

QUANTIFYING VERBAL BEHAVIOR IN SEARCH OF GRAMMATICAL CREATIVITY

Ittamar Erb¹, Shahar Spencer², Nurit Melnik¹ & Gabriel Stanovsky²

¹The Open University of Israel & ²The Hebrew University

UniDive 1st general meeting, Paris-Saclay University, 16-17 March 2023

Overview

Our project explores how NLP tools and methods can be harnessed to advance goals of theoretical linguistics, namely to uncover the mechanisms underlying the expressive power of language. We employ NLP tools for automatic lemmatization, dependency parsing and POS tagging of a dataset consisting of blogposts. We search for grammatical creativity at various linguistic levels, hypothesizing that it occurs when speakers diverge from linguistic conventions. Our goal is to develop methodologies that are largely language independent, yet as a proof of concept, we begin by exploring creative dimensions in English, where models and data are plentiful.

The current focus is on two linguistic dimensions: a *morpho-syntactic dimension* and an *argument structure dimension*. We assess the flexibility of lemmas by calculating the entropy of their distribution across categorical variables. Our working hypothesis is that creative utterances can be identified by searching for non-conventional occurrences of rigid lemmas. Preliminary findings suggest that the main challenge to improve precision is teasing apart creativity from rarity, as well as from human typos and from computational tagging errors.

Dimensions of Creativity

We focus here on two linguistic dimensions at the center of the debate: a *morpho-syntactic* dimension which targets the distribution of lemmas across different POSs, and an *argument structure* dimension which concerns the range of argument structures in which verbs appear.

Creativity in the morpho-syntactic dimension

- (1) So I had to **bobby pin** the bangs away for the longest time...
- (2) ...we were planning on doing a trip [...] but in the morning it was **bucketing down**, so that plan did not happen, no point getting needlessly wet...

Creativity in the argument structure dimension

- (3) I went to J-crew to try and **think things off**, and got a nice tie there (although it wasn't the one I wanted) and my short term memory suddenly kicked in.
- (4) no more professional bass trombone sounds will echo and **vibrate the hall**.

Quantifying Verbal Behavior

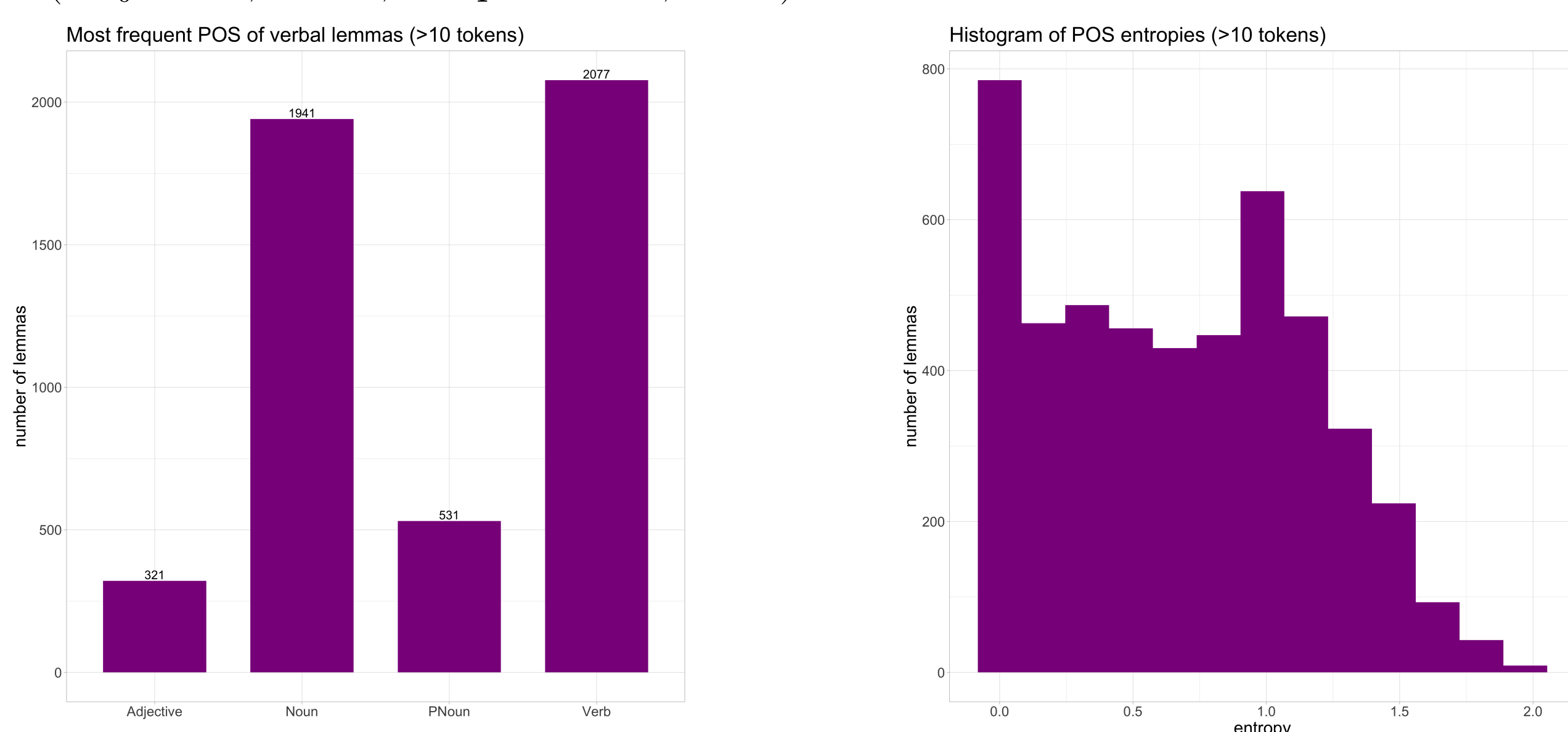
For each dimension, we rate verbal lemmas for their *rigidity* and *flexibility*.

