Development of a Word Sense Disambiguation System

Genna Cockburn, Dr. Greg Kondrak, Bradley Hauer, Amy Hua, Jacob Skitsko

Department of Computational Linguistics, University of Alberta

Supported by: Canada

Introduction

- A homonym is a word with two or more senses (meanings), where at least two of the senses come from different etymologies. Homonyms are the coarsest level of granularity (the scale or level of detail).
- Etymology is the origin of a word and the historical development of its meaning.
- Word Sense Disambiguation is the process of identifying the appropriate sense of a word with multiple senses.
- WordNet is a fine-grained online dictionary that contains every sense of a homonym.
- The Oxford English Dictionary contains a smaller range of senses, but also has etymologies of each sense.
- Navigli's clustering is an algorithm created to cluster similar WordNet senses.

Objective

To create a system that can identify the appropriate sense of a homonym based upon the word's context within a text.

Methods

Before development on a word sense disambiguation system can begin, the system needs a list of homonyms to work with.

1. To start with, queries (text files) of every Oxford sense were created.
2. These queries were then run through a program called homonym resource that organized and outputted raw data into a spreadsheet.
3. Manual annotations for etymology and synonyms were added to the spreadsheet as additional components for the system's identification of homonyms.
4. The spreadsheet was then run through another program to discard any words with a single etymology.
5. This process resulted in a list of homonyms.

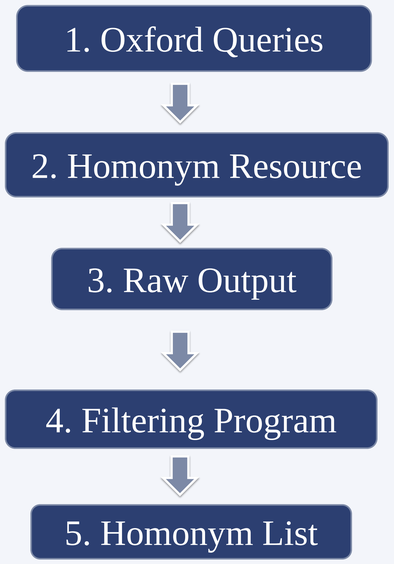

Figure 1: Visualization of the process of creating a homonym list.

Results

- With this method a list of 806 homonyms was created

	A	B	C	D	E	F	G	H	I
1	Number	English Wo	POS	Sense	Full Etymology	Abbreviatec	Etymological Wor	Synonym	French Word
2	1	affect	v	100	late Middle English (in the	L	affect	influence	affecter
3	1	affect	v	200	late Middle English: from	L	affectare	pretend	feindre
4	1	affect	n	300	late 19th century: coined	L	affectus	emotion	
5	2	agape	a	100	mid 17th century: from a-	E	a+gape	amazed	bouche bée
6	2	agape	n	200	early 17th century: from G	Gr	agapē	meal	
7	3	alight	v	100	Old English ālīhtan, from	OE	ā+līhtan	descend	
8	3	alight	a	200	late Middle English: proba	ME	a+light	shining	
9	4	alley	n	100	late Middle English: from	OF	alee	passageway	ruelle
10	4	alley	n	200	early 18th century: perhap	E	alabaster(SH)	marble	
11	5	ally	n,v	100	Middle English (as a verb	OF	alier	friend	allié
12	5	ally	n	200	VS alley, from early 18th c	E	alabaster(SH)	marble	
13	6	ancient	a,n	100	late Middle English: from	OF	ancien	old	allier
14	6	ancient	n	200	mid 16th century: alteratic	E	ensign	standard	
15	7	angle	n,v	100	late Middle English: from	L	angulus	intersection	angle
16	7	angle	v,n	200	Old English angul (noun);	OE	angul	fish	
17	7	angle	n	500	from Latin Anglus, (plural)	L	Anglus	German	
18	8	apostrophe	n	100	mid 16th century (denotin	Gr	apostrophos	punctuation	apostrophe
19	8	apostrophe	n	200	mid 16th century: via Latir	Gr	apostrophē	exclamation	apostrophe
20	9	arch	n,v	100	Middle English: from Old I	OF	arche	curve	arc
21	9	arch	a	200	mid 17th century: from arc	E	arch	playful	

Figure 2: A small sample from Homonym List: the list of 806 homonyms.

- WordNet is the standard inventory for word sense disambiguation, so we had to map the WordNet senses to our homonym classes from Oxford.
- To resolve this issue of mapping, the larger range of WordNet senses were mapped to a clustering of senses, nicknamed Navigli's clustering, which were then mapped to one of the broader Oxford senses.

