Contrast Analysis Aids the Learning of Phonological Underlying Forms

John Alderete*, Adrian Brasoveanu+, Nazarré Merchant+, Alan Prince+, Bruce Tesar+
*Simon Fraser University
+Rutgers University, New Brunswick

Phonological learning requires testing both grammars and lexical hypotheses, facing a potentially huge number of combinations of the two; see Hale & Reiss (1997), Albright & Hayes (1999), Tesar et al. (2003) and the references therein. We demonstrate the benefits of attending to surface contrasts between morphemes in order to obtain information about underlying forms prior to grammar-testing. A system of 24 abstract languages was defined via prosodic constraints in concert with the possibility of lexical specification of stress and vowel length. An algorithm for analyzing contrasts between morphemes, Contrast Analysis (Tesar, 2004), successfully determined the underlying values of most of the alternating features in all but four of the languages. Using contrast information to set underlying values significantly simplified the task of using constraint ranking / lexicon interaction to determine the rest of the grammar.

The linguistic system has six constraints. Two constraints align main stress to the left and to the right of the word (MAINL and MAINR); another is violated by long vowels on the surface (NOLONGV). Two faithfulness constraints, LENGTHF and STRESSF, are violated by segments that don't match their underlying correspondents in vowel length and stress. A sixth constraint, WEIGHT-TO-STRESS, requires that surface long vowels receive stress. This constraint causes interaction between the features, which complicates the learning of underlying feature values. For instance, one grammar has the ranking (1).

(1) LENGTHF ≫ WEIGHT-TO-STRESS ≫ STRESSF ≫ MAINL ≫ {MAINR, NOLONGV}
(2) Two of the permitted forms: /pa+ka/➔*pá+ka* /pa+ka:/➔*pa+ká:*

The output forms *pá+ka* and *pa+ká*: contrast in stress placement on the surface, but that is because of interaction with vowel length, not because of a difference in underlying stress specification. The learner needs to determine which surface differences are due to direct preservation of underlying differences, and which are due to surface feature interaction via phonological processes.

The learning situation is conceptualized as follows. The learner has access to morphologically analyzed outputs, indicating which parts of a word are associated with which morphemes. The learner starts by constructing an initial lexicon of underlying forms for morphemes, in which features that do not alternate on the surface are set irrevocably to their (unchanging) surface realization, while alternating features are marked as not yet set. Contrast Analysis (CA) then attempts to determine the correct underlying value for the alternating features by comparing the output realizations of morphemes that surface differently in the same environment. Once CA is complete, the surgery learning algorithm (Tesar et al., 2003) is used to determine the rest of the grammar via ranking / lexicon interaction.

CA compares the surface realizations of two morphemes in a particular morphological environment. If they surface differently, the learner determines the features that distinguish the surface realizations. At least one feature on which the surface realizations of the morphemes differ must be a faithful reflection of the underlying values for that feature. CA examines the lexical entries for each of the morphemes, to see if the set features can account for the contrast. If not, it checks to see if more than one of the surface differing features could possibly account for the contrast (by setting an unset feature). If only one feature could account for the contrast, then the learner sets the underlying value of that feature for each morpheme to match its surface realization.

Suppose we expand the earlier example into the following (the previous forms are r1+s1 and r1+s2).
r1+s1 *pá+ka* r1+s2 *pa+ká*: r1+s3 *pa+ká*: r2+s1 *pá:+ka* r2+s2 *pá:+ka:* r2+s3 *pa:+ká:*
Suffixes s2 and s3 differ following r2, but both are long on the surface in that environment, so they must differ in underlying stress: s2=/-ka:/, s3=/-ká:/. This solves the example in (2): r1+s2 *pa+ká:* must be due to stress attracted to weight, not to an underlying stress on s2.

Lexicon construction with Contrast Analysis is able to determine the entire lexicon for 18 of the 24 languages, setting as many as 6 alternating features. For another 2 languages, it determines all but one feature in the lexicon. In the other four languages, most or all of the features that CA cannot set are completely inert in the target language (either setting of the feature will yield the same output), and at most one active feature is left unset. CA sets features efficiently, on the basis of individual comparisons; it needn't reason across many data forms and constraint rankings, as the surgery algorithm does. These results suggest that morpheme contrast analysis plays a significant role in language learning generally, greatly reducing the amount of the lexicon determined via constraint ranking / lexicon interaction.

References

Albright, Adam, and Hayes, Bruce P. 1999. An automated learner for phonology and morphology. Ms., Linguistics Dept., UCLA.

Hale, Mark, and Reiss, Charles. 1997. Grammar Optimization: The simultaneous acquisition of constraint ranking and a lexicon. Ms., Concordia University, Montreal. ROA-231.

Tesar, Bruce, Alderete, John, Horwood, Graham, Merchant, Nazarré, Nishitani, Koichi, and Prince, Alan. 2003. Surgery in language learning. In *Proceedings of the Twenty-Second West Coast Conference on Formal Linguistics*, ed. by G. Garding and M. Tsujimura, 477-490. Somerville, MA: Cascadilla Press. ROA-619.

Tesar, Bruce. 2004. Contrast analysis in phonological learning. Ms., Linguistics Dept., Rutgers University. ROA-695.