Assuming that unexpectedness and surprise are distinctive features of creativity [2], we hypothesize that linguistic expressions with unconventional uses of rigid lemmas (in either dimension) are likely to be deemed creative.

We use SpaCy (<https://spacy.io>) to analyze the Blog Authorship Corpus [3], an open source dataset consisting of 681,288 posts and over 140 million words, parsed and tagged using a variant of Universal Dependencies [1].

Morpho-syntactic Dimension | Results

We compute a co-occurrence matrix of all verbal lemmas in the corpus and open-class POSs (Adjective, Noun, Proper Noun, Verb).



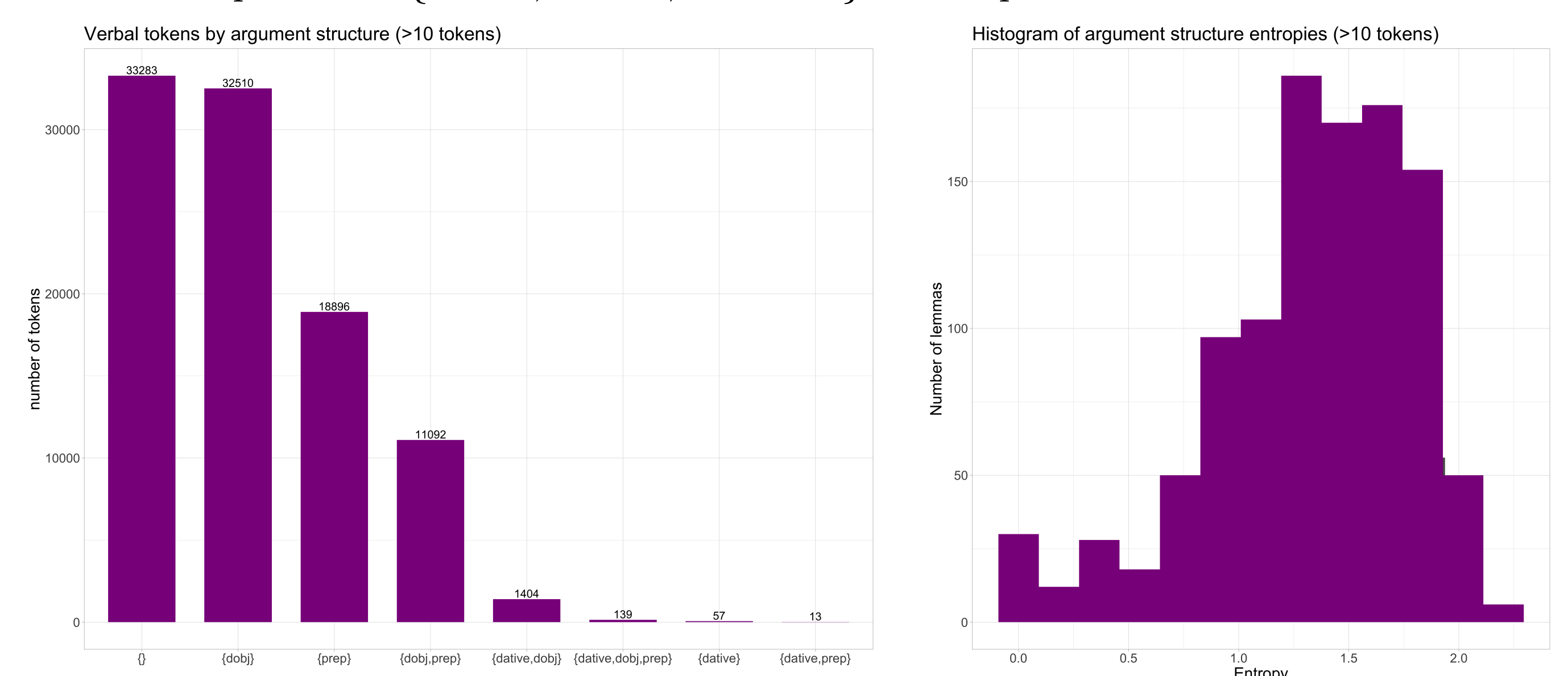
- We targeted verbal lemmas canonically associated with other POSs.
- For each of the 20 lemmas with the lowest entropy, we randomly retrieved a corpus sentence in which the lemma occurred as a verb.
- 7 were found to be potentially creative. Others were infrequent but conventional uses, typos (9), and tagging errors (10).

Morpho-syntactic creativity

- (5) ...it's hip and with it and whatever those damn kids use these days, to **lower case** your type.
- (6) Now imagine me, **shoeing** a roach the size of a full grown cat out of the apartment...

Argument Structure Dimension | Results

We compute a co-occurrence matrix of all verbal lemmas and sets of non-subject non-clausal complements {DOBJ, PREP, DATIVE} in complete clauses.



- We targeted verbal lemmas which alternate between at least two sets of complements.
- For each of the 20 lemmas with the lowest entropy score, we randomly retrieved a single corpus sentence in which the verb appears in the least frequent argument structure.
