

# FEWS (Few-shot Examples of Word Senses) Data Sheet

FEWS is a dataset for English Word Sense Disambiguation (WSD). Each example in FEWS contains an ambiguous word  $w$  and a sentence  $s$  that contains  $w$ ; the example is labeled with the dictionary definition describing the sense, or specific meaning, of  $w$  when used in that context  $s$ . For example, the sentence "I **liked** my friend's status..." contains the ambiguous word **liked**. The goal of a model trained on FEWS is to identify the corresponding sense of liked from that sentence, out of all possible senses of **liked**.<sup>1</sup>

FEWS is gathered from the definitions and example sentences provided in Wiktionary, an online crowdsourced dictionary. Due to this data source, FEWS contains a limited number of examples for each sense, and the dataset focuses on (1) few-shot learning of WSD and (2) evaluating the performance of WSD on rare senses of words.

Data Split	# Examples	# Tokens	# Annot.	# Sense Types	# Word Types	Ambiguity
Overall	121,459	3,259,240	131,278	71,391	35,416	5.62
Train	87,329	2,551,358	96,023	52,928	30,450	4.98
Ext. Train	101,459	2,683,345	111,278	61,391	31,937	5.71
Dev	10,000	287,673	10,000	10,000	8,682	5.09
Few-shot	5,000	149,791	5,000	5,000	4,417	4.77
Zero-shot	5,000	137,882	5,000	5,000	4,661	5.41
Test	10,000	288,222	10,000	10,000	8,709	5.10
Few-shot	5,000	149,384	5,000	5,000	4,449	4.71
Zero-shot	5,000	138,838	5,000	5,000	4,666	5.49

Table 1: FEWS data statistics. Reproduced from (Blevins et al., 2021)

The dataset statistics for FEWS are shown in the table above. FEWS contains approximately 1.7 examples per sense (with over 121,000 examples across more than 71,000 senses); this means that FEWS provides a few-shot learning and evaluation setting for WSD. The senses in FEWS cover many different domains of English, due to the fact that the examples in FEWS come from a crowd-sourced resource. These domains span from historical and literary meanings, to niche technical terms and new senses pertaining to contemporary culture.

We note that a small portion of FEWS (estimated  $<1\%$  of the data) contain examples of toxic or offensive meanings of words; our data analysis finds senses marked with tags such as *ethnic slurs*, *offensive*, *vulgar*, and *derogatory*. We choose not to exclude these examples because: (a) they provide real senses of English words that are beneficial for a WSD model to capture, and (b) they are potentially useful for learning to predict the use of offensive language. Appropriate uses of these examples include training a WSD or hate speech detection model; we do not intend for this data to be used as training data for a general text generation model.

For more details and analysis of FEWS, please see (Blevins et al., 2021) and the corresponding README provided with the dataset.

## References

Terra Blevins, Mandar Joshi, and Luke Zettlemoyer. 2021. FEWS: Large-scale, low-shot word sense disambiguation with the dictionary. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*.

<sup>1</sup>In this case, the correct sense of **liked** would be "to show support for, or approval of, something on the Internet by marking it with a vote."