Learning Both Variability and Exceptionality in Probabilistic OT Grammars

Learning Both Variability and Exceptionality in Probabilistic OT Grammars

Aleksei Nazarov, University of Huddersfield

The existence of both variability and exceptionality (Coetzee and Pater 2011) makes learning a (phonological) grammar more than finding a single constraint ranking compatible with all data points. Rather, this task includes finding a partition of inputs and a range of constraint rankings that together define just the attested range of variability and exceptionality in the language. (Learning lexical listing, Zuraw 2000, is set aside as a separate problem here.)

Modern Hebrew (Temkin-Martínez 2010; henceforth: TM) has an optional spirantization process that applies post-vocalically (an instance of variability; (1a)). This variable spirantization pattern is statistically dominant, but there are exceptional words in which spirantization either never applies, (1b), or always applies, (1c) (TM:Appendix B; p≤0.1 forms assumed ungrammatical). TM shows that all bolded segments must be underlyingly stops (fricatives yield a different pattern). Thus, the grammar partitions words into non-undergoing, (1b), always-undergoing, (1c), and default items, (1a). In addition, for default items, it must allow variability between stops and fricatives in just the right environment.

(1) a. /me**k**ase/ → [me**χ**ase, me**k**ase] | b. /da**k**ar/ → [da**k**ar] ***χ** | c. /ma**k**ar/ → [ma**χ**ar] ***k**

To deal with variability, learners with probabilistic ranking (e.g., Boersma 1998, Jarosz 2015) or weighted constraints (e.g., Goldwater and Johnson 2003) have been proposed. For exceptionality, implemented learners exist (Becker 2009, Coetzee 2009) for inferring lexically indexed constraints (Kraska-Szlenk 1995, Pater 2000). However, these learners use a form of Constraint Demotion (Tesar 1995), which cannot deal with variability. Thus, no existing OT learner can learn both variability and exceptionality while starting from an unindexed constraint set. I propose exactly such a learner, by showing that the logic of the existing indexed constraint learners (see above) can be implemented in Jarosz's (2015) probabilistic Expectation Driven Learning (EDL) framework, which represents variability through probabilities over constraint pair rankings:

(2) Constraints: {A,B,C} P(A>>B)=0.7 P(A>>C)=0.5 P(B>>C)=0.2

Every time the grammar is used, a full ranking is sampled (cf. Boersma 1998) that uses these pairwise probabilities as well as conditional relations between rankings (see Jarosz 2015 for details). The probabilities themselves are learned as follows: given an initial state (in this case, a uniform distribution for each ranking), it iterates the following procedure. For each /input/ and each constraint pair {A,B}, it temporarily sets A >> B and then B >> A while keeping all other probabilities constraint, and for each temporary grammar takes a fixed number (r = 50) of samples; for both sets of samples, it determines the number of matches: (3a). From these numbers of matches, it computes expected success counts: (3b). These expected success counts are then turned into probabilities as in (3c).

(3) a. N of successes in r samples from temp. grammar with categorical A >> B = $\text{Success}_{A>>B}$
N of successes in r samples from temp. grammar with categorical B >> A = $\text{Success}_{A>>B}$

b. Expected success counts: $E[\text{Success}_{A>>B}] = \text{Success}_{A>>B} \times P(A>>B)$
$E[\text{Success}_{B>>A}] = \text{Success}_{B>>A} \times [1 - P(A>>B)]$

c. $$P(A \gg B)_{dataset} = \frac{\sum_{/inputs/} E[Success_{A\gg B}]}{\sum_{/inputs/}(E[Success_{A\gg B}]+E[Success_{B\gg A}])}$$

It is these pairwise probabilities that make it possible to incorporate into a probabilistic learner the idea presented in Pater (2010), Becker (2009), and Coetzee (2009) that an indexed constraint is induced whenever inconsistency is detected, i.e., whenever two groups of words require two opposite rankings: e.g., *Stop >> Ident vs. Ident >> *Stop.

