

A standard scale of well-formedness: Why linguistics needs boiling and freezing points

Sam Featherston
Tübingen University
Sam.Featherston@uni-tuebingen.de

The debatable quality of introspective judgements of well-formedness as linguistic data has been a controversial issue since at least the sixties (Schütze 1996 for excellent review). In the follow-up of work by Cowart (1997) and Bard et al (1996), who introduced the method and practice of magnitude estimation of introspective judgements to linguistics, many of these problems have been reduced and a degree of consensus achieved. The use of experimental methods has meant that not just categorical judgements (*Is this good or bad?*) but also relative judgements (*How much better than the other one is this structure?*) can now be gathered with some degree of confidence and reliability (e.g. Keller 2000).

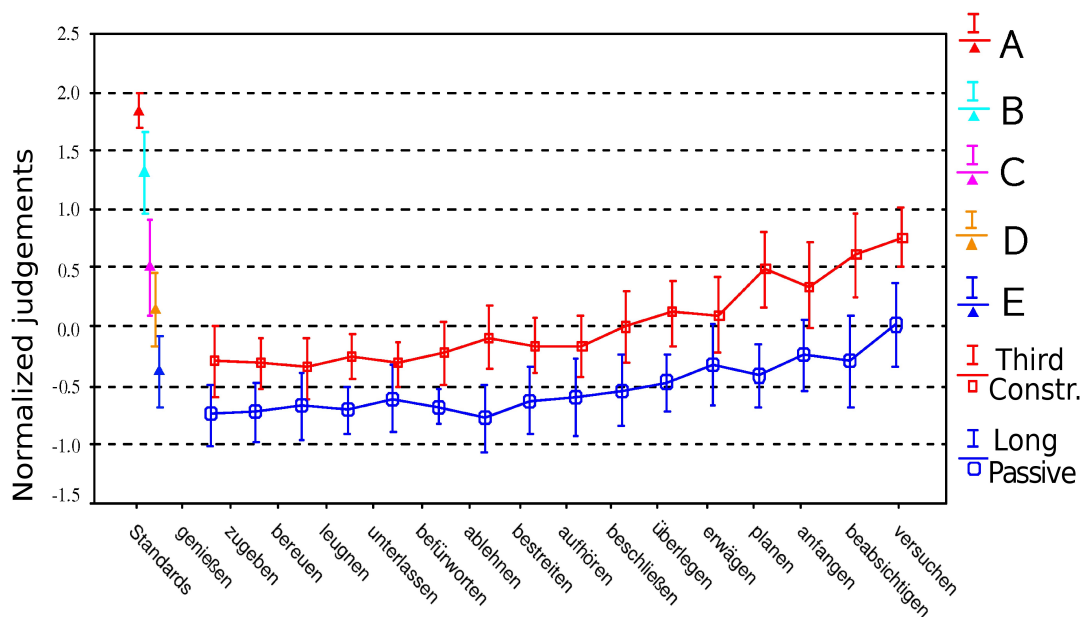
This relative measure of well-formedness does not solve all problems, however, and some questions even grow in significance as the data base of judgements becomes increasingly quantitative. While relative judgements give finer data, the traditional reference points are lost. Syntacticians often want to know where the patterns of differences obtained from studies using relative judgements lie in comparison to known examples, or relative to the traditional standards of categorical grammaticality.

This issue is related to a wider problem of the communication of intuitive judgements, due to the essentially subjective nature of the data type: how do we get one person's introspective judgements into the head of another person, when we have no way of knowing how one person's intuitions relate to those of another person? The difficulty applies to comparison of judgements between studies too: if we gather relative judgements, how do the results of one study relate to the results of a different

study? Additionally, grammatical theory has difficulty in making full use of relative data in theory construction.

In this paper we shall report our own attempt to resolve this problem by developing a set of *cardinal well-formedness values* instantiated in standard comparison items, in relation to which all other judgements may be more absolutely located. The effect is to produce a known scale, which permits speakers to locate their impressions more exactly relative to known values and thus aids the communication of values. The same effect is familiar from other fields: for example, the Celsius scale of temperature. If we hear that tomorrow will be 27° degrees, this gives us more information than if we are only told that it will be 'quite warm', in part because we can compare it with other days we have experienced with known temperature values. The anchor values of the Celsius scale - boiling and freezing - are essentially random, but perform the useful function of providing familiar known reference points. Our aim was to develop a similar scale of perceived well-formedness.

