Jakub Kozakoszczak (Uniwersytet Warszawski)

Title: Negative evidence from text corpora in semantics

Abstract: I will present two partial results of my PhD project that aims at providing an alternative linguistic tool to acceptability judgments in their evidential role in language science. On the theoretical side I will give reasons for replacing exclusion-based concepts like ungrammaticality, unacceptability, or semantic deviancy with the functional concept of intended meaning failure (IMF). An IMF occurs when a given expression fails to convey a coherent possible intended meaning (PIM). PIMs are under-discussed but present in linguistic literature, particularly in glosses. For example, some possible intended meanings of *Pi seems irrational are:

1. 'Pi seems irrational1'

2. 'Pi seems irrational2'

3. 'Paul seems irrational1'

4. 'Pi must be irrational2'

PIM is a meaning one could intend to communicate with a given expression, a meaning which is coherent (see 3., 4.) or compositionally derivable from the expression without any repairs (see 1., 2.) but not necessarily both. I will argue that the sole unacceptability of a bare expression is evidentially inconclusive unless we start with a hypothesis about its PIM and try to arrive at it compositionally. Instead of one equivocal piece of data we can obtain multiple IMFs caused by different flaws, calling for different explanations and supporting unrelated hypotheses.

Apart from higher causal transparency and taking account of polysemy IMF has the theoretical merit of being more general than unacceptability. Expressions that fail to convey a coherent meaning provide negative linguistic evidence only because they are useless and uselessness is implied by unacceptability anyway. On the other hand, unacceptability is not implied by uselessness.

On the methodological side I will outline the method of conditional structural zeros (CSZ) and I will present a pilot study on a class of Polish unacceptable verb phrases which employs this method. The method is designed to tell PIMs whose occurrence probability equals zero (structural zeros) from PIMs too rare to occur in the corpus sample (sampling zeros). This is done essentially by comparing underrepresentations of PIMs against underrepresentations of similar meanings and by comparing likelihoods of competing semantic hypotheses. The intended application of the method is to use it to obtain verifiable negative evidence from natural data alone with less appeal to unobservable intuitions as in acceptability judgments.