- 3 were found to be potentially creative. Others were mostly tagging errors, some of which misanalyzed short forms (11).

Argument-structure creativity

- (7) tHey aCtually **aRe tO** me, dEy hAve bEcOme nUtHiNg bUt
- (8) ...God's been using to teach me how to fight my feelings and **depend** Him for my strength...

Challenges

Typos

- (9) I have started making the large scroll duffel-bag that I was **planing** for a while.
planing (intended: *planning*) recognized as a denominal verb of *plane*.

Tagging errors

- (10) Wouldn't you know it the llama caught up with JAKE and **bit** him right in the hinder.
bit was analyzed as a word form of the lemma *BIT* and not *BITE*.

Short forms

- (11) And she was like, 'You know, **l-m-a-o...**
m- was tagged as VERB and *a-* its DOBJ.

Morphology

- (12) I didn't drive 35 min to class just [...] to get bended into positions that noone in their right mind wants to be **pretzelized** into...
Derivational morphology (suffix *-ized*).

Selected References:

- [1] Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of LREC'16*, pages 1659–1666.
- [2] Mark A Runco and Garrett J Jaeger. 2012. The standard definition of creativity. *Creativity Research Journal*, 24(1):92–96.
- [3] Jonathan Schler, Moshe Koppel, Shlomo Argamon, and James W Pennebaker. 2006. Effects of age and gender on blogging. In *AAAI spring symposium: Computational approaches to analyzing weblogs*, volume 6, pages 199–205.

This research was supported by a grant from the Israeli Ministry of Science and Technology (grant No. 0002336).