

Maria Ermakova/Alexander Geyken/  
Lothar Lemnitzer/Bernhard Roll

## INTEGRATION OF MULTI-WORD EXPRESSIONS INTO THE DIGITAL DICTIONARY OF GERMAN LANGUAGE (DWDS)

### Towards a lexicographic representation of phraseological variation

**Abstract** One central goal of the project ‘Zentrum für digitale Lexikographie der deutschen Sprache’ (Center for digital lexicography for the German Language, [www.zdl.org](http://www.zdl.org)) is to provide a corpus-based lexicographic component of common German multi-word expressions (MWE), including idioms, for DWDS ([www.dwds.de](http://www.dwds.de)), a general language dictionary of contemporary German. As a central challenge of this task, we have identified an adequate lexicographic representation of such common properties of MWE as variation and modification. To document the variation, we have developed a special entry-clustering model, which we call *hub-node entry*. This model comprises a core hub entry headed by a short nuclear form of the MWE and several node entries, which represent the most common variants in their full lexical forms.

**Keywords** Multi-word expressions, phraseological variation, dictionary entry structure

### Contact information

**Maria Ermakova**

Zentrum für digitale Lexikographie der deutschen Sprache  
Berlin-Brandenburgische Akademie der Wissenschaften  
[ermakova@bbaw.de](mailto:ermakova@bbaw.de)

**Alexander Geyken**

Zentrum für digitale Lexikographie der deutschen Sprache  
Berlin-Brandenburgische Akademie der Wissenschaften  
[geyken@bbaw.de](mailto:geyken@bbaw.de)

**Lothar Lemnitzer**

Zentrum für digitale Lexikographie der deutschen Sprache  
Berlin-Brandenburgische Akademie der Wissenschaften  
[lemnitzer@bbaw.de](mailto:lemnitzer@bbaw.de)

**Bernhard Roll**

Zentrum für digitale Lexikographie der deutschen Sprache  
Berlin-Brandenburgische Akademie der Wissenschaften  
[roll@bbaw.de](mailto:roll@bbaw.de)