

Cross-Lingual Word Sense Disambiguation using Wordnets and Context based Mapping

Prabhat Pandey

Rahul Arora

{prabhatp,arorar}@cse.iitk.ac.in

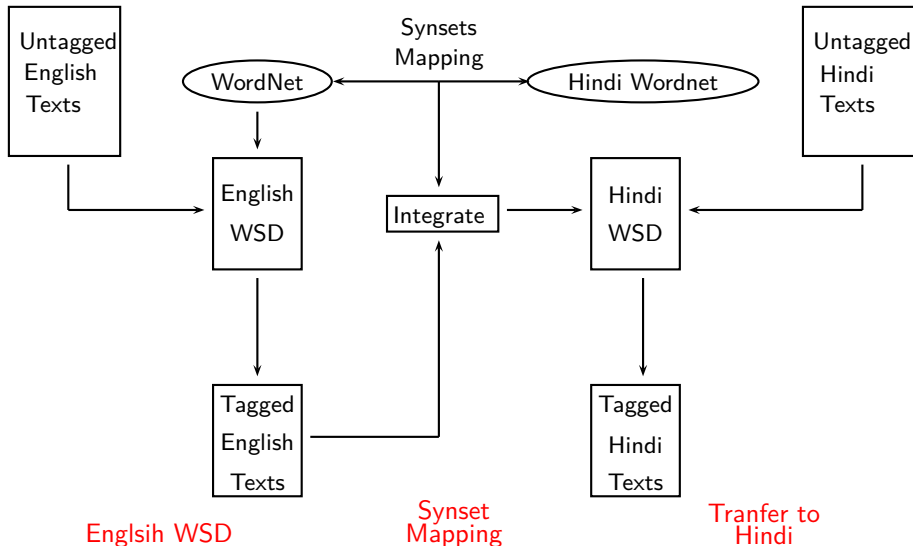
Department of Computer Science and Engineering
IIT Kanpur

Advisor : Prof. Amitabha Mukerjee

April 3, 2012

Importance

- Cross-Lingual WSD has applications in **Question Answering, Machine Translation, Information Retrieval**
- Lack of resources for languages like Hindi.
- Can be used to create sense-tagged corpus.



English WSD

- Used **WordNet::SenseRelate::AllWords** for English WSD
- Uses **Lesk Algorithm** for disambiguating senses
- Got sense tagged texts with sense number

e.g. -

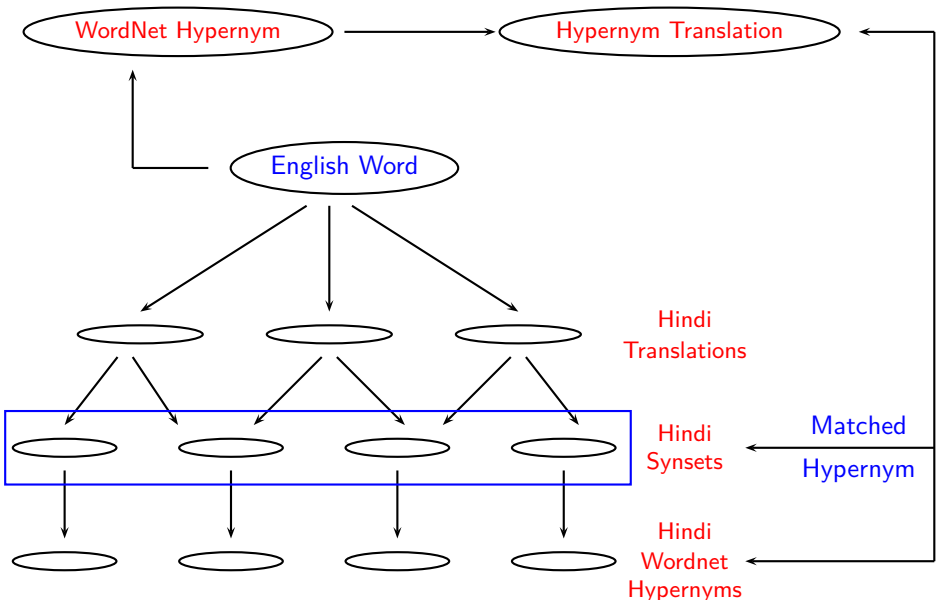
tax#n#1

[tax, taxation, revenue enhancement]

Cross-Lingual Word Sense Disambiguation using Wordnets and Context based Mapping

└ Our Approach & Results

└ Synset Mapping



Transfer to Hindi

- Search each words of Hindi synset in comparable corpus.
- Tag same sense as that of English label.

e.g.-

tax#n#1

कर[tax#n#1]#4107

Synset [2] 4107 - NOUN - [कर, टैक्स, महसूल, शालिक]

Transfer to Hindi

Three possible scenarios-

- 1 English word is polysemous while equivalent Hindi word is monosemous
e.g. - Bank बैंक
Synset obtained- [बैंक]
- 2 Both English and equivalent Hindi word are polysemous
e.g. - Branch शाखा
Synset obtained- [शाखा, डाल, डाली, टेरा, साख्र]
- 3 English Word is monosemous while equivalent Hindi word is polysemous
e.g. - Mango आम
Synset obtained- [आम, रसाल, आम्र, अम्ब, अंब, प्रियाम्बु]

Limitations and Flaws

- Restricted to **nouns** only.
- A better synset mapping resource may produce good results.
- Morphology handling.

References



Bahareh Sarrafzadeh, Nikolay Yakovets, Nick Cercone and Aijun An
Cross-lingual word sense disambiguation for languages with scarce resources
In Canadian Conference on AI'11, pages 347–358, 2011.



Dipak Narayan, Debasri Chakrabarty, Prabhakar Pande and Pushpak Bhattacharyya
An experience in building the Indo-WordNet – A WordNet for Hindi
In Proceedings of the First International Conference on Global WordNet (GWC 02), Mysore, India, 2002.



G. Miller, R. Beckwith, C. Fellbaum, D. Gross and K. Miller
Introduction to WordNet: An On-line Lexical Database.
International Journal of Lexicography Vol. 3, No. 4, pages 235-244, 1990.



J. Ramanand, Akshay Ukey, Brahm Kiran Singh and Pushpak Bhattacharyya
Mapping and Structural Analysis of Multi-lingual Wordnets
IEEE Data Engineering Bulletin, pages 30-44, 30(1), 2007.



Michael Lesk
Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone
In SIGDOC '86: Proceedings of the 5th annual international conference on Systems documentation, pages 24–26, New York, NY, USA, 1986.



Ted Pederson and Varada Kolhatkar
WordNet:SenseRelate:AllWords: A broad coverage Word Sense Tagger that maximizes semantic relatedness
In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Demonstration Session, (2009)