Proceedings of the Society for Computation in Linguistics (SCiL) 2018, pages 221-222.
Salt Lake City, Utah, January 4-7, 2018

The current learner uses a "soft inconsistency" criterion, which approximates the above criterion of opposite ranking requirements, to induce indexed constraints: if the lexicon assigns at least 60% probability to ranking A >> B, then any word that assigns 40% or less probability to ranking A >> B is exceptional w.r.t. constraint pair {A,B}. At every iteration of the EDL algorithm above, the learner finds which constraint pairs have exceptions according to this criterion, finds which of these constraint pairs has the greatest ranking preference difference between exceptions and the lexicon, and then induces a corresponding indexed constraint. While opposite tendencies between the cross-lexical trend and exceptional words thus lead to the induction of lexical constraints, variability that is consistent across the lexicon only triggers probabilistic ranking of existing constraints. This enables the learner to distinguish between variability and exceptionality.

This was confirmed with tests on a simplified abridged version of the Hebrew data, which only had variability in postvocalic position, as in (1); variability in non-postvocalic stops was leveled. The relative frequencies of each word type were estimated from the 29 roots with a single non-word-final stop in TM's experimental data and scaled to 12 data points. Only underlying stops /b,p,k/ were considered here (TM: underlying fricatives behave differently). For this version of the learner, the relative magnitude of a word's variability was not considered. In 20 runs of up to 80 iterations each with $r = 50$, the learner reached ≤5% likelihood of error within 11.35 iterations on average. The rate of exceptional words' being connected to an indexed constraint was 91% (100% for the 3 words like (1c), 65% for the single word like (1b)), while the corresponding rate for a non-exceptional word was 3% (0% for words like (1a)). Furthermore, for default words, the grammar predicted the attested range of variability in postvocalic position with 96% accuracy, and predicted lack of variation in non-postvocalic position with 95% accuracy. Thus, the learner was able to distinguish variability from exceptionality: non-exceptional words that had variable outputs, (1a), were never given an exceptionality index, while exceptional words, which did not exhibit variability post-vocalically, (1b,c), were given such an index at an overall rate of 91%. In 19 of the 20 runs, Faithfulness played no role in accounting for exceptions: the fact that only underlying stops were considered here does not seem essential to the result obtained.

References BOERSMA, Paul. 1998. "Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives." PhD dissertation, U of Amsterdam. • BECKER, M. 2009. "Phonological Trends in the Lexicon: The Role of Constraints." PhD dissertation, UMass Amherst, Amherst. • COETZEE, Andries W. 2009. "Learning Lexical Indexation." *Phonology* 26 (1): 109–45. • COETZEE, Andries W, and Joe Pater. 2011. "The Place of Variation in Phonological Theory." In: J.A. Goldsmith, J, Riggle, and A.C. Yu (eds.), *Handbook of Phonological Theory*, 2nd ed., 401–34. Wiley-Blackwell. • GOLDWATER, S., and M. Johnson. 2003. "Learning OT Constraint Rankings Using a Maximum Entropy Model." In: J. Spenader, A. Eriksson, and Ö. Dahl (eds.), Proceedings of the Stockholm Workshop on Variation within Optimality Theory, 111–20. • JAROSZ, G. 2015. "Expectation Driven Learning of Phonology" Ms., UMass Amherst. • KRASKA–SZLENK, I. 1995. "The Phonology of Stress in Polish." PhD dissertation, U of Illinois, Urbana-Champaign. • PATER, J. 2000. "Non-Uniformity in English Secondary Stress: The Role of Ranked and Lexically Specific Constraints." *Phonology* 17 (2): 237–74. • PATER, J. 2010. "Morpheme-Specific Phonology: Constraint Indexation and Inconsistency Resolution." In: S. Parker (ed.), *Phonological Argumentation: Essays on Evidence and Motivation*, 123–54. London: Equinox Press. • TEMKIN–MARTÍNEZ, M. 2010. "Sources of Non-Conformity in Phonology: Variation and exceptionality in Modern Hebrew Spirantization." PhD dissertation, USC. • TESAR, B. 1995. "Computational Optimality Theory." PhD dissertation, U of Colorado. • ZURAW, K. 2000. "Patterned Exceptions in Phonology." PhD dissertation, UCLA.