

# Adaptive word sense views for the dictionary database eWDG: The case of definition assignment

Axel Herold and Alexander Geyken

**Abstract.** We describe an approach that uses explicit and implicit references in the dictionary database eWDG to automatically assign appropriate definitions to undefined headwords. The approach is developed and evaluated in the context of a lexical information system that aims to generate adaptive views on dictionary and corpus resources.

## 1 Introduction and project context

The approach described in our paper is being developed in the context of the dictionary project “Digitales Wörterbuch der deutschen Sprache” (DWDS, en.: “Digital Dictionary of the German Language”). This long-term project was launched in 2000 at the Berlin-Brandenburg Academy of Sciences (BBAW) with the goal to build an online lexical information system that provides access to multiple German text corpora and lexical resources (Klein 2004a,b; Klein and Geyken 2000). The current release of this information system (Geyken 2005) constructed in the first project phase from 2000 to 2006 offers flexible access to four different types of resources:<sup>1</sup> (1) a corpus component with currently 800 million word tokens in total (Geyken 2007), (2) a collocation component offering options to compute collocations for a lexical unit according to common statistical measures, (3) a thesaurus component that computes synonyms, hyponyms, and hypernyms for lexical units on the basis of the dictionary data (Geyken and Ludwig 2003), and (4) the dictionary database eWDG (electronic WDG). This database builds on a six-volume print dictionary of German, the “Wörterbuch der deutschen Gegenwartssprache” (WDG, en.: “Dictionary of Present-day German”) published between 1952 and 1977 and compiled at the Deutsche Akademie der Wissenschaften<sup>2</sup>. The print dictionary comprises over 4 500 pages and contains more than 60 000 headwords (more than 120 000 if compounds are counted separately). The term “deutsche Gegenwartssprache” (en.: “German present-day language”) is understood broadly by the lexicographers; the dictionary is not restricted to the language spoken and written in the middle of the

---

1. <http://www.dwds.de>

2. since 1972: Akademie der Wissenschaften der DDR