

**Peter Fankhauser**

**Observing Surprisal through the blurry lense of Language Models**

Entropy measures the average surprisal in text on the basis of language models. It is a versatile tool for text analysis tasks such as speech recognition, collocation and phrase detection, text segmentation and corpus comparison. In this talk I will investigate how entropy varies with the domain of underlying Language Models (document, corpus, background) on a variety of corpora. Moreover, I will analyse the interaction of entropy with other contextual variables, such as position in text, sentence length, register, and field; debunking some popular myths about constant entropy rate on the way.