



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

**Any darn fool can make
something complex; it
takes a genius to make
something simple.**

— Albert Einstein

Linguistic
Background

Words, words, words,
...



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Preliminaries

- What's a **Morpheme**?
 - a morpheme is the smallest linguistic unit that has semantic meaning.
 - In spoken language, morphemes are composed of phonemes (the smallest linguistically distinctive units of sound), and in written language morphemes are composed of graphemes (the smallest units of written language).



Types of morphemes

- **Free morphemes** like *town*, and *dog* can appear with other **lexemes** (as in *town hall* or *dog house*) or they can stand alone, i.e. "free".
- **Bound morphemes** like "un-" appear only together with other morphemes to form a **lexeme**. Bound morphemes in general tend to be prefixes and suffixes. Unproductive, non-affix morphemes that exist only in bound form are known as "**cranberry**" morphemes, from the "cran" in that very word.
- **Derivational** morphemes can be added to a word to create (derive) another word: the addition of "-ness" to "happy," for example, to give "happiness." They carry **semantic** information.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Types of morphemes

- **Inflectional** morphemes modify a word's tense, number, aspect, and so on, without deriving a new word or a word in a new grammatical category (as in the "dog" morpheme if written with the plural marker morpheme "-s" becomes "dogs"). They carry **grammatical** information.
- **Allomorphs** are variants of a morpheme, e.g. the plural marker in English is sometimes realized as /-z/, /-s/ or /-ɪz/.



Other variants

- A **null morpheme** is a morpheme that is realized by a **phonologically** null **affix** (an empty string of phonological segments). In simpler terms, a null morpheme is an "invisible" affix. It's also called zero morpheme; the process of adding a null morpheme is called *null affixation*, *null derivation* or *zero derivation*.
- The **root** is the primary lexical unit of a word, which carries the most significant aspects of semantic content and cannot be reduced into smaller constituents. **Content words** in nearly all languages contain, and may consist only of, root morphemes. However, sometimes the term "root" is also used to describe the word minus its **inflectional** endings, but with its lexical endings in place. For example, *chatters* has the inflectional root or lemma *chatter*, but the lexical root *chat*. Inflectional roots are often called **stems**.



Other variants

- **Stems** may be roots, e.g. *run*, or they may be morphologically complex, as in **compound words** (cf. the compound nouns *meat ball* or *bottle opener*) or words with derivational morphemes (cf. the derived verbs *black-en* or *standard-ize*). Thus, the stem of the complex English noun *photographer* is *photo·graph·er*, but not *photo*. For another example, the root of the English verb form *destabilized* is *stabil-*, a form of *stable* that does not occur alone; the stem is *de·stabil·ize*, which includes the derivational affixes *de-* and *-ize*, but not the inflectional past tense suffix *-(e)d*. That is, a stem is that part of a word that inflectional affixes attach to.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Morphological analysis

- In natural language processing for Japanese, Chinese and other languages, morphological analysis is a process of segmenting given sentence into a row of morphemes. It is closely related to [Part-of-speech tagging](#), but word segmentation is required for these languages because word boundaries are not indicated by blank spaces. Famous Japanese morphological analysers include Juman, ChaSen and Mecab.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Words, words, words

- What's a word?
 - Definitions we will use over and over: Types, tokens, stems, roots, inflected forms, etc...
 - **Lexeme**: An entry in a lexicon consisting of a pairing of a form with a single meaning representation
 - **Lexicon**: A collection of lexemes



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Relationships between word meanings

- Homonymy
- Polysemy
- Synonymy
- Antonymy
- Hypernymy
- Hyponymy
- Meronymy



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Homonymy:

- Lexemes that share a form
 - Phonological, orthographic or both
- But have unrelated, distinct meanings
- Clear example:
 - Bat (wooden stick-like thing) vs
 - Bat (flying scary mammal thing)
 - Or bank (financial institution) versus bank (riverside)
- Can be homophones, homographs, or both:
 - Homophones:
 - Write and right
 - Piece and peace



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Homonymy causes problems for NLP applications

- Text-to-Speech
 - Same orthographic form but different phonological form
 - *bass* vs *bass*
- Information retrieval
 - Different meanings same orthographic form
 - QUERY: *bat care*
- Machine Translation
- Speech recognition
 - *Why?*



Polysemy

- The *bank* is constructed from red brick
I withdrew the money from the *bank*
- Are those the same sense?
- Or consider the following example
 - While some *banks* furnish sperm only to married women, others are less restrictive
 - Which sense of *bank* is this?
 - distinct from (homonymous with) river *bank* sense?
 - How about the savings *bank* sense?



