Creativity, productivity and diversity: The case of Hebrew possessive constructions

Ittamar Erb
The Open University of Israel
ittamar.erb@gmail.com

Nurit Melnik
The Open University of Israel
nuritme@openu.ac.il

Relevant UniDive working groups: WG4

1 Introduction

The notion of *creativity* is often invoked in the linguistics literature. Almost every introductory textbook emphasizes the human ability to produce and understand sentences that have never been made or heard before. What enables this creativity, according to generative approaches in linguistics, is the grammar – a system of rules which generate all and only the grammatical sentences of a language. This, it is assumed, is one property that makes human language unique and sets it apart from non-human communication systems.

Recently, however, Sampson (2016) questioned the appropriateness of associating rule-based generation with creativity. The generativists' notion of creativity, he argues, is comparable to ascribing creativity to the generation of a mathematical equation, for example 4792 * 5306 = 25426352. Although this particular well-formed equation may have never been carried out before, would its generation be considered creative?

However, as speakers, we recognize that not all linguistic expressions are created equal. Consider, for example, the *way*-construction, which can host a wide range of verbs (Perek, 2018). Nevertheless, we find (1) to be more creative than an alternative expression where *Tindersurfed* is replaced with *drove*, while replacing it with *believed* would render it rather unintelligible.

(1) Meet the man who Tindersurfed his way around Europe... (Hoffmann, 2020, ex.3)

The goal of our project is to develop a model which captures speakers' intuition regarding constructions and their extensibility. To achieve this we aim to (i) characterize constructions with regards to their usage, as attested in linguistic corpora, and (ii) to predict speakers' evaluation of newly coined instantiations of these constructions. As a test case we compare two variant pronominal possessive constructions in Hebrew.

2 Constructions and their extensibility

2.1 Productivity and diversity

Linguistic productivity refers to the extensibility of a schema, or, in other words, the range of lexical items that can instantiate it. In morphology, where it was first defined, it involves the application of morphological rules (e.g., the nouns to which the suffix *-hood* can be added). In Construction Grammar the notion is extended from morphology to syntax, where it relates to the extent to which constructions are selective (or permissive) with regards to the lexical items that may fill their slots (Goldberg, 1995; Perek, 2016, 2018).

Different corpus-based measures have been proposed in the literature for assessing the productivity of morphological and syntactic schemas. We find that they are tightly connected to the three dimensions that are often employed in various domains for modelling and measuring *diversity*: variety, balance and disparity (Stirling, 1998; Morales et al., 2021; Lion-Bouton et al., 2022).

Variety The number of types into which items can be classified is one measure of diversity. For morphological rules, Bybee (2001) proposes that a clear indicator of their productivity is their *realized productivity*, that is the type count of their category members in a corpus. This measure is also applied to constructions and their class of attested slot fillers (e.g., Perek, 2016)

Balance A second dimension targets the uniformity of the type–item distribution. In the linguistic literature, a higher Type Token Ratio (TTR) is a sign of a high degree of lexical richness. An additional linguistic balance-based measure is *potential productivity*, which is the proportion of hapax legomena among the total number of tokens of the pattern in the corpus (Baayen, 2009).

Disparity The third dimension relates to the degree to which types differ from each other within a category. Goldberg (1995), in her discussion of the productivity of constructions, envisions a semantic space in which verbal lexemes are projected

onto two dimensions, and semantically close lexemes cluster together. This conceptualization is operationalized by *semantic vector-space models* and *distributional semantic models* which take into account the relations between attested types (e.g., Perek, 2016, 2018; Watson et al., 2021).

2.2 Evaluating coinages

Constructions exhibit varying degrees of productivity, yet even the most productive constructions impose some restrictions on the lexical items that can instantiate them. This is what Goldberg (1995) terms *partial productivity*.

Furthermore, as mentioned above, when faced with a novel utterance, speakers' judgements vary; coinages may be perceived as ordinary, surprising, ungrammatical, unintelligible, among other reactions. Goldberg (1995, p.133) suggests that "[n]ew or previously unclassified verb forms are attracted to existing clusters on the basis of similarity to existing cases". Yet Watson et al. (2021) show that similarity is not a sufficient predictor and that the acceptability of creative extensions of constructions depends on the interaction of category diversity, token frequency and item similarity.

3 Hebrew possessive constructions

Two competing Hebrew possessive constructions serve as test cases in our project. We explore the category structure of each construction individually, and use their (partial) synonymy to test and compare each construction's extensibility with regards to various coinages.

3.1 Preliminary findings

Hebrew offers two variant pronominal possessive constructions: a prepositional construction, in which the possessive suffix is attached to the preposition *ʃel* 'of' (2), and a suffixed construction in which a pronominal suffix is attached to the possessed nominal (3).

