Michael Becker (Indiana University) **Friday 11:15**
& Maria Gouskova (New York University)

Source-oriented generalizations as grammar inference in yer deletion

Introduction. In constraint-based theories and schema-based theories, markedness constraints and schemas express generalizations that are *product-oriented*: a process applies if the output satisfies certain requirements (Bybee & Slobin 1982, Bybee & Moder, 1983, Bybee 2001). In rule-based grammars, generalizations are *source-oriented*: a rule applies to inputs that have a certain phonological shape (Albright & Hayes 2003). In Russian, for example, mid vowels ("yers") are deleted from the last syllable of some stems when a vowel-initial suffix is added, but high vowels are never deleted, e.g. [mox ∼ mx-a] 'moss NOM/GEN', *[mux ∼ mx-a] (Lighner 1965, Halle 1973, Yearley 1995, inter alia). The product of the deletion lacks the vowel; the generalization has to be stated over the source of the derivation.

In this paper, we show that deletion of a yer (a mid vowel) in Russian is subject to both source-oriented and product-oriented generalizations. We model the result using multiple constraint-based grammars, capturing an opaque generalization using phonotactic grammars learned over subsets of the lexicon.

The Russian lexicon. Yer deletion is governed by several product-oriented generalizations. For example, when yers are deleted, resulting CCC clusters predominantly have a medial obstruent, e.g. [kastʲor ∼ kastra] 'fire NOM/GEN'. Deletion is normally blocked if it puts a sonorant between consonants, e.g. [mudr̥ets ∼ mudr̥etsa] 'wise NOM/GEN', *[mudr̥tsa] (Yearley 1995, Gouskova & Becker to appear).

On the other hand, there is a constraint on the source of deletion: yer deletion can create tri-consonantal clusters, as in [kastʲor ∼ kastra], but only if the base ends in a simple coda, not a complex coda, *[kasotr̥ ∼ kastra]. This is a source-oriented generalization, since once the yer vowel is deleted, the produced cluster no longer retains the source syllabification.

Experiment. To test whether these generalizations are productive, we used a "wug test" (Berko 1958). A group of 115 Russian speakers each rated 48 nonce words, created from a pool of 403 consonant combinations (∼14 responses per consonant combination). Each nonce word was presented in the nominative base form, e.g. [s̪om], and the participant rated it on a scale of 1–5. Then, two genitives were shown, the faithful [s̪oma] and the yerless [s̪ma], in random order, and the participant gave each genitive a binary judgment as acceptable or unacceptable. The binary judgments of the yerless genitives are plotted in Figure 1 by the shape of the base.

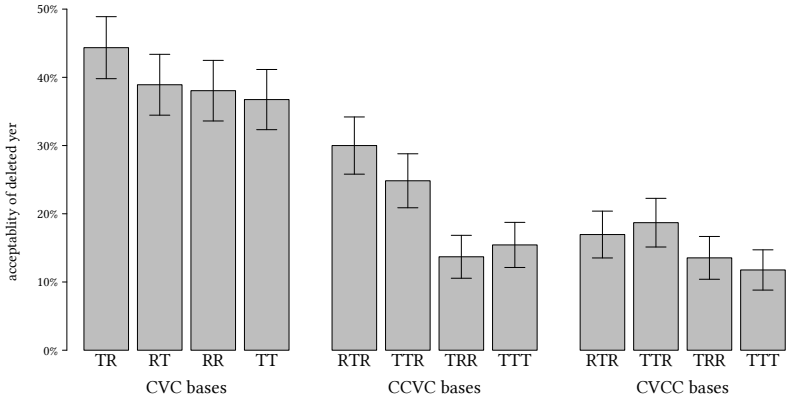

Figure 1: Acceptability of yerless genitives, organized by the cluster position and sonority profile. T=obstruent, R=sonorant.

As Figure 1 shows, the participants found deletion most acceptable when it created TR, RTR & TTR clusters (ending in an obstruent followed by a sonorant), confirming the productivity of this product-oriented generalization. They also preferred deletion that created CC clusters over CCC clusters (s̪om ∼ s̪ma > p̪s̪om ∼ p̪s̪ma). Among the CCC clusters, they preferred clusters that originated from bases with a simple coda over those with a complex coda (p̪s̪om ∼ p̪s̪ma > po̪s̪m ∼ p̪s̪ma), confirming the productivity of this source-oriented generalization. All of these effects were highly significant ($p < .0001$) in a mixed-effects logistic regression model using the *lme4* package in R.