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Polysemy

- A single lexeme with multiple *related* meanings (bank the building, bank the financial institution)
- Most non-rare words have multiple meanings
 - The number of meanings is related to its frequency
 - Verbs tend more to polysemy
 - Distinguishing polysemy from homonymy isn't always easy (or necessary)



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Metaphor and Metonymy

- Specific types of polysemy
- Metaphor:
 - Germany will pull Slovenia out of its economic slump.
 - I spent 2 hours on that homework.
- Metonymy
 - The White House announced yesterday.
 - This chapter talks about part-of-speech tagging
 - Bank (building) and bank (financial institution)



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Parts of Speech Table

part of speech	function or "job"	example words	example sentences
Verb	action or state	(to) be, have, do, like, work, sing, can, must	EnglishClub.com is a web site. I like EnglishClub.com.
Noun	thing or person	pen, dog, work, music, town, London, teacher, John	This is my dog . He lives in my house . We live in London .
Adjective	describes a noun	a/an, the, 69, some, good, big, red, well, interesting	My dog is big . I like big dogs.
Adverb	describes a verb, adjective or adverb	quickly, silently, well, badly, very, really	My dog eats quickly . When he is very hungry, he eats really quickly.
Pronoun	replaces a noun	I, you, he, she, some	Tara is Indian. She is beautiful.
Preposition	links a noun to another word	to, at, after, on, but	We went to school on Monday.
Conjunction	joins clauses or sentences or words	and, but, when	I like dogs and I like cats. I like cats and dogs. I like dogs but I don't like cats.
Interjection	short exclamation, sometimes inserted into a sentence	oh!, ouch!, hi!, well	Ouch! That hurts! Hi! How are you? Well , I don't know.



Words with More than One Job

- Many words in English can have more than one job, or be more than one part of speech. For example, "**work**" can be a verb and a noun; "**but**" can be a conjunction and a preposition; "**well**" can be an adjective, an adverb and an interjection. In addition, many nouns can act as adjectives.
- To analyze the part of speech, ask yourself: "What **job** is this word doing in this sentence?"
- In the table below you can see a few examples. Of course, there are more, even for some of the words in the table. In fact, if you look in a good dictionary you will see that the word **but** has six jobs to do:

verb, noun, adverb, pronoun, preposition and conjunction!



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Words with More than One Job

word	part of speech	example
work	noun	My work is easy.
	verb	I work in London.
but	conjunction	John came but Mary didn't come.
	preposition	Everyone came but Mary.
well	adjective	Are you well ?
	adverb	She speaks well .
	interjection	Well! That's expensive!
afternoon	noun	We ate in the afternoon .
	noun acting as adjective	We had afternoon tea.



Part-of-speech tagging

- In corpus linguistics, *part-of-speech tagging* (POS tagging or POST), also called grammatical tagging or word category disambiguation, is the process of marking up the words in a text (corpus) as corresponding to a particular *part of speech*, based on both its definition, as well as its context —i.e., relationship with adjacent and related words in a phrase, sentence, or paragraph. A simplified form is the identification of words as nouns, verbs, adjectives, etc.
- Once performed by hand, POS tagging is now done in computational linguistics using algorithms which associate discrete terms, as well as hidden parts of speech, in accordance with a set of descriptive tags.



Principle

- Part-of-speech tagging is harder than just having a list of words and their parts of speech, because some words can represent more than one part of speech at different times. This is not rare—in natural languages (as opposed to many artificial languages), a large percentage of word-forms are ambiguous. For example, even "dogs", which is usually thought of as just a plural noun, can also be a verb:

The sailor dogs the hatch.

- "Dogged", on the other hand, can be either an adjective or a past-tense verb. Just which parts of speech a word can represent varies greatly.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

History

- Research on part-of-speech tagging has been closely tied to corpus linguistics. The first major corpus of English for computer analysis was the **Brown Corpus** developed at Brown University by Henry Kucera and Nelson Francis, in the mid-1960s. It consists of about 1,000,000 words of running English prose text, made up of 500 samples from randomly chosen publications. Each sample is 2,000 or more words (ending at the first sentence-end after 2,000 words, so that the corpus contains only complete sentences).



