



He is CEO of Lexical Computing, a research company working in the area of language technologies, primarily at the intersection of corpus and computational linguistics and computer lexicography. By profession He is an NLP researcher and software engineer. His research interests are devoted to effective processing of very large text corpora and computer lexicography. He has been involved in numerous projects related to automatic corpus processing and annotation as well as dictionary production.

Since 2008 He has been involved in the development of Lexical Computing's flagship product, the Sketch Engine corpus management suite. Since 2011 He has been director of the Czech branch of Lexical Computing leading the local development team of Sketch Engine and He became CEO of Lexical Computing in 2014.

He is also a fellow of the NLP Centre at Masaryk University, where his interests lie mainly in morphosyntactic analysis and its practical applications.



Guest Speaker : JAKUBÍČEK Miloš Sketch Engine - Lexical Computing CZ s.r.o. & Masaryk University in Brno

[See the announcement](#)

Title: Post-editing lexicography

Language : English

Date : 15/11/2024

Time : 17:00-18:30 (Beijing time) / 10:00-11:30 (Paris time)

Venue : University of Huazhong & Online (France and China)

Organizers : GENG Yundong 耿云冬 & CHEN Lian 陈恋 & DAO HuyLinh [Registration](#)

Abstract:

In the talk I will review the development of methods for automating dictionary production from corpus data. I will start with a brief overview of the "historical" corpus revolutions (as identified e.g. by Rundell, 2008) and their implications on the lexicographic process, with the main part of the talk focusing on the most recent change leading to a lexicographic workflow called « Post-editing lexicography". I will discuss in detail to what extent this is a revolution or evolution, how individual parts of the lexicographic workflow are affected and what are the main implications on the lexicographic business process. Finally, I will review these changes in the light of the availability of large language models and general AI development.