

42: Managing Semantic Norms for Cognitive Linguistics, Corpus Linguistics, and Lexicon Studies

- Class: Linguistics and Data Science
- Presenter: Yuna Park
- Date: 2023-10-30

Table of content

1. Introduction
2. Norm-less Semantics
3. Norm data set and examples of norm-based linguistics
4. Methodological issues and reproducibility in the context of norms

Note

This chapter provides an overview of the concept of semantic norm rather than introducing specific data. Therefore, please focus on the bold keywords.

Section 1. Introduction

Norms: This term is used by psycholinguists when they collect **ratings** for linguistic items, generally words.

Examples:

The emotional valence norms (Warriner, Kuperman, and Brysbaert 2013)

They asked hundreds of native English speakers to rate words for how good or bad they are.

The word vacation received a rating of 8.53 on the nine-point rating scale for these emotional valence norms, which means the word vacation overall appeared to be very positive to the raters.

Section 2. Norm-less Semantics

Some of the problems that may arise for norm-less semantics with respect to the reproducibility

Corpus linguistics: Semantic prosody

Semantic Prosody: The idea that words consistently occur in certain types of attitudinal or emotional contexts

- For examples, the verb *to cause* tends to occur in only bad things get caused from the Corpus of Contemporary American English.

The problem is that concordances are usually **hand-classified** for their emotional connotation, such as whether they are overall positive or negative.

And linguistic intuitions are by definition **not reproducible** because they are contingent on one person's beliefs.

This time-consuming hand-classifying process results in focusing on the meaning of only a few isolated headwords.

Cognitive linguistics: sensory language and perceptual metaphor

Perceptual language, including the study of perception verbs or the study of perceptual adjectives as they occur in metaphorical expressions, needs a **categorization of sensory words** according to perceptual modality, such as classifying smooth as a touch word, or classifying squealing as a sound word.

But numerous examples like thick, long and harsh are **hard to classify**. The literature in this field is ripe with examples where different researchers have classified the same words differently.

Therefore, criteria need to be made explicit in order for research on perceptual language to be reproducible.

The advantage of Norms

Norms allow **reproducible** research which means different researchers will reach the same conclusions if they use the same norm data. They are furthermore **scalable**, which affords more generalizable claims that replicate many more words. Thus, norms provide an opportunity to free from the shackles of a time-consuming classification process that is difficult to reproduce and extend to larger chunks of the English lexicon.

Section 3. Norm data set and examples of norm-based linguistics

1960 ~ 1970s:

- *The Measurement of Meaning* (Osgood, Suci, & Tannenbaum 1957) / Paivio's, which included concreteness, imageability, and meaningfulness ratings for 925 English words (Paivio, Yuille, & Madigan 1968)

These days :

Norms are often collected via crowdsourcing platforms in large **megastudies**

- Concreteness norms collected for 40,000 English words using data from over 4,000 participants. (Brysbaert, Warriner, & Kuperman 2014)
- emotional valence norms collected 14,000 English words. (Warriner, Kuperman, & Brysbaert 2013)

Some **unexpected dimensions of meanings** that have been normed.

- Roughness, hardness, and size of touch adjectives (Stadtlander & Murdoch 2000)
- The color and motion-relatedness of nominal concepts (Medler et al. 2005)
- The graspability and painrelatedness of object terms (Amsel, Urbach, & Kutas 2012)
- Bochum English Countability Lexicon that includes expert annotator's ratings for whether nouns are mass or count (Kiss et al. 2016).

A lot of norming data sets are traditionally published in the journal *Behavior Research Methods*. However, many particular data sets have **low visibility**, which also means that researchers may not know that norms are

an available methodological option for answering their research questions.

websites that allow easy access to norm data sets:

- LAB, the Linguistic Annotated Bibliography (Buchanan, Valentine, & Maxwell 2018)
- language goldmine (languagegoldmine.com)

Norm-based studies:

- *Pollyanna hypothesis* (speakers have a prosocial need to talk about ositive things more often than about negative things) tested by Warriner and Kuperman (2015) using emotional valence.
- Lupyan and Winter (2018) used concreteness norms to argue that language is much more abstract than is commonly assumed by “embodied” approaches to cognition.
- Iconicity disproportionately resides in the perceptual part of the English vocabulary (Sidhu & Pexman 2018; Winter et al. 2017) (using Iconicty norms)
- Children’s language is relatively more iconic compared to adult’s language (Perry, Perlman, & Lupyan 2015) (using Iconicty norms)
- adults increase the frequency of iconic words when talking to their children (Perry et al. 2017) (using Iconicty norms)

Section 4. Methodological issues and reproducibility in the context of norms

There are few methodological issues about norms.

1. Norm data are still subjective, as they rely on native language user judgments and **wisdom of the crowd** effect.
2. It is important to consider **construct validity** for norming studies to avoid situations of evaluating the sense of words like freedom.
3. Analyses should not be performed on **infrequent and little-known words** that participants cannot rate.
4. Norming studies should also compute **inter-rater reliability statistics** which is the degree of agreement among participants who rate the same phenomenon.
5. It is desirable to **replicate** specific analyses with different norm data sets to ensure that a particular claim rests on a firm foundation.
6. Norm based research can be **preregistered** which is researcher specifies their analysis plan in advance to allow clearly demarcating the boundary between confirmatory (hypothesis-testing) and exploratory (hypothesis-generating) research.

Quiz

Fill the blank.

- Norms allow researchers to avoid hand-classification and have the advantages that they are _____ (if using the same norm data set) and _____ (giving the opportunity to analyze many more words automatically).