

Lexicons encode differently what people do differently. Computational studies of the pragmatic motivations of lexical typology.

Barend Beekhuizen

University of Toronto, Toronto, Ontario, Canada

Abstract

Languages differ in what meanings their lexical items encode: The meaning covered by English 'blue' is famously split into 'sinij' (darkblue) and 'goluboj' (lightblue) in Russian. Recent years have seen novel interest in functional explanations of such variation, pointing to a correlation between greater communicative need of a lexical field and a finer-grained lexical inventory. Here, I develop the position that rather than the mere difference in "need" to mention lexical field, it is the field's discourse-pragmatic diversity that predicts whether languages "lump" or "split" more. I will demonstrate this with computational techniques and a typologically diverse corpus of spontaneous spoken data from 51 languages (DoReCo), first for the field of verbs of visual perception ('see'-'look'), then on a lexicon-wide level. There are implications: our notions of what a comparable concept is in lexical semantics, what lexical knowledge entails, and the dimensions along which languages differ require re-examining.