HOW TO CONSTRAIN THE SET OF POSIBLE CAUSAL EXPLANATIONS

Ramon Ferrer-i-Cancho
Complexity and Quantitative Linguistics Lab
LARCA Research Group
Universitat Politècnica de Catalunya

The dream of any language researcher is to understand the causes of linguistic phenomena. However, deciding which explanation from a set is really the cause, or closer to the cause, is a very complex problem in my modest opinion. In the era of publish-or-perish, it is tempting to produce simple models tailored to particular problems leaving the issue of causality aside. That helps to increase the length of one’s CV but does not help significantly to the progress of science.

The goals of my talk are very modest. I cannot offer a guide to identify or design causal models. I can only offer limited help to constrain the set of candidates for a causal explanation by means of some considerations that are often neglected when evaluating candidate hypotheses or models:

1. A good model is one with a special combination between parsimony and quality of fit (Burnham & Anderson, 2002). Be careful about models whose only virtue is parsimony or their predictive power, but not both at the same time. The simplest model is an empty model and a model that fits the data nicely is at a risk of suffering from overfitting.

2. Simple models that only apply to a narrow domain (even those showing a nice balance between parsimony and quality of fit) cause the illusion of being right by their local suitability but they can be a burden when integrated into a general theory (Ferrer-i-Cancho, 2014a). As Mario Bunge puts it: "Scientific knowledge is systematic: a science is not an aggregation of disconnected information, but a system of ideas that are logically connected among themselves”. This point is intimately related to 1.

3. Be careful with strong hypotheses that are presented as default or null hypotheses. What a null or parsimonious hypothesis is may depend on one’s hidden assumptions. This problem can also be metatheoretical, e.g., the belief that specialization and fragmentation of knowledge is the natural way of science (which is related to point 2). However, the belief in that some degree of unity across disciplines, levels and domains is necessary
can lead to more parsimonious solutions at a global scale. Fragmentation of knowledge should be the last resort, not the default one.

4. Claims about the optimization in language require (a) a clear statement of a cost function, (b) a baseline and, (c) a clear connection between the minimization of the cost function and observed phenomena that are interpreted as direct (or indirect) evidence of optimization. (Ferrer-i-Cancho, 2014a; Ferrer-i-Cancho, 2014b). The risk of failing to identify the causes is higher when at least one of these components is missing.

Each point will be illustrated with examples of hypotheses and models.

REFERENCES


