CHAPTER I

INTRODUCTION

1. Summary of the dissertation

This dissertation presents Error-Selective Learning, an error-driven model of phonological acquisition in Optimality Theory which is both *restrictive* and *gradual*. It is restrictive in that it chooses grammars that can generate observed outputs but as few others as possible; it is gradual in that it requires numerous errors of the same kind to learn a new grammar. Together these two properties provide a model that can derive many observed intermediate stages in phonological development, while still explaining how learners eventually converge on the target grammar.

Error-Selective Learning is restrictive because its ranking algorithm is a version of Biased Constraint Demotion (BCD: Prince and Tesar, 2004), in which learners are biased to prefer rankings between classes of constraints, e.g. Markedness >> Faithfulness. BCD learners store the errors made by their current grammars in a table called the Support, and use their biases to choose the most restrictive ranking compatible with the Support. To account for a range of restrictiveness problems, the proposed learner uses a BCD algorithm with three ranking biases: (i) the well known Markedness >> Faith bias (Smolensky, 1996); (ii) a bias for high-ranking paradigm uniformity constraints (i.e. OO-Faith: Benua, 1997; McCarthy 1998); and (iii) a bias for ranking more specific IO-Faith constraints above more general ones (Smith, 2000; Hayes, 2004).

Error-Selective Learning is gradual because it uses a novel mechanism for introducing errors into the Support. As errors are made they are not immediately used to

learn new rankings, but rather stored temporarily in an Error Cache. Learning via BCD is only triggered once some constraint overcomes a Violation Threshold: that is, when some constraint has caused too many errors to be ignored. Once learning is triggered, the error-selective learner assesses the violation profiles of the errors currently in the Cache, and chooses the *best* error to add to the Support – an error that will cause minimal changes to the current grammar. Once the Support has been updated, the error-selective learner uses BCD to build a new ranking, empties their Error Cache and begins again.

In using Violation Thresholds and the proposed method of choosing best errors, the error-selective learner is sensitive to frequency in a way that accords with the connection between order of acquisition and input frequencies in child-directed speech. This reliance on frequency is also used to propose an extension of ESL which derives variation between intermediate stages, by introducing a stochastic notion of Violation Threshold.

Evidence for the Error-Selective BCD model comes from a wide survey of data in the literature, including new results from corpus data, as well as a novel artificial language experiment with children. Error-selective analyses are provided for several intermediate stages, relying on the M >> F and Spec-F >> Gen-F biases, including a detailed examination of the onset cluster acquisition of two children in the Compter/Streeter database. The artificial language study provides novel evidence of the third bias – for high-ranking OO-Faith – by showing that four year old children's repairs of unfamiliar coda-onset clusters in an 'alien' language were skewed in ways that kept derived plural words similar to their singular bases.

At the heart of all the dissertation's proposals is the Support, a stored repository of the data from which the BCD learner has built its current grammar. Through continual updates and revisions to the Support, the BCD learner remains restrictive even in the face of wrong structural analyses and missing data. The error-selective learner is therefore unlike the Gradual Learning Algorithm (Boersma, 1997 *et subsqt*), which has until now been the only OT learning algorithm to model gradual phonological learning. It is shown that the GLA is not well-suited to ongoing restrictive learning, even when equipped with a similar series of ranking biases, principally because it does not store its errors.

2. Structure of the dissertation

The first two chapters of the dissertation present the error-selective BCD learning model. Chapter 1 introduces the OT learning approach of Tesar and Smolensky (2000), the problems of restrictiveness in phonotactic learning, and the Biased Constraint Demotion solution. It synthesizes much recent work on restrictiveness in OT learnability theory, and focuses in detail on the implementation of the Specific >> General IO-Faith bias. A thorough method for discovering specific-to-general IO-faith relations is proposed, which compares the *contexts* of faithfulness constraints on a dynamic language-specific basis and uses them to impose the F-specificity bias.

Chapter 2 moves on to the facts of intermediate stages of phonological acquisition, which BCD algorithms on their own are not designed to model. It presents a body of evidence illustrating two kinds of intermediate stages, which both fall between initial and target grammars in their tolerance of marked structures, and then presents the Error-Selective model that derives these stages. It then spells out the role of frequency in

error-selective learning, as embodied in the Error Cache and its constraint violations, and demonstrates the connection between violation frequency and order of acquisition using cross-linguistic data from Germanic, Romance and English (e.g. Roark and Demuth, 2000.) Error frequencies and the Error Cache are then also used to propose a variable version of Error-Selective Learning.

Chapter 3 compares the error-selective BCD learner with an alternative, the stochastic OT learner that uses the Gradual Learning Algorithm (Boersma, 1997; Boersma and Hayes, 2001; Curtin and Zuraw, 2001.) Here it is shown through OTSoft simulations that the GLA must be augmented with all the same biases assumed above: both to learn a restrictive final grammar, and to pass through a full range of attested intermediate stages. Furthermore, the GLA learner can still be tricked into learning superset grammars if the learner makes incorrect assumptions about the learning data. The GLA is also demonstrated to fall short in learning restrictive grammars with lexical exceptions; the current state of learning with regard to exceptionality vs. free variation in the two models remains a central question for further research.

Chapter 4 returns to the notion of ranking biases in learning, using a novel experimental paradigm for artificial language research with children. The data from this experiment, which used a wug-test (Berko, 1958) with both novel morphological bases and a novel suffix, provides evidence that young learners bring a bias for paradigm uniformity to the task of learning novel phonotactic patterns. More generally, these positive results suggest that young children are both willing and able to participate in artificial language learning, pointing to a new source of data in the study of phonological acquisition.