Discovery and justification in linguistics

Pauli Brattico

Proposal

Methodological principle:

Scientific theories and hypotheses must be justified by showing that the data follows from them (17th century scientific method)

Alternative

Methodological principle ("divine illumination"):

The human mind can see truths without objective checking (classical and medieval periods)

- **The 17th century scientific method is motivated by the skeptic who claims that**
  - the assumptions might be internally inconsistent, and the attested data might not follow while some non-attested data follows;
  - the whole idea, hypothesis or theory can be dismissed on these grounds without looking into anything else;
  - as so in the advanced sciences they are dismissed without consideration. "First show me the proof."
- **To answer the skeptic**
  - all subjective gaps and ambiguities are removed from the data, theory and the connections between the two;
  - no pseudo-formalizations, proof sketches, phrase structure storyboarding, informal English, intuitions, hand-waving – all eliminated

Step 1

- To calculate linguistic data, it must be provided in some quantifiable form

- We create a file which contains all the data (whether phonological morphological, syntactic, semantic, pragmatic) representing, in a systematic, fully explicit and possibly gapless way, the phenomenon of interest

- Examples...

Step 2

- To use computation, we formalize the theory/hypothesis in some machine-readable formalism (i.e. provide definitions for Merge, Move, Agree…)
- Examples…

Step 3

- We write functions which calculate through every empirical consequence of the theory/hypothesis (Step 2) and compare the results with the datatset (Step 1)
- The procedure constructs rigorous proofs and answers the skeptic once and for all:
  - The data is unambiguous and provided in an explicit format;
  - The theory is formal, every notion in it is defined logically, nothing is unambiguous, no concept remain unclear, there can be no confusion over anything;
  - Inferences from the latter to the former are unambiguous and can be replicated in their entirety.