

How does the distribution of words in language relate to meaning?

From collocations to distributional semantic models



Workshop

Prof. Dr. Stephanie Evert

FAU Erlangen-Nürnberg

Chair: Prof. Dr. Raphael Berthele, UNIFR

Thursday 01/10/2026, 09:00 – 17:30

Friday 02/10/2026; 09:00 – 17:30

Institut de plurilinguisme

Université de Fribourg

Rue de Morat 24, 1700 Fribourg (Suisse)

☞ Salle K103

Corpus linguists have long been paying attention to the close connection between the meaning of a word and the contexts in which it is used, i. e. its distribution in language. In this workshop, we will explore two important quantitative approaches for investigating this relation empirically: collocations and distributional semantic models (DSMs).

Collocations are habitual word combinations that reflect where a word typically occurs in language. They have found important applications e. g. in language acquisition research and teaching, in computational lexicography, and for identifying discursive patterns in corpus-assisted discourses studies. Distributional semantic models encode the distribution across contexts into high-dimensional vector representations of word meaning. As they provide a flexible approach to word meaning, DSMs have been used e. g. to measure semantic similarity or to visualize word lists as a semantic map.

Through this hands-on introduction to the foundations of collocations and DSMs, participants will explore how collocational profiles and distributional vector representations are related, and what they can tell us about language and word meaning.

A follow-up workshop, focusing on large language models and context-sensitive embeddings, is planned in Fribourg as part of the 2027 CUSO doc-

toral program Sciences du langage. No prior knowledge of R or Python is required for these workshops.

Biography

Stephanie Evert is Chair of Computational Corpus Linguistics at FAU Erlangen-Nürnberg. After studying mathematics, physics and English linguistics, she received a PhD in computational linguistics from the University of Stuttgart. Her research interests encompass the quantitative methodology of corpus linguistics, multivariate analysis and distributional semantics, applied corpus studies and digital humanities, tools for processing large text corpora, the combination of human interpretation with machine learning (digital hermeneutics), as well as language technology and its applications.

PhD programme on Multilingualism

The event is part of the doctoral programme Multilingualism: Acquisition, Education, and Society (more information here: <https://institute-multilingualism.ch/en/doctoral-programme>)

Participation

To register, please contact Dr. Marie-Anne Morand (marie-anne.morand@unifr.ch) by September 20, 2026.