

Parameters for Word Sense Disambiguation Used in Multilingual Settings

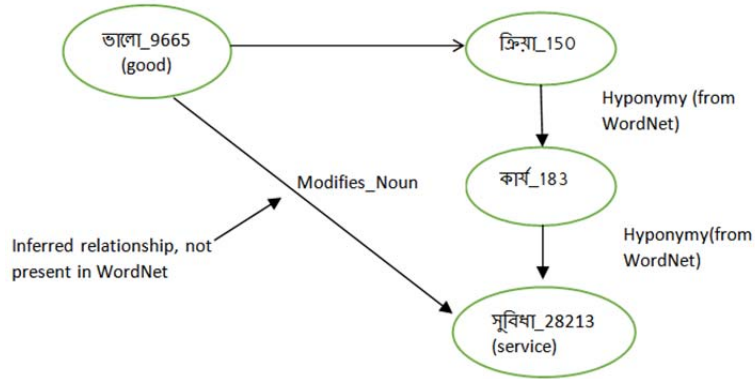
Arindam Roy and Sunita Sarkar

Assam University, Silchar

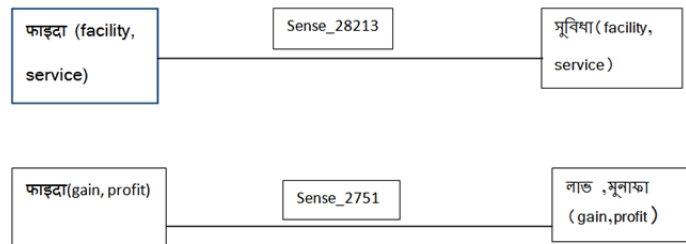
Abstract—*Word Sense Disambiguation (WSD) is a mechanism to obtain the sense of target words (words meant to be disambiguated) or all words against a sense repository like the WordNet or a Thesaurus using the context in which the word appears. A basic axiom of WSD is words around the target word, known as context words, provide strong and consistent clues for disambiguation.*

There are important parameters which facilitate WSD to a great extent. These parameters may be domain specific or span across languages. These are as follows:-

1. **Domain Specific FrequentSense**:- It can be gleaned from sense tagged corpora. The domain specific frequent sense can be different from the WordNet first sense. For e.g. the most frequent sense of the word “উপনিবেশ” in the Tourism corpus is in the sense of “a geographical area politically controlled by a distant country” whereas the WordNet first sense is “an area where a group of families(of a different nationality) live together”.
2. **Conceptual Distance** :- Conceptual distance between two synsets (S1,S2) is given by Conceptual Distance increases with path length between two synsets and it is inversely proportional to the height of the common ancestor . As the conceptual distance between two synsets become more, the synsets tend to get more separated in terms of meaning. Conceptual distance can only take place in a hypernymy-hyponymy relationship.
3. **Semantic Graph distance**:-It is defined as the shortest path length between any two synset nodes in the WordNet graph. An edge on the shortest path can be any semantic relation in the WordNet (as opposed to conceptual distance where the edges are only hyponymy-hypernymy relations.Let us consider the sentence “কোহিমাতেমোকোচুসমার্কিটহাউসেপৰ্যটকদেৰথাকাৰভালোসুবিধাআছে” (from Gyan Nidhi Corpus, TDIL). WordNet captures the semantic relation (MODIFIES_NOUN) between the synset {ভালো_32 (having desirable properties)} and {ক্রিয়া_150 (denoting an action or a state) } as well as the semantic relation (HYPONMY) between the synsets {সুবিধা_28213 (service)} and {ক্রিয়া_150}. This helps us in inferring the relationship between synsetsভালো_32 and সুবিধা_28213. The figure below shows an example of Semantic Graph Distance.



4. Sense Distribution across two languages within a domain:-Let us consider the example of two senses of the Nepali word फाइदा {faiidaa} and the corresponding cross linked words in Bengali.



Therefore if we have a sense tagged document in a resource rich language, we can infer sense distribution in a resource poor language