Appendix E: Interaction terms and conjoined constraints

Appendix to "Deriving the Wug-shaped curve:
A criterion for assessing formal theories of linguistic variation

Bruce Hayes
UCLA

March 2021

It is standard for scholars using MaxEnt/logistic regression for purposes of statistical testing to check *interaction terms*. For example, given two choices A vs. B, C vs. D, it can turn out to be the case that the choice between C and D comes out different when A is true than when B is. Testing for the interaction of the A/B and C/D factors can inform us how likely it is this scenario is in effect.

Interaction terms are also widely proposed in Optimality Theory, under the title of "conjoined constraints" (Smolensky 1995); i.e. constraints that are composed of two preexisting constraints X and Y, and are violated only when both X and Y are violated. Deployed in an OT model in which the selection procedure is MaxEnt, these are very close to being conjoined constraints.[1] An argument that conjoined constraints are needed specifically in MaxEnt is given by Shih (2017); and I have checked the examples studied in this paper (using the xxx package in R) for whether a model with improved fit can be obtained by adding interaction terms to the analysis, and this is indeed sometimes the case. Given that interaction terms are useful, do they vitiate my claim about MaxEnt deriving wug-shaped curves?

To be sure, totally free use of local conjunction will wipe out the possibility of making strict predictions about quantitative signatures. For any of the data points in any of the cases covered in this paper, we could add a conjoined constraint that covers exactly that data point, shifting the prediction made by the model to any value we please; hence at some level the framework ends up saying nothing. But this seems like a weak, purely formal, objection. Any model that provided a constraint for every data point would be considered to be gross *overfit*; nothing is explained when there are as many points to be fit parameters to fit them.

Second, wug-shapedness is threatened only when the conjoined contains elements separate derive from the xxx and Perturber families; but this would likely create an implausible, perhaps

[1] However, the OT version of conjunction is usually *local* conjunction: *A & B is violated only when the violations of A and B are in the same location in the string, in some formally defined sense.

uninteresting partition of the constraints. Indeed, typically when we use conjoined constraints, the theoretical framework provides a *substantive reason* why the presence of (say) C makes A better relative to B harder.

To give an example, in Japanese and other languages, there is a tendency to avoid voiced obstruent geminates (double consonants). Is this because the goal is to avoid two bad things at once? Many languages do indeed avoid voiced obstruents, and many languages avoid geminates, so the interaction-term approach has some appeal. Yet closer scrutiny of the phonetics of voicing difficulty, as in Westbury and Keating (1986), gives a *substantive* reason, based on airflows and vocal tract properties, to think that voiced obstruent geminates pose a special extreme of difficulty in maintaining voicing, which has been documented experimentally by Kawahara (2006). In other words, we would have good phonetic justification for setting up the "conjoined" constraint *[+voice, +obstruent, +long] even in the absence of any information that [+voice, +obstruent] and [+long] are themselves avoided to a lesser extent. *[+voice, +obstruent, +long] is plausible as a *simplex* constraint with richer internal specification, and this fits the facts of phonetics more directly than an appeal to constraint conjunction would.[2]

Lastly, there is some evidence (e.g. xxx) that language learners are pre-equipped with a bias against complexity, and fail to detect patterns in the language data that required unmotivated conjoined constraints to analyze; an example is given in Hayes and Londe (2006, Fig. 5). It seems possible that non-wug-shaped data patterns could be a reality of the language environment, but not of the grammar that the learner shapes to those data.

[2] For a much earlier insistence on substantive support when positing interaction terms, see Sankoff and Labov (1979:205).