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Note

It is worth remembering, as Eugene Charniak points out in *Statistical techniques for natural language parsing*, that merely assigning the most common tag to each known word and the tag "proper noun" to all unknowns, will approach 90% POST accuracy because many words are unambiguous.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Major algorithms

Some current major algorithms for part-of-speech tagging include the [Viterbi algorithm](#), [Brill Tagger](#), [Constraint Grammar](#), and the [Baum-Welch algorithm](#) (also known as the forward-backward algorithm). Hidden Markov model and visible Markov model taggers can both be implemented using the [Viterbi algorithm](#).



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Some references

1. Charniak, Eugene. 1997. "Statistical Techniques for Natural Language Parsing". *AI Magazine* 18(4):33–44.
2. Hans van Halteren, Jakub Zavrel, Walter Daelemans. 2001. Improving Accuracy in NLP Through Combination of Machine Learning Systems. *Computational Linguistics*. 27(2): 199–229.
3. DeRose, Steven J. 1990. "Stochastic Methods for Resolution of Grammatical Category Ambiguity in Inflected and Uninflected Languages." Ph.D. Dissertation. Providence, RI: Brown University Department of Cognitive and Linguistic Sciences.
4. DeRose, Steven J. 1988. "Grammatical category disambiguation by statistical optimization." *Computational Linguistics* 14(1): 31–39.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

How do we know when a word has more than one sense?

- ATIS examples
 - Which flights serve breakfast?
 - Does America West serve Philadelphia?
- The “zeugma” test:
 - ?Does United serve breakfast and San Jose?



Synonyms

- Word that have the same meaning in some or all contexts.
 - filbert / hazelnut
 - couch / sofa
 - big / large
 - automobile / car
 - vomit / throw up
 - Water / H₂O
- Two lexemes are synonyms if they can be successfully substituted for each other in all situations
 - If so they have the same **propositional meaning**



Synonyms

- But there are few (or no) examples of perfect synonymy.
 - Why should that be?
 - Even if many aspects of meaning are identical
 - Still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.
- Example:
 - Water and H₂O



More Terminology

Lemmas and wordforms

- A *lexeme* is an abstract pairing of meaning and form
- A *lemma* or *citation form* is the grammatical form that is used to represent a *lexeme*.
 - *Carpet* is the lemma for *carpets*
 - *Dormir* is the lemma for *duermes*.
- Specific surface forms *carpets*, *sung*, *duermes* are called *wordforms*
- The lemma *bank* has two *senses*:
 - Instead, a *bank* can hold the investments in a custodial account in the client's name
 - But as agriculture burgeons on the east *bank*, the river will shrink even more.
- A *sense* is a discrete representation of one aspect of the meaning of a word



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Synonymy is a relation between senses rather than words

- Consider the words *big* and *large*
- Are they synonyms?
 - How *big* is that plane?
 - Would I be flying on a *large* or small plane?
- How about here:
 - Miss Nelson, for instance, became a kind of *big* sister to Benjamin.
 - ?Miss Nelson, for instance, became a kind of *large* sister to Benjamin.
- Why?
 - *big* has a sense that means being older, or grown up
 - *large* lacks this sense



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Antonyms

- Senses that are opposites with respect to one feature of their meaning
- Otherwise, they are very similar!
 - dark / light
 - short / long
 - hot / cold
 - up / down
 - in / out
- More formally: antonyms can
 - define a binary opposition or at opposite ends of a scale (*long/short, fast/slow*)
 - Be **reversives**: *rise/fall, up/down*



Hyponymy

- One sense is a **hyponym** of another if the first sense is more specific, denoting a subclass of the other
 - *car* is a hyponym of *vehicle*
 - *dog* is a hyponym of *animal*
 - *mango* is a hyponym of *fruit*
- Conversely
 - *vehicle* is a hypernym/superordinate of *car*
 - *animal* is a hypernym of *dog*
 - *fruit* is a hypernym of *mango*

superordinate	vehicle	fruit	furniture	mammal
hyponym	car	mango	chair	dog



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Hypernymy more formally

- Extensional:
 - The class denoted by the superordinate
 - extensionally includes the class denoted by the hyponym
- Entailment:
 - A sense A is a hyponym of sense B if being an A entails being a B
- Hyponymy is usually transitive
 - (A hypo B and B hypo C entails A hypo C)



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

WordNet

- *A hierarchically organized lexical database*
- *On-line thesaurus + aspects of a dictionary*
 - *Versions for other languages are under development*



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

WordNet

- Where it is:
 - <http://www.cogsci.princeton.edu/cgi-bin/webwn>



Format of Wordnet Entries

The noun “bass” has 8 senses in WordNet.

