

Stefan Th. Gries

Dimensions of phraseologisms, their measurement, and their motivation

Much phraseological work in the past has been based in particular on lexicographic resources and/or statistical measures such as association strength to approach the question of phraseologism identification. In this talk, I will discuss a variety of dimensions relevant to this process as well as their corpus-based approximations/operationalizations, arguing that corpus data have much more distributional information to offer than association strengths such as MI or LLR alone. Specifically, I will advocate an analytical process that separates frequency, directional associations, entropy, and dispersion to illuminate different characteristics of phraseologisms and will motivate them with reference to cognitive or psycholinguistic work in corpus-based usage-based/cognitive linguistics.

Beyond frequency: how else to count what counts

This talk will be a brief hands-on session exemplifying how to compute from corpus data measures of directional association, entropy, and dispersion. We will use the open-source software R to study - in a very small corpus - the behavior of several words to provide participants with a first practically-informed understanding of how to apply such measures to their own work.