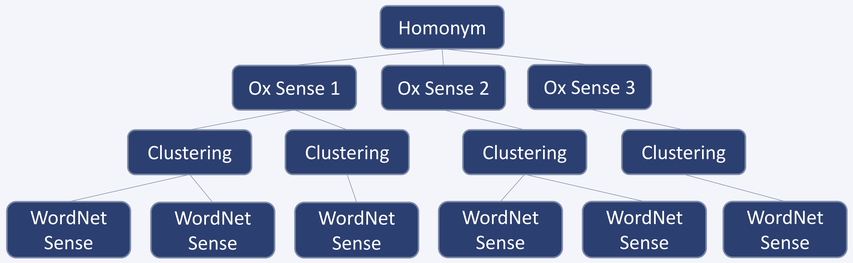

Figure 3: Visualization of Oxford to WordNet Mapping

	A	B	C	D	E	F	G	H
1	WORD	POS	WN SENSE #	WN KEY	OED HOM. #	GLOSS		
83	fuzz	Noun	fuzz#n#2	fuzz%1:18:00::	200	uncomplimentary terms for a policeman		
84	fuzz	Noun	fuzz#n#3	fuzz%1:09:00::	100	a hazy or indistinct representation		
85	fuzz	Noun	fuzz#n#4	fuzz%1:08:00::	100	the first beard of an adolescent boy		
86	gob	Noun	gob#n#1	gob%1:18:00::	200	a man who serves as a sailor		
87	gob	Noun	gob#n#2	gob%1:14:00::	100	a lump of slimy stuff		
88	gob	Noun	gob#n#3	gob%1:08:00::	300	informal terms for the mouth		
89	halt	Noun	halt#n#1	halt%1:26:00::	100	the state of inactivity following an interruption		
90	halt	Noun	halt#n#2	halt%1:11:00::	100	the event of something ending		
91	halt	Noun	halt#n#3	halt%1:04:00::	100	an interruption or temporary suspension of progress		
92	halt	Verb	halt#v#1	halt%2:38:05::	100	cause to stop		
93	halt	Verb	halt#v#2	halt%2:38:01::	100	come to a halt, stop moving		
94	halt	Verb	halt#v#3	halt%2:41:00::	100	stop from happening or developing		
95	halt	Verb	halt#v#4	halt%2:30:02::	100	stop the flow of a liquid		
96	halt	Adjective	halt#a#1	halt%5:00:00:unf	200	disabled in the feet or legs		

Figure 4: A small sample of the spreadsheet produced by Oxford to WordNet mapping

Conclusions

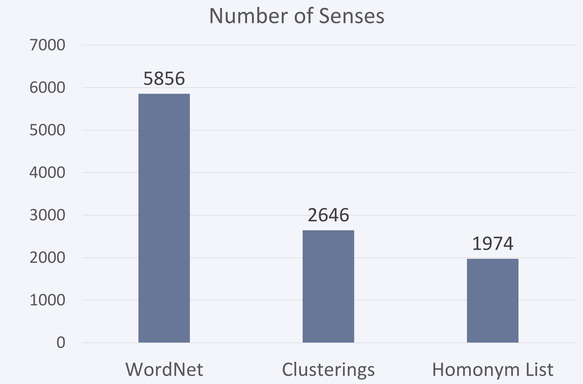

Figure 5: Bar graph illustrating the reduction in granularity level of senses

- As illustrated by figure 5, the overall creation and mapping of the homonym list resulted in a reduction in the granularity of WordNet senses to the coarsest linguistically feasible level.
- The additional components of the homonym list, such as synonyms, French translations and etymology, will be used to identify the appropriate meaning.
- One example of how the components will be used is with French translations. As seen in our data, word senses with distinct etymologies tend to translate differently across languages. This is illustrated in rows 2 and 3 of figure 2.
- By using this idea, we can look at the French translation of the sentence, and figure out what homonym class that instance belongs to through the differences in translation.
- Additional homonyms must be also added to the list as there are more than 806 homonyms within the English language.

Citations

- Data found on the spreadsheet within figure 2 was taken from the English online Oxford Dictionary.
- Data found on the spreadsheet within figure 3 was taken from WordNet online Dictionary.
- Roberto Navigli. "Meaningful Clustering of Senses Helps Boost Word Sense Disambiguation Performance" 2006. PDF file.
- Etymology | Definition of Etymology in English by Oxford Dictionaries." *Oxford Dictionaries | English*, Oxford Dictionaries, en.oxforddictionaries.com/definition/etymology.

Acknowledgements

The researcher would like to thank:

- Principle Investigator Dr. Greg Kondrak, Direct Supervisor Bradly Hauer, Amy Hua and Jacob Skitsko
- Sponsors Canada Summer Jobs and WISEST
- The WISEST Summer Research Program 2018