

Semantic Similarity: What for?

Ido Dagan

Bar Ilan University

dagan@macs.biu.ac.il

Abstract

Linguistic similarity has been a prominent notion and tool in computational linguistics and related areas, as elaborated nicely in the announcement of this workshop. Yet, what exactly counts as “similarity”, or when two linguistic concepts should be regarded as similar, often remains rather vague and ill posed, which is in fact quite typical for unsupervised notions. This talk will focus on similarity at the semantic level, and will explore the perspective that different notions of similarity may be defined relative to concrete modeling goals. In particular, I will refer to the two major goals in semantic modeling: predicting likelihood of occurrence, which is the typical goal in disambiguation and language modeling, and recognizing target meanings, which is the typical semantic goal in text understanding applications such as question answering, information extraction, summarization and information retrieval. We will discuss each goal and present corresponding semantic similarity approaches.