Modeling language data and evaluating linguistic analyses with mathematical methods: Implications for construction grammar

Stela Manova
University of Vienna, stela.manova@univie.ac.at

Keywords: construction grammar, affix ordering, complexity measuring, mathematical modeling, language processing

In science, a problem often allows for different solutions. The so-called Big O notation serves for assessment of the complexity of those solutions in mathematics and computer science. The Big O notation tells us how an algorithm slows as data grow. Thus, complexity is not a property of data (as is the case in linguistics) but of analysis. I explain how such an understanding of complexity can be applied to linguistic analyses and evaluate the complexity of a recent approach to affix order, the so-called Complexity-Based Ordering (CBO), Plag (2002), Hay & Plag (2004), Plag & Baayen (2009). CBO elaborates on the Parsability Hypothesis (Hay 2001, 2003) and relies on data from corpora (Hay & Baayen 2002). The approach was originally formulated for English, enjoyed popularity but fails to account for the order of affixes even in languages closely genealogically related to English (Manova 2010, 2015, Zirkel-Hinkelbach 2011, Talamo 2015). I therefore introduce an alternative approach that is based on a mathematical method, Gauss-Jordan elimination, and is simpler than CBO.

Gauss-Jordan elimination serves for solving systems of linear equations numerically ((1) is an example of a system of linear equations), that is, only with the help of elementary operations such as substitution, addition or multiplication.

(1)
$$2x + y + 2z = 10$$
$$x + 2y + z = 8$$
$$3x + y - z = 2$$

The goal of Gauss-Jordan is, based only on well-known facts and elementary operations with them, to come to a single option for a variable (the unknown); x, y and z are the variables in (1). If there is only one option for a variable, this option is the variable's value, i.e. the solution to the problem.

With respect to affix order, the well-known information is information about the lexical category specification of an affix, i.e. whether the affix derives nouns (N), adjectives (A) or verbs (V); a single option for a variable means one affix combination of a kind. Thus, I model derivational suffix combinations in terms of bigrams of the type SUFF1–SUFF2, where SUFF1 has three valency positions for further suffixation: $SUFF2_N$, $SUFF2_A$ and $SUFF2_V$. As can be seen from Table 1, this assumption allows data to be distributed so that in most cases there is one option of a kind, see for N (*-ist*$_N$*-dom*$_N$) and for V (*-ist*$_N$*-ize*$_V$).

Table 1: Combinability of the English suffix -ist (data from Aronoff & Fuhrhop 2002, based on OED, CD 1994)

SUFF1	Lexical category of SUFF1	SUFF2 according to lexical category
-ist	N	N: *-dom* (2) A: *-ic* (631), *-y* (5) V: *-ize* (3)

If more than one SUFF2 of the same lexical category is available (see for A in Table 1), one of the SUFF2 suffixes attaches by default, suffix *-ic* in our case: in English, the combination *-ist*$_N$*-ic*$_A$ derives 631 types, while *-ist*$_N$*-y*$_A$ derives only 5 types. Regarding default suffixes, having counted suffix combinations in large dictionaries and corpora for different languages, Manova (2011, 2015), Manova & Talamo (2015) and Manova & Knell (2021) maintain that a default suffix derives more than ten types, while SUFF2 suffixes that compete with the default suffix derive ten or fewer types each.

Finally, I report the results of a psycholinguistic experiment with native and advanced non-native speakers of English and German, and with native speakers of Italian, Spanish, Polish and Slovene. The participants in the experiment did not need semantic cues to process suffix combinability, i.e. they could differentiate between existing and non-existing suffix combinations presented to them without lexical bases such as roots/stems/words, and the participants were better in recognizing productive than unproductive combinations (Manova & Knell 2021 for English).

All these, taken together, have consequences for our understanding of what a (morphological) construction is (Croft 2001, Goldberg 2006, Booij 2010, Jackendoff & Audring 2020, Hoffmann 2022) as well as of how constructions are stored and accessed in the mental lexicon.

References

Aronoff, Mark & Nanna Fuhrhop. 2002. Restricting Suffix Combinations in German and English: Closing Suffixes and the Monosuffix Constraint. *Natural Language and Linguistic Theory* 20(3). 451-490.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Jackendoff, Ray & Jenny Audring. 2020. *The Texture of the Lexicon: Relational Morphology and the Parallel Architecture*. New York: Oxford University Press.

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39(6). 1041-1070.

Hay, Jennifer. 2003. *Causes and Consequences of Word Structure*. London: Routledge.

Hay, Jennifer & Harald Baayen. 2002. Parsing and Productivity. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2001,* 203-255. Dordrecht: Kluwer.

Hay, Jennier & Ingo Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22(3). 565-596.

Hoffmann, Thomas. 2022. *Construction Grammar: The Structure of English*. Cambridge: Cambridge University Press.

Manova, Stela. 2010. Suffix combinations in Bulgarian: Parsability and hierarchy-based ordering. *Morphology* 20(1). 267–296.

Manova, Stela. 2011. A cognitive approach to SUFF1-SUFF2 combinations: A tribute to Carl Friedrich Gauss. *Word Structure* 4(2). 272–300.

Manova, Stela. 2015. Affix order and the structure of the Slavic word. In S. Manova (ed.), *Affix Ordering Across Languages and Frameworks*, 205-230. New York: Oxford University Press.

Manova, Stela & Luigi Talamo. 2015. On the significance of the corpus size in affix-order research. *SKASE Journal of Theoretical Linguistics* 12(3). 369-397.

Manova, Stela & Georgia Knell. 2021. Two-suffix combinations in native and non-native English: Novel evidence for morphomic structures. In Sedigheh Moradi, Marcia Haag, Janie Rees-Miller & Andrija Petrovic (eds.), *All Things Morphology: Its Independence and Its Interfaces,* 305-323. Amsterdam: Benjamins.

Plag, Ingo. 2002. The role of selectional restrictions, phonotactics and parsing in constraining suffix ordering in English. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2001*, 285-314. Dordrecht: Kluwer.

Plag, Ingo & Harals Baayen. 2009. Suffix ordering and morphological processing. *Language* 85(1). 109-152.

Talamo, Luigi. 2015. Suffix combinations in Italian: Selectional restrictions and processing constraints. In Stela Manova (ed.), *Affix Ordering Across Languages and Frameworks*, 175–204. New York: Oxford University Press.

Zirkel-Hinkelbach, Linda. 2011. Can Complexity-Based Ordering be extended from English to German?. Talk at The 3rd Vienna Workshop on Affix Order: Advances in Affix Order Research, University of Vienna, 15-16 January 2011.