

Expanding a Lexicon to be Multi-Lingual Using Cognates

Bryan Wilkinson

Problem

How to make a static knowledge resource multi-lingual using automatic or semi-automatic techniques

Linguistic Background

- Historical Linguistics
- Word Relations
 - Cognates
 - Loan-Words

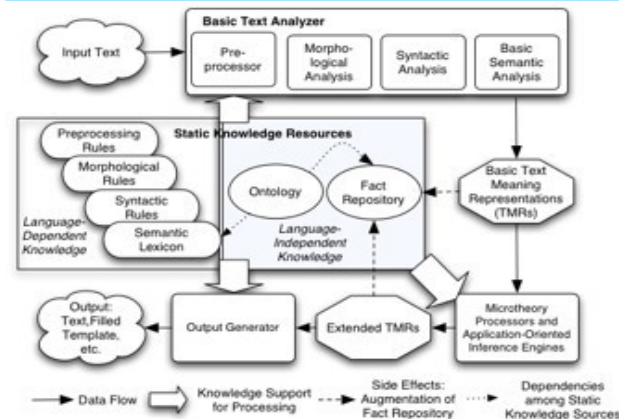
Latin	campus
Spanish	campo
Italian	campo
French	champ

Related Work

- Various Merging Algorithms
 - BiFrameNet
 - Padó's FrameNet Work

OntoSem

- Text Processing System
- Ontology (Language Independent)
- Lexicon (Language Dependent)



Cognate Recognition

- Phonetic Based
- Orthographically Based

Cysouw's cognate alignment

Rappoport and Levent-Levi's affix and letter correspondence

photo	foto
vibration	vibracion
tourism	tourismo

Bootstrapping The Lexicon with Cognates

- Associate cognates with same concept as word in L1
- Use syntactic information such as part of speech to get arguments for word
- Use lexical restrictions found in ontology to associate concepts for new words

Evaluation

- For cognate identification use precision and recall
- For Bootstrapping, look only at precision

Future Work

- Expand to use loan words
- Continue looking for better cognate recognition