

MIT Open Access Articles

Weak Quantitative Standards in Linguistics Research

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Gibson, Edward, and Evelina Fedorenko. "Weak Quantitative Standards in Linguistics Research." *Trends in Cognitive Sciences*, vol. 14, no. 6, June 2010, pp. 233–34.

As Published: <https://doi.org/10.1016/j.tics.2010.03.005>

Publisher: Elsevier

Persistent URL: <https://hdl.handle.net/1721.1/121245>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-NonCommercial-NoDerivs License



Weak quantitative standards in linguistics research

Edward Gibson (egibson@mit.edu)

Department of Brain and Cognitive Sciences, Department of Linguistics and Philosophy,
Massachusetts Institute of Technology
46-3035, MIT, Cambridge, MA 02139

Evelina Fedorenko (evelina9@mit.edu)

Department of Brain and Cognitive Sciences, McGovern Institute for Brain Research
Massachusetts Institute of Technology
46-4141C, MIT, Cambridge, MA 02139

Corresponding author: GIBSON, E.A.F. (Email: egibson@mit.edu)

A serious methodological weakness affecting much research in syntax and semantics within the field of linguistics is that the data presented as evidence are often not quantitative in nature. In particular, the prevalent method in these fields involves evaluating a single sentence/meaning pair, typically an acceptability judgment performed by just the author of the paper, possibly supplemented by an informal poll of colleagues. Although acceptability judgments are a good dependent measure of linguistic complexity (results from acceptability-judgment experiments are highly systematic across speakers and correlate with other dependent measures—but see [1]), using the researcher's own judgment on a single item/pair of items as data sources does not support effective testing

of scientific hypotheses for two critical reasons. First, as several researchers have noted [2,3,4], a difference observed between two sentences could be due to lexical properties of the materials rather than syntactic or semantic properties [5,6]. Multiple instances of the relevant construction are needed to evaluate whether an observed effect generalizes across different sets of lexical items [7]. The focus of this letter, however, is on a second problem with the standard linguistic methodology: because of cognitive biases on the part of the researcher, the judgments of the researcher and his/her colleagues cannot be trusted (see Box 1). As a consequence of these problems, multiple items and multiple *naïve* experimental participants should be evaluated in testing research questions in syntax/semantics, which therefore require the use of quantitative analysis methods.

The lack of validity of the standard linguistic methodology has led to many cases in the literature where questionable judgments have led to incorrect generalizations and unsound theorizing, especially in examples involving multiple clauses, where the judgments may be more subtle and possibly more susceptible to cognitive biases. As one example, a well-cited observation in the syntax literature is that an object-subject-verb question asking about three elements is more natural than one asking about only two (e.g., *What did who buy where?* is claimed to sound better than *What did who buy?*). Several theories explain and build on this purported phenomenon [10,11]. However, it turns out that the empirical claim is not supported by quantitative measurements [12,13]. There are many other such examples of questionable judgments leading to unsound theorizing [2,3,4; including an example from the first author's PhD thesis (Gibson, E., 1991, PhD Thesis, Carnegie Mellon University), which is discussed along with other

examples in (Gibson, E. & Fedorenko E., *The need for quantitative methods in syntax*, unpublished)]. Without quantitative data from naïve participants, cognitive biases affect *all* researchers.

Furthermore, even if such cases were rare, the fact that this methodology is not valid has the unwelcome consequence that researchers with higher methodological standards will often ignore the current theories from the field of linguistics. This has the undesired effect that researchers in closely related fields are unaware of interesting hypotheses in syntax and semantics research.

In order to address this methodological weakness, future syntax/semantics research should apply the quantitative standards from cognitive science, whenever feasible. Of course, gathering quantitative data does not guarantee the lack of confounding influences or the correct interpretation of data. However, adopting existing standards for data gathering and analyses will minimize reliance on data that are spurious, driven by cognitive biases on the part of the researchers.

Corpus-based methods provide one way to quantitatively test hypotheses about syntactic and semantic tendencies in language production. A second approach involves controlled experiments. Experimental evaluations of syntactic and semantic hypotheses should be conducted with participants who are naïve to the hypotheses and samples large enough to make use of inferential statistics. A variety of experimental materials should be used in order to rule out effects due to irrelevant properties of the experimental items (e.g.,

particular lexical items). It is our hope that strengthening methodological standards in the fields of syntax and semantics will bring these fields closer to related fields, such as cognitive science, cognitive neuroscience and computational linguistics.

Box 1: Cognitive biases and linguistic judgments (177 words)

There are at least three kinds of unconscious cognitive biases [8,9] that can adversely affect the results of intuitive judgments, given the way that they are currently typically gathered in the syntax/semantics literature:

1. Confirmation bias on the part of the researcher: Researchers will often have a bias favoring the success of the predicted result, with the consequence that they will tend to treat data that do not support the hypothesis as flawed in some way (e.g., from a not quite native speaker, or from a speaker of a different dialect).
2. Confirmation bias on the part of the participants: Individuals that the researcher asks to provide a judgment on a linguistic example – including the researcher him/herself – may be biased because they understand the hypotheses. When faced with complex materials, they may then use these hypotheses to arrive at the judgment.
3. Observer-expectancy effects (the “clever Hans” effect): Individuals that the researcher asks to provide a judgment may be biased because they subconsciously want to please the researcher and are consequently affected by the researcher’s subtle positive/negative reactions.

References

- [1] Edelman, S. & M. Christiansen. 2003. How seriously should we take Minimalist syntax? *Trends in Cognitive Sciences* 7: 60-61.
- [2] Schutze, C. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- [3] Cowart, W. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.
- [4] Wasow, T. & Arnold, J. 2005. Intuitions in linguistic argumentation. *Lingua*, 115, 1481-1496.
- [5] MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676-703.
- [6] Gibson, E. & Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, 2, 262-268.
- [7] Clark, H.H. (1973). The Language-as-Fixed-Effect Fallacy: A Critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.
- [8] Evans, J. St. B. T., Barston, J.L., & Pollard, P. 1983. On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, 11, 295-306.
- [9] Nickerson, R.S. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175-220.
- [10] Kayne, R. 1983. Connectedness. *Linguistic Inquiry*, 14, 223-249.
- [11] Pesetsky, D. 2000. *Phrasal movement and its kin*. Cambridge, MA: MIT Press.

[12] Clifton, C., Jr., Fanselow, G., & Frazier, L. 2006. Amnestying superiority violations: Processing multiple questions. *Linguistic Inquiry*, 37, 51-68.

[13] Fedorenko, E. & Gibson, E. (2010). Adding a third wh-phrase does not increase the acceptability of object-initial multiple-wh-questions. *Syntax*.