

## Keynote Lecture

### How to Identify Multi-Word Expressions in Corpora?

Authors: Vaclav Brezina

Affiliation Lancaster University

#### Outline

Multi-word expressions (MWEs) such as *of course*, *in the light of*, *as good as new* or *take into account* are central to fluent communication, yet relatively difficult to identify systematically. This lecture draws on the forthcoming *Frequency Dictionary of Multi-Word Expressions in British English* (Brezina & Gablasova, Routledge, 2026) to present a clear, corpus-based method for analysing a wide range of MWEs across genres.

I begin by contrasting corpus evidence with current AI language models. Modern AI generates language by answering a simple question – *What is the next word?* – a principle long used in corpus linguistics to measure collocation. Yet while AI can imitate fluent usage, corpora remain more transparent and reliable for identifying MWEs because they trace patterns directly to authentic human interaction in speech and writing.

The lecture outlines a practical framework combining frequency, association strength and dispersion to capture the core phraseology of contemporary British English. Examples from the dictionary illustrate how this method reveals stable, meaningful MWEs and supports applications in language teaching, language testing, lexicography and applied linguistic research. The central claim is this: AI can model language, but corpora allow us to understand it.

Visit [Vaclav Brezina keynote page](#) to learn more