(2) axot ʃel-i
sister of-my
'my sister'

(3) axot-i
sister-my
'my sister'

An analysis of the two constructions based on small-scale spoken language corpora,[1] revealed considerable differences between the two (Erb, 2022). Based on the formal type frequency (Baayen's realized productivity) and TTR measures, the prepositional construction was found to be much more frequent and more productive.

	Prepositional	**Suffixed**
Token frequency	653	172
Type frequency	286	55
Type–Token Ratio	0.437	0.319

Table 1: Productivity measures

A functional analysis examined the range of functions which the constructions serve, with the prototypical function being definite reference to inalienable entities, most notably kinship terms and body parts. The suffixed construction was found to be restricted to prototypical instances. Conversely, the prepositional construction was vastly extended to non-prototypical ones, thus exhibiting a larger degree of productivity (Erb, 2022).

	Prepositional	**Suffixed**
Inalienable	307/653 (47%)	144/172 (83.7%)
Definite	420/653 (64.3%)	172/172 (100%)

Table 2: Prototypical functions

3.2 The next steps

The preliminary findings presented above constituted the first step in our investigation. We are currently working on the next step, which is to further enhance our analysis of the productivity of the two constructions by applying additional measures, most notably relating to disparity. Our goal is to develop a comprehensive, theoretically motivated, quantitative/geometrical description of the semantic space occupied by each construction. To achieve this, we are using AlphaBERT (Seker et al., 2021) to obtain the semantic vectors and we are experimenting with different disparity metrics to characterize the semantic space.

Subsequently, we aim to employ the models to predict speakers' responses to coinages, i.e. nominals not attested in either constructions in the spoken Hebrew corpora. We plan to assemble a collection of theoretically motivated measures, as a combination of different factors to determine whether a novel instance is perceived by speakers to be creative, productive or ungrammatical, and to what degree.

[1] The Corpus of Spoken Israeli Hebrew (Izre'el et al., 2002) & The Haifa Corpus of Spoken Hebrew (Maschler et al., 2021)

Acknowledgements

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

References

R. Harald Baayen. 2009. Corpus linguistics in morphology: morphological productivity. In Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics: An International Handbook*, pages 900–919. Mouton de Gruyter, Berlin/New York.

Joan Bybee. 2001. *Phonology and language use*. Cambridge University Press, Cambridge.

Ittamar Erb. 2022. From synchrony and diachrony and back: the case of hebrew pronominal possessives. MA thesis, Tel-Aviv University.

Adele Goldberg. 1995. *Constructions. A Construction Grammar approach to argument structure*. University of Chicago Press, Chicago.

Thomas Hoffmann. 2020. Construction grammar and creativity: Evolution, psychology, and cognitive science. *Cognitive Semiotics*, 13(1).

Shlomo Izre'el, Benjamin Hary, and Giora Rahav. 2002. Designing CoSIH: The corpus of Spoken Israeli Hebrew. *International Journal of Corpus Linguistics*, 6(2):171–197.

Adam Lion-Bouton, Yagmur Ozturk, Agata Savary, and Jean-Yves Antoine. 2022. Evaluating diversity of multiword expressions in annotated text. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3285–3295, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Yael Maschler, Hilla Polak-Yitzhaki, Stav Fishman, Carmit Miller Shapiro, Netanel Goretsky, Gallith Aghion, and Ophir Fofliger. 2021. The Haifa Corpus of Spoken Hebrew. https://sites.google.com/humanities.haifa.ac.il/corpus.

Pedro Ramaciotti Morales, Robin Lamarche-Perrin, Raphaël Fournier-S'Niehotta, Rémy Poulain, Lionel Tabourier, and Fabien Tarissan. 2021. Measuring diversity in heterogeneous information networks. *Theoretical Computer Science*, 859:80–115.

Florent Perek. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics*, 54(1):149–188.

Florent Perek. 2018. Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory*, 14(1):65–97.

Geoffrey Sampson. 2016. Two ideas of creativity. In Martin Hinton, editor, *Evidence, Experiment and Argument in Linguistics and Philosophy of Language*, pages 15–26. Peter Lang, Bern.

Amit Seker, Elron Bandel, Dan Bareket, Idan Brusilovsky, Refael Shaked Greenfeld, and Reut Tsarfaty. 2021. AlephBERT: A Hebrew large pretrained language model to start-off your Hebrew NLP application with.

Andrew Stirling. 1998. On the economics and analysis of diversity. *Science Policy Research Unit (SPRU), Electronic Working Papers Series, Paper*, 28:1–156.

Julia Watson, Farhan Samir, Suzanne Stevenson, and Barend Beekhuizen. 2021. Coin it up: generalization of creative constructions in the wild. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43.