

From the first lexicographic endeavours, lexicographers have been quite cautious about claiming that representations of meaning in their dictionaries are absolute. Unfortunately, one usually finds such warnings only in prefaces, dictionary front matter, commentaries and other more obscure works, and they are in most cases ignored. However, it is important to take these warnings seriously, especially when lexicographic resources are used as absolute truths for semantic tasks in Natural Language Processing.

Natural language understanding (NLU) is one of the most important goals in Artificial Intelligence. In the race to achieve the NLU goal, the fact that knowledge is essentially defined by humans sometimes tends to be forgotten. In this sense, understanding what constitutes shared meaning in a language community, or between different language communities, is still quite relevant. As a consequence, modern lexicography needs to adapt and provide adequate data (also) for NLU tasks. In the talk I will try to provide some answers to the following questions: are we actually measuring anything meaningful in NLU benchmarks (e.g. WiC in SuperGLUE)? Can modern lexicography help with this? Will lexicographic data look the same in the next 5-10 years and will it (have to) be openly accessible?