

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Information Distribution Depends on Language-Specific Features

Permalink

<https://escholarship.org/uc/item/2zc7p9vf>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Klafka, Josef

Yurovsky, Dan

Publication Date

2019

Peer reviewed

Information Distribution Depends on Language-Specific Features

Josef Klafka

University of Chicago, Chicago, Illinois, United States

Dan Yurovsky

University of Chicago, Chicago, Illinois, United States

Abstract

Language can be thought of as a code: A system for packaging a speaker's thoughts into a signal that a listener must decode to recover some intended meaning. If language is a near-optimal code, then speakers should structure information in their utterances to minimize the impact of errors in production or comprehension. To examine the distribution of information within utterances, we apply information-theoretic methods to a diverse set of languages in various spoken and written corpora. We find reliably non-uniform and cross-linguistically variable information distributions across languages. These distributions are consistent across contexts, and are predictable from typological features, most notably canonical word order. However, when we include even a small amount of predictive context (bigrams or trigrams), the language-specific shapes disappear, and all languages are characterized by uniform information distribution. Despite cross-linguistic variability in communicative codes, speakers structure their utterances to preserve uniform information distribution and support successful communication.