Strict domination in probabilistic phonology

Edward Flemming (MIT) and Giorgio Magri (CNRS)

We consider a variable phonological process (say, t-deletion) whose rate of application depends on some grammatical factors $X_1, X_2, \ldots, X_n$. We denote by $\mathbb{R}(x_1, x_2, \ldots, x_n)$ the rate of application to target underlying forms that have value x_1 for factor X_1, value x_2 for factor X_2, and so on. A ranking $\gg$ of the grammatical factors $X_1, \ldots, X_n$ and a ranking $\gg_k$ of the values of each factor X_k jointly order the tuples of factor values lexicographically: $(x_1, \ldots, x_n) \succ (\widehat{x}_1, \ldots, \widehat{x}_n)$ iff there exists a factor X_k such that $x_k \gg_k \widehat{x}_k$ and $x_h = \widehat{x}_h$ for every factor X_h such that $X_h \gg X_k$ (if any). The **Strict Domination Generalization** (SDG) says that for any phonological process and any speaker of any dialect in any register, there exist rankings $\gg, \gg_k$ such that *the rates of application of the process match the corresponding lexicographic order* $\succ$: if $(x_1, \ldots, x_n) \succ (\widehat{x}_1, \ldots, \widehat{x}_n)$, then $\mathbb{R}(x_1, \ldots, x_n) \geq \mathbb{R}(\widehat{x}_1, \ldots, \widehat{x}_n)$. Thus, "each [factor] in the hierarchy [specified by the ranking $\gg$] outweighs the effects of all [factors] below it" (Labov 1969). This SDG was indeed introduced by Labov and played a prominent role in the early variationist literature (Kay & McDaniel 1979), but has faded away. This abstract brings this intriguing but 'lost' generalization back to the fore of the discussion: we provide new empirical evidence and start to discuss its theoretical implications.

Our first example is taken from Fasold (1978). He looks at the process that variably deletes a word final voiced alveolar stop after a vowel, namely /d/ → ∅/V__##. He focuses on three grammatical factors that affect the rate of d-deletion, all with values *yes*/*no*: *F(ollowing)V(owel)* has value *yes* when the word final /d/ is followed by a vowel; *ST(ress)* has value *yes* when the syllable that hosts the word final /d/ is stressed; *P(ast)M(orpheme)* has value *yes* when the word final /d/ realizes past tense. We rank these three factors as *FV* $\gg$ *ST* $\gg$ *PM* and their values as *no* $\gg$ *yes*. Fig. 1 lists the tuples of factor values from top to bottom in increasing lexicographic order together with the corresponding rates of d-deletion (from Wolfram 1974). As the rates increase from top to bottom (the bottom two rates are indistinguishable), this test case complies with the SDG. We now supplement this old example with three new pieces of evidence.

lexicographic order ↓	rates ↓
(*yes*, *yes*, *yes*)	.171
(*yes*, *yes*, *no*)	.186
(*yes*, *no*, *yes*)	.263
(*yes*, *no*, *no*)	.393
(*no*, *yes*, *yes*)	.412
(*no*, *yes*, *no*)	.666
(*no*, *no*, *yes*)	.704
(*no*, *no*, *no*)	.703

Figure 1

Many schwa vowels are optional in French ([gæʁdəʁi∼gæʁdʁi] 'kindergarten'). Smith & Pater (2020) investigate three grammatical factors that affect the rate of realization of schwa at the end of a word, all with values *yes*/*no*: *CCC* has value *yes* when omission of schwa yields a cluster of three consonants; *CLI(tic)* has value *yes* when the potential word-final schwa belongs to a clitic; *CLA(sh)* has value *yes* when omission of schwa yields a clash between stresses of adjacent words. We rank these factors as *CCC* $\gg$ *CLI* $\gg$ *CLA* and their values as *yes* $\gg$ *no*. Fig. 2 lists the tuples of factor values from top to bottom in increasing lexicographic order together with the corresponding rates of schwa realization from Smith & Pater. As also the rates increase from top to bottom, this test case complies with the SDG.

