MorphAdorner Futures

Possible Future Facilities

- Key Phrase Extraction
- Text Summarization
- Semantic categorization
- Word sense disambiguation
- Sentiment assignment
- Web services
- Other languages besides English

Key Phrase Extraction

- Extract "interesting" phrases which characterize a text
- Find the most likely categories for the text (semantic categories, using, e.g., Roget categories)
- Locate words that contribute to forming these categories and compute a weight based on relevance of the categories
- Apportion weights to surrounding words
- Phrases with high weightings are the interesting/key phrases
- LocalMaxs algorithm

Text Summarization

- Summarize contents for a paragraph, section, chapter, etc.
- Find key phrases
- Replace "high-falutin" words with simpler synonyms
- Perform anaphora (dangling pronoun reference) resolution for proper names

Semantic Categorization

- Categorize text according to topic and semantic notion
- We assign (old) Roget's Thesaurus categories to lemmata which provides possible topical categories for each word
- Allows comparison of texts written in different languages
- Need word sense disambiguation to distinguish different uses of a word

Word Sense Disambiguation

- A "bank" can be both a financial institution or a geographic location next to a river
- Need to disambiguate particular occurence to allow searches and comparisons to refer same type of bank in different contexts

Sentiment Assignment

- Training-based methods as in Nora and Monk
- Automatic methods based on scoring text extracts for words conveying dimensions of sentimentality. Examples: positive -> negative, active -> passive, and angry -> friendly, sacred -> secular
- Semantic categorization with word sense disambiguation will allow (we hope) for automatic sentiment assignment to text sections

Web Services

- Provide MorphAdorner facilities as web services
- WSRP, Hessian/Burlap
- Portlet versions of existing servlets
- "Drop Box" adornment

Other Languages

- Extend NUPOS to other languages than English
- Use existing part of speech tag sets
- Create training data, lexicons
- Create tokenizers, sentence splitters
- Create lemmatizers, standardizers
- Use different tagger algorithms