Relax, lean back, and be a linguist

Sam Featherston

February 16, 2009

In this text I wish to raise some questions about the quantity and quality of data required in grammar research. In previous work I have argued, often emphatically, in favour of using more and better data, so that grammatical hypotheses are more descriptively adequate, and theory building less speculative. My arguments have been just a small part of a wider intellectual trend in this direction, since others have had similar thoughts, and research in the field is increasingly using more evidence-based argumentation. I therefore think that we can assume that the case for greater empirical input into grammatical generalizations and theory building is established.

My aim here is to question how far this trend need go. The title of this piece is intended to sum up my suggestion that, while it was necessary and indeed urgent for researchers in grammar to consider data and respect data more, it is neither necessary nor perhaps desirable for them to spend too much time worrying about the finer points of data collection and analysis. Some researchers will take a particular interest in methodology and will innovate and set new standards, but others will be more conservative and rely to a greater extent on traditional methods, while still accepting the basic premise that a theory is an account of data.

My basic point is that we need to regard both paths as valid, if we wish the field of research in grammar to remain a single unit, and not drift apart into separate discourses, to the disadvantage of both parts. One of the insights of the SFB 441 Linguistic Data Structures was the extent to which the data types and approaches represented had significant features in common. In particular, it became clear that linguistic data is always complex and requires filtering, interpretation, and location within a wider model to yield its full evidential value. The developing different wings of grammar research therefore need each other; neither all-data nor all-theory can have as much value as a judicious combination of the two.
This paper therefore contains three suggestions what a middle way might look like, a consensus which recognizes data as a prerequisite of description and explanation, but still remains accessible to more theory-oriented colleagues. All of them have to do with aspects of exactness in linguistic data. As the title implies, I suggest that empirical grammar is possible without worrying too much about precision.

**Relax, . . .**

The current interest in experimental work in grammar research makes it seem likely that a new consensus will soon emerge on the experimental standards required for a linguistic claim to be taken seriously. I wish to make a plea that the new consensus should be fairly tolerant, at least in the area of judgements, where I have experience.

Tolerance is important because it can save linguists time and effort. The challenge in setting standards is to distinguish a) those factors which need to be carefully controlled for to avoid type I error (false positives) from b) those which have negligible effects and those which can be controlled by allowing them to vary freely, at a small risk of type II error (false negatives). Type I error is the main problem, as standards should prevent false claims being supported; type II error, on the other hand, can be largely left to the discretion of the individual researcher, especially since it usually takes the form of finding no difference where there really is one, and this sort of no-difference result offers only weak evidence for or against anything.

So where can we exercise tolerance? In his seminal work Carson Schütze (1996) discusses very many subject-related and task-related factors which might distort or falsify linguistic judgements (also Cowart 1997). But in our own experience of gathering relative judgements, none of the personal variables has ever had any meaningful effect on the results. We would therefore suggest that many variables need not be controlled for, all other things being equal and as long as extreme values are avoided. Please note these two conditions carefully; some of these factors might have significant effects in certain specific conditions: dialect background, for example, if (but only if) the language tested contains strongly dialect-sensitive material. Similarly, linguists can deliver results indistinguishable from non-linguists', but it would be unwise to gather data from a linguist with an interest in the point being studied. Again, lexical frequency has little effect in judgements over a broad range of values, but unfamiliar words will cause distortion.

Extreme values should be avoided, but usually simply allowing them to vary controls for them quite adequately. One might add that allowing subject-related variables to vary freely has the advantage that it strength-
ens the claim that the findings are generalizable to the whole population of speakers. These variables do not seem to make a great difference:

- social, educational and professional status of subjects
- dialect background of subjects (except . . .)
- the sex, age, and handedness of subjects
- a degree of linguistics training among subjects (but . . .)
- frequency in lexis (as long as extremes are avoided)
- speed of response required
- structure type being tested (statements, questions, dialogues)
- carrying out experiments over the net or on paper
- the precise methodology, as long as it asks subjects about their receptive responses to examples

To summarize, many of the worries about possible distorting factors in Schütze (1996) and elsewhere are less problematic than was feared; additionally various contributions (eg Weskott & Fanselow 2008, Bader & Häussler 2008) have shown convincingly that the differences between methodologies gathering judgements are in practice fairly small, since all give satisfactory results in simple cases.

On the other hand, other factors must be carefully controlled. The linguistic materials need to carefully constructed in order to gather comparable data across items and conditions:

- plausibility of content of experimental materials
- sentence or phrasal length and complexity, for example ...
- ... referential abstractness (eg too many pronouns)
- unclear meaning
- low accessibility of intended interpretation

My suggestion is thus that linguists can legitimately avoid worrying too much about methodology and their subjects (though they need to take care with the linguistic materials) and can just concentrate on arguing from some data. Methodologically, we can ... relax.

... lean back ...