lexicographic order ↓	rates ↓
(*no*, *no*, *no*)	.09
(*no*, *no*, *yes*)	.12
(*no*, *yes*, *no*)	.56
(*no*, *yes*, *yes*)	.65
(*yes*, *no*, *no*)	.68
(*yes*, *no*, *yes*)	.83
(*yes*, *yes*, *no*)	.91
(*yes*, *yes*, *yes*)	.94

Figure 2

Storme (2021) investigates three grammatical factors that affect the rate of realization of schwa at the boundary between roots and suffixes: *CLU(ster)* has value *lcl* when omission of schwa yields a liquid+C+liquid cluster ([ˌgæʁd(ə)-ˈʁi]) and value *olc* when it yields an obstruent+liquid+C cluster ([ˌʁegl(ə)-ˈʁa] 'adjust-FUT.3SG'); *M(orphology)* encodes the morphological status of the suffix as *infl(ection)* ([ˌgæʁd(ə)-ˈʁa] 'keep-FUT.3SG') and *deriv(ational)* ([ˌgæʁd(ə)-ˈʁi]); *CLA* is as above. We rank these factors as *CLU* $\gg$ *M* $\gg$ *CLA* and their values as *olc* $\gg$ *lcl*, *deriv* $\gg$ *infl*, and *yes* $\gg$ *no*. Fig. 3 lists the tuples from top to bottom in increasing lexicographic order together with the rates of schwa realization from Storme. Again, we observe that the rates increase from top to bottom. The only glitch is that the rate for (*olc*, *infl*, *yes*) is larger than the rate for (*olc*, *deriv*, *no*). Since this reversal is not statistically significant, we conclude this test case complies with the SDG.

lexicographic order ↓	rates ↓
(*lcl*, *infl*, *no*)	.50
(*lcl*, *infl*, *yes*)	.57
(*lcl*, *deriv*, *no*)	.64
(*lcl*, *deriv*, *yes*)	.76
(*olc*, *infl*, *no*)	.83
(*olc*, *infl*, *yes*)	.90
(*olc*, *deriv*, *no*)	.86
(*olc*, *deriv*, *yes*)	.89

Figure 3

We turn to the classical t-deletion process /d, t/ → ∅/C__##. We extract rates from Coetzee and Kawahara (2003; CK). We focus on two factors (the difference between mono-morphemic versus past tense forms is known to be another relevant factor, but CK's data exclude past tense forms). One factor is the *F(ollowing)C(ontext)*, for which CK distinguish three values: *consonant*, *vowel*, and *pause* (they exclude words followed by a liquid, due to low rates of deletion in this context). The other factor is the manner of the *P(receding)C(onsonant)*, for which we distinguish three values: *sibilant*, *stop*, and *nasal* (we exclude words where the preceding consonant is a non-sibilant fricative due to lack of data). We rank these factors as *FC* $\gg$ *PC* and their values as *sibilant* $\gg$ *stop* $\gg$ *nasal* and *cons* $\gg$ *pause* $\gg$ *vowel* (in accordance with a large literature). Fig. 4 lists the tuples from top to bottom in increasing lexicographic order. The central column lists the rates of t-deletion we have computed from CK's data (we only count words with final /t/, because /d/ cannot occur in the full range of preceding contexts). Unfortunately, the match between lexicographic order and rates is tenuous (other lexicographic orders do worse).

lexicographic order ↓		fitted rates ↓
(*vow*, *nas*)	.73	.73
(*vow*, *stop*)	.10	.16
(*vow*, *sibi*)	.32	.24
(*pau*, *nas*)	.16	.14
(*pau*, *stop*)	.29	.41
(*pau*, *sibi*)	.56	.36
(*con*, *nas*)	.46	.47
(*con*, *stop*)	.50	.57
(*con*, *sibi*)	.70	.59