Michael Becker (Indiana University) **Friday 11:15**
& Maria Gouskova (New York University)

Analysis. Russian speakers accept yer deletion not only based on the goodness of the created product, but also on the plausibility of the base as a yer word. We propose that once speakers identify words that undergo a process, such as vowel deletion, these words are separated from the general lexicon, and trigger the formation of a new constraint-based grammar. The lexicon is partitioned, and phonotactic generalizations are learned on each partition separately (cf. Ito & Mester 1995, Zuraw 2000, Pater 2006, 2008, et seq).

In Russian generally, complex codas are quite common, i.e. the constraint *COMPLEXCODA has a low weight/ranking. Among yer words, however, complex codas are not allowed; learning a separate phonotactic grammar for the yer words allows *COMPLEXCODA to remain high for those words. When the speakers have two sublexicons/grammars, a new word needs to be associated with one of them. Given a word like [poʂm], it is quite acceptable as a word of Russian generally, and therefore [poʂma] is an acceptable genitive. However, [poʂm] is unacceptable as a yer word due to its complex coda, thus making [pʂma] an unacceptable genitive. This is a *grammar inference* mechanism: when the speakers judge [poʂm ∼ pʂma], they need to decide how likely this paradigm is given the yer grammar. Given Bayes' theorem, this likelihood is proportional to the likelihood that the yer grammar assigns to the paradigm. Since the yer grammar has a high-ranking *COMPLEXCODA, the yer grammar assigns a low probability to the base [poʂm], and therefore also to its genitive [pʂma]. There is nothing in the yer grammar that prohibits the produced cluster in [pʂma]; the generalization is source-oriented, and it is expressed by the grammar inference mechanism.

This mechanism can also extend to other phenomena, such as the propensity of English ablauting verbs (e.g. *drink* ∼ *drank*) to end either in a velar (*sneak, dig*) or in a nasal (*swim, win*), but ideally both – a velar nasal (Albright & Hayes 2003). These preferences can be captured with a separate phonotactic grammar that covers just these verbs. The multiple phonotactic grammar approach can also offer a solution to otherwise puzzling results, such as the tendency in the Hungarian dative to prefer the [nɛk] allomorph with bases that end in a complex coda, or a sibilant, or a coronal sonorant (Hayes et al. 2009).

Learning simulation. To diagnose which of the generalizations that speakers extend to nonce words were source-oriented and which were product-oriented, we used the UCLA Phonotactic Learner (Hayes & Wilson 2008), trained on a list of 1,902 real yer words from Zaliznjak's (1977) dictionary, and tested on the same nonce words that the participants judged. When the learner was trained and tested on the nominative sources, it learned a constraint against complex codas, but it overestimated the acceptability of TRR items. When the learner was trained and tested on yerless genitives, it failed to find a difference between simple and complex codas, but it correctly found the preference for RTR and TTR clusters. Trained this way, then, the phonotactic learner can identify generalizations that are present only in the source or only in the product. The two grammars were combined (simply adding their penalties), yielding a grammar that captured both kinds of generalizations.

Conclusion. We identified several generalizations about the distribution of yer deletion in Russian, and showed that at least one is source-oriented. The existence of source-oriented generalizations has been controversial (Bybee 2001, Becker & Fainleib 2009, Kapatsinski 2011), and there is some evidence that people have a bias against learning such generalizations. This paper shows that they are learned by speakers.

Source-oriented generalizations such as the limitation of deletion to simple codas are not a problem for rule-based theories: a rule such as $\mathrm{V} \rightarrow \varnothing \ / \ \mathrm{C} \_\_ \ \mathrm{C} + \mathrm{V}$ would work. But the analysis would be ill-equipped to model product-oriented generalizations, such as the preference for TR clusters in both CC and CCC outputs, while RT, RR, and TT clusters are all equally acceptable.

The constraint-based phonotactic grammars we use are product-oriented, but they successfully capture the generalizations in the data by learning separate phonotactic grammars for subparts of the lexicon, using existing and well-understood tools. The partitioning of the lexicon is principled, as it is based on the paradigmatic behavior of the lexical items involved. When the speaker encounters a novel word, the grammar inference mechanism assigns it to the most likely sublexicon, and this inference expresses the observed source-oriented generalization.