In order to select the items we first tested a wide range of example sentences across the full range of perceived well-formedness, using the magnitude estimation methodology. We then divided the range of judgements we obtained into five bands, which we labelled A – E. From the centre of each of these bands we selected items which were fairly consistently graded as belonging to the level. Examples at level A were perfect and examples at level E were very ill-formed, but not so much so as to be



incomprehensible, since the aim is to provide a scale of structural well-formedness, not degrees of word salad.

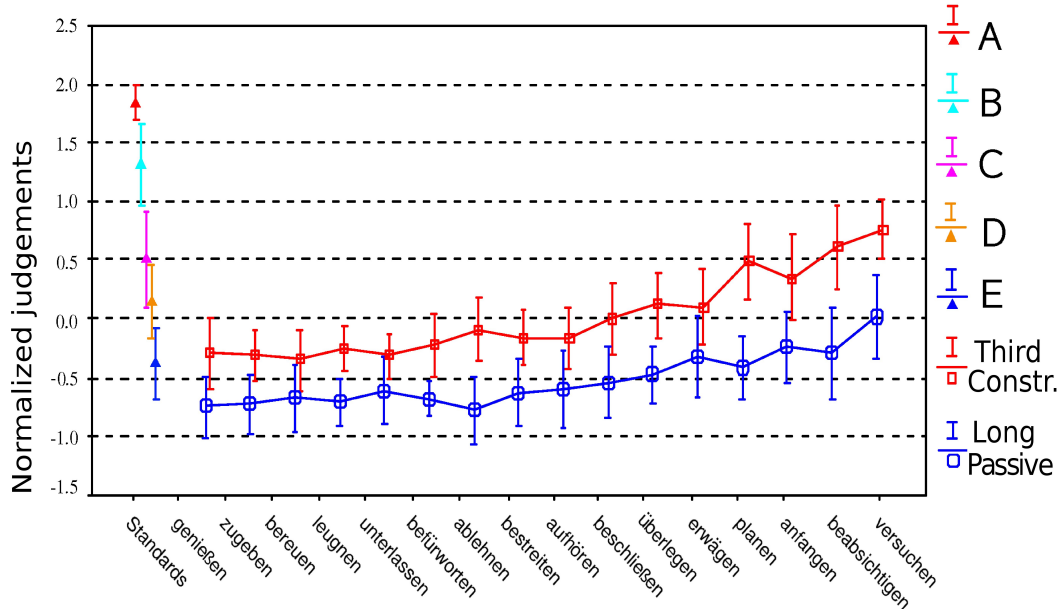


Figure 1: The use of standard comparison items in a judgement experiment.

To test whether speakers would be able to use the scale, we carried out a further study in which we explained the five bands of well-formedness to informants and instructed them to assign example sentences to the bands. We then assessed, first, the degree of reliability with which speakers are able to attribute examples to five bands of well-formedness, and second, the consistency between informants. The results showed that subjects were easily able to perform the task and mostly assigned examples to the correct group or to an adjacent group (91%). Informants agreed to 93% within 1.25 of a grading band.

Figure 1 shows the use of these standard comparison items as anchor values in a magnitude estimation of well-formedness study of the effects of matrix verb on the third construction and long passive in German. The five items A – E are shown in on the left in the graphic. The experimental conditions (third construction, long passive) form two continuums across the centre. The comparison items reveal that even though the two experimental conditions become better when more extraction-friendly predicates are used, these structures are only ever at best marginally well-formed.

We round off the talk with a review of the beneficial effects of the use of cardinal well-formedness values:

- introspective judgements attain a degree of objectivity.
- results from one study using relative judgements can be compared to those of another, via the tertium comparationis of the standard items.
- the availability of known reference points can sharpen the definition even of the single person's judgements.
- the data from relative judgements becomes more usable as the empirical underpinning of gradient grammatical theory.

Outlook: Might it be possible to develop cardinal well-formedness values in more than one dimension, like the cardinal vowels from phonetics?

References

- Bard, Ellen; Robertson, Dan & Sorace, Antonella (1996). Magnitude Estimation of Linguistic Acceptability. *Language* 72 (1), 32-68
- Cowart Wayne (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgements*. Thousand Oaks, California: Sage
- Keller, Frank (2000). Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality. PhD Thesis, University of Edinburgh
- Schütze, Carson (1996). *The Empirical Base of Linguistics: Grammaticality Judgements and Linguistic Methodology*. Chicago: University of Chicago Press