constraint ranking and the hidden constituent structure of linguistic stress systems, it rapidly converges on the right result 97% of the time. There is obviously much more to be done, but this is a good beginning on a problem that might have seemed hopeless.

Many developments in linguistic theory are initially driven by problems that arise in

analyzing and comparing languages. Considerations of language learnability enter the picture considerably later. Optimality Theory is different: the first work on language learnability, in Tesar's 1995 doctoral dissertation, is almost simultaneous with the emergence of the theory itself. The reason for this is not far to seek. LOT shows that learning an OT grammar is based on the central premise of the theory itself, the ranking of defeasible constraints. Special learning algorithms are not needed, nor are special parameter-setting cues in the ambient language. This tight integration of a linguistic theory and learnability is perhaps unprecedented and is LOT's greatest strength.

References

- 1 Bertolo, S., ed. (in press) *Language Acquisition and Learnability*, Cambridge University Press
- 2 Prince, A. and Smolensky, P. (1993) *Optimality Theory: Constraint interaction in generative grammar* (RuCCS-TR-2) Rutgers University Center for Cognitive Science
- 3 McCarthy, J.J. (in press) *A Thematic Guide to Optimality Theory*, Cambridge University Press
- 4 Grimshaw, J. (1997) Projection, heads, and optimality. *Linguistic Inquiry* 28 (3), 373-422

John J. McCarthy

Department of Linguistics

University of Massachusetts

Amherst, MA 01003 USA

Phone: 413-545-6830

Fax: 413-545-2792

E-mail: jmccarthy@linguist.umass.edu