Learning Phonological Grammars for Output-Driven Maps

Bruce Tesar
Department of Linguistics, Rutgers University, New Brunswick
tesar@rutgers.edu

The challenge of simultaneously learning a lexicon of underlying forms and a constraint ranking has been addressed by several scholars in recent work (Apoussidou 2007, Jarosz 2006, Merchant 2008, Tesar 2006). In particular, the proposal of Merchant, the Contrast Pair and Ranking information algorithm (CPR), avoids having to explicitly enumerate all possible underlying forms for each morpheme (in contrast to Apoussidou and Jarosz), and also avoids having to explicitly enumerate all possible constraint rankings (in contrast to Jarosz).

While CPR avoids those computational traps, there are still some components of CPR (and of the related work by Tesar) that pose computational difficulties. (1) The focus of CPR on lexical hypotheses for only a pair of related words at a time (a contrast pair) is a vast improvement over simultaneous consideration of all possible lexica, but the space of lexical hypotheses for a single contrast pair still grows exponentially in the number of unset underlying features for the morphemes involved in the pair. (2) The technique of initial lexicon construction, setting in advance features that do not alternate, can restrict further the number of lexical hypotheses that need to be considered, but at the cost of requiring that the learner have a complete paradigm of surface form data before learning of underlying forms can begin. (3) The extraction of ranking information performed by CPR is able to obtain ranking information from contrast pairs for which complete underlying forms have not yet been determined, but also faces exponential computational complexity, due in part to the fact that the procedure is separately computing the ranking implications of each lexical hypothesis in the (exponentially growing) set of possible hypotheses for the contrast pair.

The current paper demonstrates that each of these computational concerns can be significantly improved upon by taking the structure of grammars into greater consideration. The key grammatical structure lies in Tesar's proposal of output-driven map (Tesar 2008). Intuitively, an output-driven map is a phonological map in which all disparities introduced between the input and the output are motivated by conditions on the output. This notion is formalized by the requirement that any grammatical input-output mapping A→C entails the grammaticality of B→C whenever B has "greater similarity" to C than A does (A→C has every input-output disparity that B→C does, but B→C may lack some disparities of A→C); see Example 1. An output-driven map is necessarily a restricted identity map (Prince and Tesar 2004), meaning that every grammatical form maps to itself, a property assumed to hold of grammars in much learnability work, including that of Merchant. Output-driven maps can be viewed as a strengthened version of restricted identity maps.

The structure of output-driven maps can be exploited in learning via the contrapositive: B↛C entails A↛C. Given a grammatical output C, it is a given that C→C (restricted identity map property). Suppose B has one disparity with C (e.g., they differ in the value of one feature on one segment). If the learner possesses sufficient information to determine that B cannot map to C, then the learner need not bother checking to see if A maps to C; because the map is output-driven, any input which has, relative to C, all of the disparities of B plus additional ones cannot be grammatical. All lexical hypotheses which include all of the disparities of B→C may be dismissed without evaluation. Instead of needing to evaluate all combinations of possible values for all unset features of a word (exponential in the number of unset features), the learner can obtain the same information while only evaluating a single unset feature at a time (linear in the number of unset features), having the other unset features match (temporarily) the values of their output correspondents, addressing concern (1). If a word has eight unset binary features,

this means evaluating 8 lexical hypotheses instead of 256. Even greater benefit is realized when obtaining ranking information from forms with unset features, addressing concern (3).

The speed-up realized by exploiting the structure of output-driven maps is significant enough that initial lexicon construction is no longer needed. This frees the learner from needing an entire paradigm before learning commences; the learner can begin learning about underlying forms from even a single datum, addressing concern (2). This algorithm has the notable property that features of underlying forms which cannot be shown to require a particular value remain unset; non-contrastive features are never set, without any need for the learner to separately construct an "inventory of contrastive features".

Examples

Example 1: Output-driven map relationships using stress and length. Stress is indicated with an accent mark on the vowel, while length is indicated with [ː]. B has "greater similarity" to C than A does to C because the only disparity between B and C, B's lack of stress on the first syllable, is shared by A, while A has an additional disparity with C, length on the first syllable.

Output-Driven Map Property	A = paka B = paːka C = páːka
Direct: A→C entails B→C	/paka/→[páːka] entails /paːka/→[páːka]
Contrapositive: B↛C entails A↛C	/paːka/↛[páːka] entails /paka/↛[páːka]

References

Apoussidou, Diana. 2007. The Learnability of Metrical Phonology. PhD. dissertation, University of Amsterdam, Amsterdam.

Jarosz, Gaja. 2006. Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory. PhD. dissertation, The Johns Hopkins University, Baltimore, MD. ROA-884.

Merchant, Nazarré. 2008. Discovering underlying forms: Contrast pairs and ranking. PhD. dissertation, Rutgers University, New Brunswick. ROA-964.

Prince, Alan, and Tesar, Bruce. 2004. Learning phonotactic distributions. In *Constraints in Phonological Acquisition*, ed. by René Kager, Joe Pater, and Wim Zonneveld, 245-291. Cambridge: Cambridge University Press.

Tesar, Bruce. 2006. Learning from paradigmatic information. In *Proceedings of the 36th Meeting of the North East Linguistics Society*, ed. by Christopher Davis, Amy Rose Deal, and Youri Zabbal, 619-638. GLSA. ROA-795.

Tesar, Bruce. 2008. Output-driven maps. Ms., Linguistics Dept., Rutgers University. ROA-956.