The problem. Affixes may impose selectional restrictions on the phonological shape of the base. For example, some affixes combine only with bases that have certain phonological characteristics (e.g., the English suffix [-n] attaches only to obstruent-final monosyllabic adjectives such as stiff-en, black-en but not to sonorant-final ones, *full-en, *calm-en or disyllabic ones, *aloof-en; Siegel 1974). Alternatively, an affix may be realized as one of several suppletive allomorphs (Carstairs 1988 et seq.) depending on the phonological context (e.g., in Russian, “about” is [a] before consonant-initial words, as in [a nɪˈnɪj] ‘about Nina’, vs. [ab] before vowel-initial ones, as in [ab ənɪˈi] ‘about Anna’). The question is how such phonological selectional restrictions are expressed in the grammar.

Approaches to formalizing phonological selectional restrictions include subcategorization frames (Lieber 1980, Paster 2004), universal markedness constraints in an emergence of the unmarked ranking (Mascaró 1996), and Generalized Alignment (McCarthy and Prince 1993). We argue that these theories fail to capture aspects of phonological selectional restrictions. For example, universal markedness can explain the Russian case: ONSET rules out *[a ənɪˈi], and NoCODA rules out *[ab ēnɪˈni]. But this approach fails for Haitian Creole determiner allomorphy [-a/-la], a pattern that is the opposite of Russian: [-la] after consonants [liːˈna] ‘the book’ and [-a] after vowels [bɔk-a] ‘the sorcerer’. Such arbitrary selectional restrictions can be stated in terms of Generalized Alignment or subcategorization frames, but these approaches struggle with restrictions that are negative or non-local. The English superlative “-est” has both properties: it attaches to trochaic feet (e.g., (tall)est (prettiest)est), but not if the foot is preceded by an unstressed syllable, as in *ob(scene)est, and “-est” tends to not attach to adjectives that end in [st], e.g. ?moist-est (Mondorf 2009). Alignment can capture the requirement to attach to the right edge of a foot, but it cannot explain why ending in [st] makes affixation less likely, or why an initial unstressed syllable should prevent suffixation. What these cases do have in common is phonotactic generalizations about bases for affixation: they tend to begin in a consonant, or constitute trochees and not end in certain consonant clusters.

The proposal: sublexical phonotactic grammars. We extend the proposal in Becker and Gouskova (2012), whereby the lexicon is partitioned based on morphological and phonological subpatterns. For the Russian example, words that combine with [ab] form one sublexicon, and words that combine with [a] will form another. Generalizations are then learned for each sublexicon based on which phonological structures are under- or unattested (Hayes and Wilson 2008). The sublexicon for [ab] respects the restriction that word-initial words begin in a vowel; all of the words in the [a] sublexicon begin in a consonant. These restrictions are not universal or substantively grounded but learned from the lexicon, so they can be arbitrary. They can also be non-local, because they are learned over the entire base rather than from the edge of attachment: the English sublexicon for “-est” obeys a non-local generalization, i.e., the ban on initial unstressed syllables or multiple stresses, as well as the local dispreference against st#. Well-formedness of a base with respect to the sublexical phonotactic grammar predicts the probability of its combining with the affix.

Case study: Russian diminutives. Russian has four allomorphs for diminutivizing masculine nouns, [-ik/~fik/~ok/~ok]: [butarbróːt]~[butarbróːd-ik] ‘sandwich (+dim)’, [ajfon]~[ajfon-fik] ‘iPhone (+dim)’, [fɪɡiːl]~[fɪɡiːl-ik] ‘weathervane (+dim)’, and [bɪbɔnok]~[bɪbɔnof-ak] ‘child (+dim)’. Polivanova (1967) identifies many phonotactic generalizations about the stress, syllable structure, and segmental makeup of the base nouns: [-ik] does not combine with dorsal-final bases and prefers final stress; [-ok] does not combine with bases that end in a consonant cluster or have medial stress; [-fik] prefers stress-final bases that end in a labial or a sonorant but not a cluster. Polivanova’s description was verified in an elicitation study with Russian speakers. The study confirmed most of Polivanova’s generalizations and uncovered a new, non-local constraint: words with hiatus tend to not form [-ok] diminutives (e.g., [kloun] ‘clown’~*[kloun-ok], [opɪnum] ‘opium’~*[opɪnum-ok]).

For an independent confirmation of these patterns, we tested whether a machine learner would
identify non-local and arbitrary constraints. We constructed sublexicons of Russian nouns by searching the Russian portion of the Google Ngrams corpus, and we trained a machine phonotactic learner (Hayes and Wilson 2008) on each sublexicon of bases. The learner induced a constraint against hiatus for the [-ok] sublexicon but not for Russian nouns in general; it also dispreferred words with medial stress in the [-ok] sublexicon. For the [-ik] and [-ifik] sublexicons, the learner induced constraints against word-final dorsals and non-final stress; it did not mind word-final clusters in the [-ik] sublexicon but dispreferred them when trained on [-ok] bases, mirroring people’s behavior:

![Figure 1: People vs. the UCLA Phonotactic Learner: the treatment of base-final clusters in the [-ik] and [-ok] subgrammars](image)

Our case study reveals the existence of non-local and arbitrary selectional restrictions in Russian diminutive allomorphy, suggesting that Russian speakers have detailed phonotactic knowledge of sublexicons from which these restrictions are derived. Our proposed sublexical phonotactic grammars thus capture selectional properties that are unaccounted for by existing formalizations focusing on word edges.

References


