

Measuring the variational homogeneity of English as a world language: probabilistic indigenization effects in four syntactic alternations

(Tamaredo, Szmrecsanyi, Grafmiller, Heller & Röthlisberger)

Szmrecsanyi et al. (2016:133) define probabilistic indigenization as the process “whereby stochastic patterns of internal linguistic variation are reshaped by shifting usage frequencies in speakers of post-colonial varieties”. Put differently, probabilistic indigenization is about the extent to which probabilistic constraints come to shape variation patterns in different ways. The outcome of such indigenization will manifest in more heterogeneity, and a concomitant loss of homogeneity in the factors governing syntactic variation within different national varieties of English. The present study extends our previous knowledge of probabilistic grammars’ homogeneity (see, for instance Bernaisch et al. 2014) by sketching a corpus-based variationist method for calculating the similarity between varieties: what counts is not if and/or how often people use particular constructions, but how – that is, subject to which probabilistic constraints – they choose between “alternate ways of saying ‘the same’ thing” (Labov 1972:188). As a case study, we discuss similarity patterns between three varieties of English around the world (British English, Indian English, and Singapore English), fueled by a variationist analysis of four grammatical alternations:

- (1) The dative alternation (see, e.g., Bresnan & Hay 2008) ($N=3,012$)
 - a. *I’d given Heidi my T-Shirt* (the ditransitive dative variant)
 - b. *And I’d given the key to Helen* (the prepositional dative variant)
- (2) The genitive alternation (see, e.g., Rosenbach 2014) ($N=3,108$)
 - a. *the country’s economic crisis* (the *s*-genitive)
 - b. *the economic growth of the country* (the *of*-genitive)
- (3) The particle placement alternation (see, e.g., Gries 2003) ($N=2,480$)
 - a. *you can just cut the tops off* (verb-object-particle order)
 - b. *cut off the flowers* (verb-particle-object order)
- (4) The subject omission alternation (see, e.g., Torres Cacoullos and Travis 2014) ($N=1,225$)
 - a. *I will call again this week* (overt subject)
 - b. \emptyset *will call again this week* (null subject)

Based on materials from the International Corpus of English, we ascertain the degree of homogeneity versus heterogeneity of syntactic variables using the following procedure: relevant observations of the (a) and (b) variants above were annotated for approximately 10 probabilistic constraints including e.g. the principle of end weight (longer constituents tend to follow shorter constituents; see e.g. Wasow & Arnold 2003) and animacy effects (animate constituents tend to occur early; see e.g. Rosenbach 2008), and then submitted to variationist modeling. To evaluate the similarity between region-specific variation patterns, the method draws inspiration from the comparative sociolinguistics literature (e.g. Tagliamonte 2001): are the same constraints significant across varieties? Do the constraints have similar effect sizes? Is the overall ranking of constraints similar? Regression and conditional random forest analysis indicate that of the four alternations under study, the dative alternation is the most homogeneous one, followed – in increasing order of heterogeneity – by the genitive

alternation, the particle placement alternation, and the subject omission alternation. Against this backdrop we will evaluate claims in the literature (Szmrecsanyi et al. 2016:133) according to which the extent of probabilistic indigenization is proportional to the lexical specificity of the alternation under study.

References

- Bernaish, Tobias, Stefan Th. Gries & Joybrato Mukherjee. 2014. The dative alternation in South Asian Englishes: Modelling predictors and predicting prototypes. *English World-Wide* 35(1). 7–31. doi:10.1075/eww.35.1.02ber
- Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua* 118(2). 245–259. doi:10.1016/j.lingua.2007.02.007.
- Gries, Stefan Th. 2003. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. New York: Continuum Press.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.
- Rosenbach, Anette. 2008. Animacy and grammatical variation: Findings from English genitive alternation. *Lingua* 118. 151–171.
- Rosenbach, Anette. 2014. English genitive variation – the state of the art. *English Language and Linguistics* 18(02). 215–262. doi:10.1017/S1360674314000021.
- Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109–137. doi:10.1075/eww.37.2.01szm.
- Tagliamonte, Sali. 2001. Comparative sociolinguistics. In Jack Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *Handbook of Language Variation and Change*, 729–763. Malden and Oxford: Blackwell.
- Torres Cacoullos, Rena & Catherine E. Travis. 2014. Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics* 63. 19–34. doi:10.1016/j.pragma.2013.08.003
- Wasow, Thomas & Jennifer Arnold. 2003. Post-verbal constituent ordering in English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of Grammatical Variation in English*, 119–154. Amsterdam: Mouton de Gruyter.