Languages optimize the trade-off between lexicon size and average utterance length: A case study of numeral systems

Human languages vary in terms of which meanings they lexicalize, but there are important constraints on this variation. It has been argued that languages are under pressures to be simple (e.g., to lexicalize few meanings) and to be informative (i.e., to allow for a precise communication), and that a good compromise between these two pressures determines which meanings get lexicalized (Kemp and Regier 2012, a.o.). We will argue for a modification of this proposal: we show that there is another pressure influencing which meanings languages lexicalize, the pressure to produce morphosyntactically not too complex utterances. We conduct a case study on numeral systems, showing that simplicity/informativeness approach cannot explain which number meanings are lexicalized cross-linguistically. We proceed to show that natural languages numeral systems are finding a good compromise between the pressures to lexicalize few meanings (i.e., to minimize lexicon size) and to produce morphosyntactically not too complex utterances (i.e., to minimize average utterance length). This case study in conjunction with previous work suggests that, to explain which meanings get lexicalized across semantic domains, a more general approach may be that languages are finding a compromise between three pressures: be simple, be informative, and minimize average utterance length.

Click here to see the colloquium program.