Figure 4

To make sense of this apparent counter-example, we bring word frequency into the picture. We constructed a logistic regression model of deletion rates with *FC*, *PC*, log of word frequency, and interactions between these factors as predictors. We used this model to calculate the deletion rates when frequency is held constant at the average frequency in the corpus. These fitted rates listed on the rightmost column of Fig. 4 better match the lexicographic order. Yet, two contexts remain problematic: (*vow*, *nas*), where the rate is too high; and (*pau*, *nas*), where the rate is too low. To explain them, we recall that the nasal in word-final /nt/ clusters is optionally deleted in English (with nasalization of the preceding vowel): [wãt] 'want' (Cohn 1993). If nasal deletion applies, the final /t/ is no longer preceded by a consonant and thus is not eligible for t-deletion. This likely results in a lower than expected rate of t-deletion in the context (*pau*, *nas*) between nasal and pause. Furthermore, nasal deletion creates the environment for optional flapping of the /t/ if it is followed by a vowel: [wãɾ̃ə] 'want a' (Jensen 1993). This likely results in a larger than expected rate of t-deletion in the context (*vow*, *nas*) between a nasal and a vowel, because the realization of /nt/ as a nasalized flap is coded by CK as an instance of t-deletion (although the flap is best analyzed as the correspondent of /t/). Once these two contexts are set aside, the only glitch is that the fitted rate for (*pau*, *sibi*) is slightly lower than the fitted rate for (*pau*, *stop*). As it is not clear that this difference is significant, we conclude that this test case is compatible with the SDG.

To understand how SOT/NHG/ME cope with the SDG, we consider a minimal system with only three factors X_1, X_2, X_3 that take only two values *yes* and *no*. We assume that *yes* promotes while *no* inhibits the application of the process considered. The SDG is then equivalent to the entailment $\boxed{\text{if } \mathbb{R}(yes, no, no) > \mathbb{R}(no, yes, no) > \mathbb{R}(no, no, yes), \text{ then } \mathbb{R}(yes, no, no) > \mathbb{R}(no, yes, yes)}$ together with all the analogous entailments obtained by changing the roles of the factors. Failure of this entailment yields a probabilistic gang effect: factor X_1 promotes the process more than factors X_2 and X_3 do individually but less than they do jointly. Since ME/NHG are based on categorical HG and gang effects are possible in categorical HG, it is unsurprising that ME/NHG fail at the boxed entailment and thus flout the SDG. The case of SOT is more complicated.

To illustrate, we consider four constraints: C_1 penalizes non-application of the process when factor X_1 has value *yes*; C_2, C_3 are defined analogously, with X_2, X_3 in place of X_1; C_0 penalizes application of the process. We consider only two candidates, representing application and non-application of the process. Fig. 5a-b represent the rates $\mathbb{R}(no, yes, no)$ and $\mathbb{R}(no, yes, yes)$ predicted by SOT as a function of the ranking values θ_2 (horizontal axis) and θ_3 (vertical axis) when $\theta_0 = 5$. Because of the boxed assumption $\mathbb{R}(no, yes, no) > \mathbb{R}(no, no, yes)$, we focus on the region $\theta_2 \geq \theta_3$ below the red line. In this region, comparison of the two panels shows that these two rates are (almost: Jäger & Rosenbach 2006) identical. The boxed entailment thus (almost) holds in SOT: since $\mathbb{R}(yes, no, no) > \mathbb{R}(no, yes, no)$ and $\mathbb{R}(no, yes, no) \simeq \mathbb{R}(no, yes, yes)$, then also $\mathbb{R}(yes, no, no) > \mathbb{R}(no, yes, yes)$. Thus, SOT (almost) complies with the SDG. Yet, it complies for the wrong reason, namely because it predicts the identity $\mathbb{R}(no, yes, no) \simeq \mathbb{R}(no, yes, yes)$ for which we see no evidence in the data. We conclude that no framework for probabilistic constraint-based phonology successfully predicts the SDG.

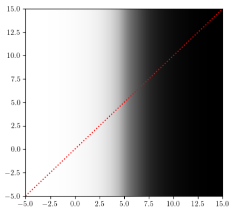

Figure 5a

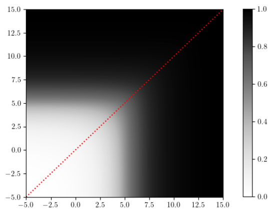

Figure 5b