

# Using WordNet and Short-Term Memory for Contextual Disambiguation

William T. F. Strachan

Word Count - PLACEHOLDER  
Supervisor - Dr Dimitar Kazakov  
Department of Computer Science  
University of York  
20 November 2016

# Contents

<b>1</b>	<b>Literature Review</b>	<b>2</b>
1.1	Natural Language Processing . . . . .	2
1.1.1	Text Preprocessing . . . . .	2
1.1.2	Lexical Analysis . . . . .	2
1.1.3	Syntactic Parsing . . . . .	3
1.1.4	Semantic Analysis . . . . .	3
1.1.5	Latent Semantic Analysis . . . . .	3
1.2	Psycholinguistics . . . . .	3
1.2.1	Long-term Store . . . . .	3
1.2.2	Short-term Store . . . . .	4
1.3	Previous Work . . . . .	4

# 1 Literature Review

In order to define a problem, we must first establish the previous works, so as to build upon them effectively.

## 1.1 Natural Language Processing

The study of natural language processing aims to allow a computer to understand natural language, and formulate a relevant response based upon its input. Within this, the problem of input text analysis has traditionally be broken down into smaller sub-problems [1]:

- Text Preprocessing
- Lexical Analysis
- Syntactic Parsing
- Semantic Analysis

The subsequent subsections will discuss each of these in more detail.

### 1.1.1 Text Preprocessing

Before any analysis can take place, the inputted raw text must be converted into a usable format. This, once again, can be broken down into multiple steps [1]:

- **Document Triage**
  - Character encoding must be identified.
  - The language can then be identified.
  - Non-useful data, such as images and html formatting must be removed.
- **Text Segmentation**
  - Individual words (tokens) must be separated from one another.
  - Text Normalisation; replacing multiple equivalent tokens with one token (e.g. "Ave." and "Avenue").
  - Identifying sentences, i.e. locating where a sentence begins and ends.

### 1.1.2 Lexical Analysis

One word can have multiple forms, for example "judge" (the lemma) has the forms {"judge", "judges", "judging", "judged"} (morphological variants). The job of Lexical Analysis is to replace all morphological variants of a word, with their corresponding lemma, a process known as stemming [1].

### 1.1.3 Syntactic Parsing

When deriving meaning from sentences, the grammatical structure can provide important insight. The Syntactic parsing technique extracts this information using two processes [2]:

- **Part-of-speech Tagging**
  - Each word is given tags denoting their syntactic role (e.g. noun/adjective/verb).
- **Chunking**
  - Noun phrases and verb phrases are detected and tagged using "begin chunk" and "inside chunk" labels

### 1.1.4 Semantic Analysis

The semantics of a sentence, is the meaning given by its tokens [3]. The topic of semantic analysis will be expanded upon in the proceeding subsections.

### 1.1.5 Latent Semantic Analysis

## 1.2 Psycholinguistics

Language understanding is a problem which, it can be stated, is solved by the human brain. From this statement, it can be derived that a computational solution could be effectively built around knowledge of the processes at work in the brain. The process of language comprehension can be described using the working memory model [4].

According to Baddeley et al. [4] there exist multiple, special purpose, memory structures within two main categories, the Short-term Store and the Long-term Store.

### 1.2.1 Long-term Store

The long-term store (LTS) contains semi-permanent information. Within the LTS, there exist Explicit and Implicit memory structures. The contents of the Implicit memory describe skills and methods of doing things, whereas the Explicit memory contains factual information [4]. When considering these structures, it can be seen that Explicit memory is of greater interest in the context of NLP.

Within the Explicit memory exists knowledge of semantics [4]. The information held here not only defines concepts (meanings of word forms), but also their attributes and rules of use. In 1966, M. Quillian proposed a model of Semantic Memory [5]. The model consists of a graph of nodes, each representing a concept, connected by edges of differing types, each representing a different syntactic feature (for example, hypernym).

### 1.2.2 Short-term Store

The short-term store (STS) is a structure of limited capacity, used to store items for periods usually of no more than a few seconds [4]. In 1955, G. Miller, based upon previous experimental results, concluded that the size of the STS existed in the realm of  $7 \pm 2$  items of information [6].

In 1971, R. Atkinson and R. Shiffrin proposed a model of the STS [7]. In this model, the STS can both send information to, and draw information from the LTS. Inputs from the sensory registers (memory structures holding information relating to inputs from senses) are also sent to the STS. Atkinson and Shiffrin proposed that, over time, the activation of items in the STS decreased; they went on to theorise that items could be lost from the STS, only when a new, more highly activated item could take its place. To counter this loss of activation, the authors discussed the control process, rehearsal. This process makes use of repetition to increase the activation of items in memory, decreasing their chance of loss.

## 1.3 Previous Work

## References

- [1] F. Damerau. N. Indurkha, *Handbook of Natural Language Processing*, 2nd ed. Taylor and Francis Group, 2010.
- [2] R. Collobert. et al., “Natural language processing (almost) from scratch,” *Journal of Machine Learning Research*, vol. 12, no. no. 76, pp. 2493–2537, Aug 2011.
- [3] C. Goddard, *Semantic Analysis: A Practical Introduction*, 2nd ed. Oxford University Press, 2011.
- [4] M. Eysenck. A. Baddeley and M. Anderson, *Memory*, 2nd ed. Psychology Press, 2015.
- [5] M. Quillain, “Semantic memory,” *BBN Report*, vol. 675, no. no. 2, pp. 227–270, Oct 1966.
- [6] G. Miller, “The magical number seven, plus or minus two some limits on our capacity for processing information,” *Psychological Review*, vol. 101, no. no. 2, pp. 343–352, May 1955.
- [7] R. Shiffrin. R. Atkinson, “The control processes of short term memory,” *Technical Report*, vol. 173, Apr 1971.