1. bass¹ - (the lowest part of the musical range)
2. bass², bass part¹ - (the lowest part in polyphonic music)
3. bass³, basso¹ - (an adult male singer with the lowest voice)
4. sea bass¹, bass⁴ - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass¹, bass⁵ - (any of various North American freshwater fish with lean flesh (especially of the genus *Micropterus*))
6. bass⁶, bass voice¹, basso² - (the lowest adult male singing voice)
7. bass⁷ - (the member with the lowest range of a family of musical instruments)
8. bass⁸ - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective “bass” has 1 sense in WordNet.

1. bass¹, deep⁶ - (having or denoting a low vocal or instrumental range)
*”a deep voice” ; ”a bass voice is lower than a baritone voice” ;
”a bass clarinet”*



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

WordNet Noun Relations

Relation	Also called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	<i>breakfast</i> ¹ → <i>meal</i> ¹
Hyponym	Subordinate	From concepts to subtypes	<i>meal</i> ¹ → <i>lunch</i> ¹
Member Meronym	Has-Member	From groups to their members	<i>faculty</i> ² → <i>professor</i> ¹
Has-Instance		From concepts to instances of the concept	<i>composer</i> ¹ → <i>Bach</i> ¹
Instance		From instances to their concepts	<i>Austen</i> ¹ → <i>author</i> ¹
Member Holonym	Member-Of	From members to their groups	<i>copilot</i> ¹ → <i>crew</i> ¹
Part Meronym	Has-Part	From wholes to parts	<i>table</i> ² → <i>leg</i> ³
Part Holonym	Part-Of	From parts to wholes	<i>course</i> ⁷ → <i>meal</i> ¹
Antonym		Opposites	<i>leader</i> ¹ → <i>follower</i> ¹



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

WordNet Verb Relations

Relation	Definition	Example
Hypernym	From events to superordinate events	<i>fly</i> ⁹ → <i>travel</i> ⁹
Troponym	From a verb (event) to a specific manner elaboration of that verb	<i>walk</i> ¹ → <i>stroll</i> ¹
Entails	From verbs (events) to the verbs (events) they entail	<i>snore</i> ¹ → <i>sleep</i> ¹
Antonym	Opposites	<i>increase</i> ¹ ↔ <i>decrease</i> ¹



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

WordNet Hierarchies

Sense 3

bass, basso --

(an adult male singer with the lowest voice)

=> singer, vocalist, vocalizer, vocaliser

=> musician, instrumentalist, player

=> performer, performing artist

=> entertainer

=> person, individual, someone...

=> organism, being

=> living thing, animate thing,

=> whole, unit

=> object, physical object

=> physical entity

=> entity

=> causal agent, cause, causal agency

=> physical entity

=> entity

Sense 7

bass --

(the member with the lowest range of a family of musical instruments)

=> musical instrument, instrument

=> device

=> instrumentality, instrumentation

=> artifact, artefact

=> whole, unit

=> object, physical object

=> physical entity

=> entity



How is “sense” defined in WordNet?

The set of near-synonyms for a WordNet sense is called a *synset* (*synonym set*); it’s their version of a sense or a concept

- Example: *chump* as a noun to mean

- ‘a person who is gullible and easy to take advantage of’

{*chump*¹, *fool*², *gull*¹, *mark*⁹, *patsy*¹, *fall guy*¹, *sucker*¹,
*soft touch*¹, *mug*²}

- Each of these senses share this same gloss
- Thus for WordNet, the meaning of this sense of *chump* is this list.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Word Sense Disambiguation (WSD)

- Given
 - a word in context,
 - A fixed inventory of potential word sense
- decide which is the sense of this word.
 - English-to-Spanish MT
 - Inventory is set of Spanish translations
 - Speech Synthesis
 - Inventory is homographs with different pronunciations like *bass* and *bow*
 - Automatic indexing of medical articles
 - MeSH (Medical Subject Headings) thesaurus entries



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Two variants of WSD task

- Lexical Sample task
 - Small pre-selected set of target words
 - And inventory of senses for each word
 - We'll use *supervised machine learning*
- All-words task
 - Every word in an entire text
 - A lexicon with senses for each word
 - Sort of like part-of-speech tagging
 - Except each lemma has its own tagset