Elsewhere I have said (Featherston 2007) that the quality of a grammatical analysis can never exceed the quality of the data base. Am I here contradicting this? No, because ‘quality’ in data does not necessarily demand fineness of detail. Quality can also mean relevance: to establish the relative positions of German verbs and their objects requires careful thought and a good look at lots of different examples in different structural contexts. It does not, however, require very much fineness of detail in the data. I would argue that
many of the big questions in grammar are fairly robustly visible in data, and many of the smaller questions depend crucially on the answers to the big questions: but since we haven’t yet answered the big questions, there is still room for significant progress using just the masses of data beyond serious question. For many grammatical issues therefore, we need not worry about the finer detail. Let me state quite clearly here that this should not be taken as devaluing data of fine detail. There are plenty of questions which resoundingly do require finer detail, but also many which do not, so at least some of the time, we can lean back and look at the clear patterns.

The idea of leaning back is relevant in a second way too. An important part of the current trend towards empirical grammar research is the way that linguists look at data. If you use individual examples to validate points made in an article, then you look at data from within a theoretical perspective. You will also tend look at fragments of data rather than the wider patterns. This approach can continue even to experimental data: if we design an experiment with the simple aim of confirming or falsifying a hypothesis, we may be happy with the appearance of a single significant difference predicted by one account and not the other. There is of course nothing wrong with this.

But it is possible to look at data with a wider perspective. When we gather data more broadly, the data patterns gain an independent existence as evidence in themselves: we can print out the results as bar charts and hang them on the wall. Linguists interested in grammar architectures can look at wider sets of data and ask themselves what sort of system could produce the data patterns that they find. This is data-driven linguistics, and while it can usefully be theoretically informed, it need not be constrained by theory. The work done on the relationship between data and theory within the SFB 441 has demonstrated the value of this approach (eg Featherston 2005). To find out what grammar is supported by what data type, I can recommend that linguists lean back and look at the landscape of the data.

\[\ldots\text{and be a linguist.}\]

This final point relates to the use of statistics in papers reporting experimental work. I would like to suggest that statistics are a tool for testing how well a set of results supports a linguistic conclusion. They are very valuable in this function, but they need not be more than this.

The Results section in a paper often contains two parts: a graphic illustrating the results and a section of text describing them directly below it. This text section is often quite unreadable, because it consists of a list of the significance values of main effects and interactions, with only a very little text between the brackets opening and closing and the various $F$-values, $p$-values
and degrees of freedom. Placing this text section directly after the results graphic implies that this paragraph (or page ...) is the true description of the results of the experiment. This seems to me to be giving these statistical tests a prominence that they do not deserve. The results of statistical tests would be better located within the text, at the point where the linguistic point that they refer to is discussed. Those statistics which are not used as part of an argument, for instance, whether a main effect is significant when the predictions relate to an interaction, could be relegated to an endnote.

Papers on grammar require more than just the reporting of data; they require the analysis and the interpretation of data, as well as some discussion showing what new insight can be gained. Notice that linguistics contrasts here with the sciences. Scientific papers in journals are often short and very factual, containing a set of findings plus a description of their relation to previous work. Linguistics papers demand far more: a discussion of why the issue is important in the first place, the motivation for the additional work, an account of the wider implications; many of these need to be argued for.

This is a consequence of the nature of linguistics as a subject. Core linguistics perhaps most closely resembles economics in its subject matter, given that both address a fundamental aspect of human behaviour. Sub-parts of both disciplines can usefully be modelled with quantitative data, but these establish merely the facts of the case; the wider implications do not automatically follow. Highly complex models of the economy are run on powerful computers in economics research centres world-wide, but the conclusions of the various institutes still vary quite widely. In both linguistics and in economics, the contribution of human analysis and interpretation is essential: the computer models of the economy were not predicting the global financial meltdown of 2008 even after the ‘sub-prime’ scandal hit the headlines. Some economists, on the other hand (eg the Guardian’s economics editor Larry Elliot), had been warning the financial world for some time.

Conclusion: quantitative data in general and statistical tests in particular do not in themselves tell us the answer to linguistic questions. They are tools: powerful tools, but still just tools. All they can do is produce a quantitative measure of how well some data supports our hypotheses. I therefore suggest that linguists use data and apply statistical tests, but do not forget that both the starting point and the end point of a study must be a grammatical analysis. Be a linguist.

In this text I have put forward arguments for methodological tolerance, for the value of looking at the bigger picture as well as the detail in data, and for the primacy of linguistic analysis over statistics. Not everyone will agree with the approach to data which I have advocated, but I would like
readers to understand why I think that that it is justified.

The main reason is what I consider to be the kernel of the trend to more empiricism. The important step is the acceptance that an explanatory theory presupposes an observationally adequate data base. Once grammarians are arguing from some data rather than arguing from elegance or economy, the essential conceptual step towards empirical adequacy is taken. Most of us now would recognize that research can contain imaginative leaps – such leaps are in fact the generation of hypotheses – but that these need to be tested and sometimes rejected. Empirical grammar research demands only the acceptance that hypothesis generation and hypothesis testing are equally valid and valuable parts of the research cycle. What data is used is less crucial: experience in the *SFB 441 Linguistic Data Structures* has taught the value of plurality in data types and approaches.

The second reason is that paradigm shift is necessarily slow and probably partial. If we wish to change the attitude to data of the whole of the grammar research community, this must be done slowly, and through persuasion and consent. A sharp break with previous practice will cause a split between grammarians who do and those who do not adopt the more empirical approach. I think we should aim at the higher goal of changing the practice of grammar research from the inside, not just forming a break-away group.

References