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Supervised Machine Learning Approaches

- Supervised machine learning approach:
 - a **training corpus** of words tagged in context with their sense
 - used to train a classifier that can tag words in new text
 - Just as we saw for part-of-speech tagging, statistical MT.
- Summary of what we need:
 - the **tag set** (“sense inventory”)
 - the **training corpus**
 - A set of **features** extracted from the training corpus
 - A **classifier**



Word sense disambiguation sketch – not ML

- Consider a simple, informal, anything but robust knowledge-based approach to word sense disambiguation. We know that many English sentences can map onto a template that looks like the following:

<agent> <action> <actionable item>

which normally correspond to the *subject*, *verb* and *object* respectively. This simple observation leads to an intriguing method for disambiguating word senses. As an example consider the sentence “*The banker banks his plane over the river bank near the bank that he banks on for good banking service.*” – whew



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Word sense disambiguation sketch – not ML

What do we know about this somewhat contrived sentence? We know that “...*that he banks on for good banking service.*” is a relative clause and can be treated as a separate sentence. That’s good – and the same techniques we will now discuss for the “*The banker banks his plane over the river bank near the bank*” part will also serve well for the clause. We need to find the verb - <action>. Morphological analysis reveals that the content words *banker, banks, plane, river, bank* and *bank* can be as follows:



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Word sense disambiguation sketch – not ML

WORD	ROOT	MORPH-1	MORPH-2
Banker	Bank	Noun singular	Comparative adjective
Banks	Bank	Noun plural	Verb present tense
Plane	Plane	Noun singular	Verb present tense
River	River	Noun singular	
Bank	Bank	Noun singular	Verb present tense
Bank	Bank	Noun singular	Verb present tense



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Word sense disambiguation sketch – not ML

- The function words also provide us with useful information, thus over and near are prepositions and hence their phrases *over the river bank* and *near the bank* will need to be attached to the sentence structure (remember grammar school English classes and diagramming?). So now we need to find the verb in the fragment *The banker banks his plane*. Morphological analysis has revealed two candidates: *banks* and *plane*.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Word sense disambiguation sketch – not ML

- If we consider *banks* we find that *banker* is a compatible <agent> (a subject), that is *banker* has the right features (selectional restrictions) to be compatible with a verb <action> of *banks*. Also *plane* is an acceptable <actionable item> (an object). Thus the sense of *banks* is determined by the constraints imposed (selectional restrictions) that *banker* and *plane* impose.
- If we consider the alternative and choose the second candidate *plane* as the verb, we find that it does not work so well.



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

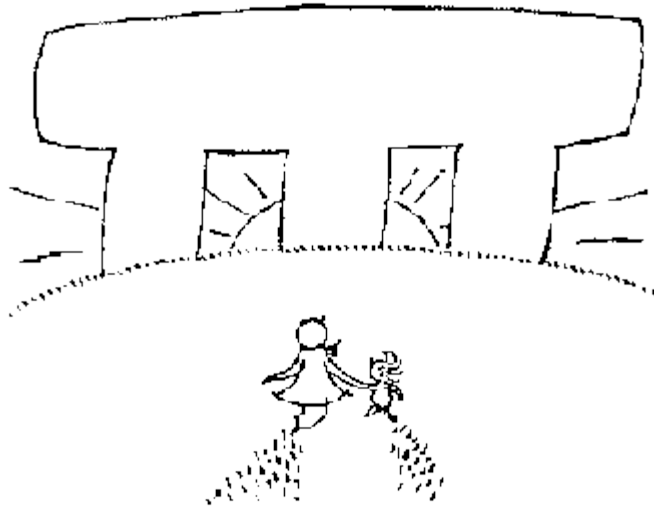
Conclusion

- Morphemes, words, ...
- Lexical Semantics
 - Homonymy, Polysemy, Synonymy
 - Thematic roles
- Computational resource for lexical semantics
 - WordNet
- Task
 - Word sense disambiguation



CSE6339 3.0 Introduction to Computational Linguistics
Instructor: Nick Cercone – 3050 LAS – nick@cse.yorku.ca
Mondays, Wednesdays 10:00-11:20 – North Ross 836A
Winter Semester, 2014

Other Concluding Remarks



T. T. T.

Put up in a place where it's easy to see
the cryptic admonishment T. T. T.

When you feel how depressingly slowly you climb,
it's well to remember that Things Take Time.