

More than meets the eye: Study of Human Cognition in Sense Annotation

Salil Joshi

IBM Research India
Bangalore, India
saljoshi@in.ibm.com

Diptesh Kanojia

Gautam Buddha Technical University
Lucknow, India
dipteshkanojia@gmail.com

Pushpak Bhattacharyya

Computer Science and Engineering Department
Indian Institute of Technology, Bombay
Mumbai, India
pb@cse.iitb.ac.in

Abstract

Word Sense Disambiguation (WSD) approaches have reported good accuracies in recent years. However, these approaches can be classified as weak AI systems. According to the classical definition, a strong AI based WSD system should perform the task of sense disambiguation in the same manner and with similar accuracy as human beings. In order to accomplish this, a detailed understanding of the human techniques employed for sense disambiguation is necessary. Instead of building yet another WSD system that uses contextual evidence for sense disambiguation, as has been done before, we have taken a step back - we have endeavored to discover the cognitive faculties that lie at the very core of the human sense disambiguation technique.

In this paper, we present a hypothesis regarding the cognitive sub-processes involved in the task of WSD. We support our hypothesis using the experiments conducted through the means of an eye-tracking device. We also strive to find the levels of difficulties in annotating various classes of words, with senses. We believe, once such an in-depth analysis is performed, numerous insights can be gained to develop a robust WSD system that conforms to the principle of strong AI.

1 Introduction

Word Sense Disambiguation (WSD) is formally defined as the task of computationally identifying senses of a word *in a context*. The phrase ‘in a context’ is not defined explicitly in the literature. NLP researchers define it according to their convenience. In our current work, we strive to unravel

the appropriate meaning of contextual evidence used for the human annotation process. Chatterjee et al. (2012) showed that the contextual evidence is the predominant parameter for the human sense annotation process. They also state that WSD is successful as a *weak AI* system, and further analysis into human cognitive activities lying at the heart of sense annotation can aid in development of a WSD system built upon the principles of strong AI.

Knowledge based approaches, which can be considered to be closest form of WSD conforming to the principles of strong AI, typically achieve low accuracy. Recent developments in domain-specific knowledge based approaches have reported higher accuracies. A domain-specific approach due to Agirre et al. (2009) beats supervised WSD done in generic domains. Ponzetto and Navigli (2010) present a knowledge based approach which rivals the supervised approaches by using the semantic relations automatically extracted from Wikipedia. They reported approximately 7% gain over the closet supervised approach.

In this paper, we delve deep into the cognitive roles associated with sense disambiguation through the means of an eye-tracking device capturing the gaze patterns of lexicographers, during the annotation process. In-depth discussions with trained lexicographers indicate that there are multiple cognitive sub-processes driving the sense disambiguation task. The eye movement paths available from the screen recordings done during sense annotation conform to this theory.

Khapra et al. (2011) points out that the accuracy of various WSD algorithms is poor on certain

Part-of-speech (POS) categories, particularly, verbs. It is also a general observation for lexicographers involved in sense annotation that there are different levels of difficulties associated with various classes of words. This fact is also reflected in our analysis on sense annotation. The data available after the eye-tracking experiments gave us the fixation times and saccades pertaining to different classes of words from the analysis of this data we draw conclusive remarks regarding the reasons behind this phenomenon. In our case, we classified words based on their POS categories.

In this paper, we establish that contextual evidence is the prime parameter for the human annotation. Further, we probe into the implication of context used as a clue for sense disambiguation, and the manner of its usage. In this work, we address the following questions:

- *What are the cognitive sub-processes associated with the human sense annotation task?*
- *Which classes of words are more difficult to disambiguate and why?*

By providing relevant answers to these questions we intend to present a comprehensive understanding of sense annotation as a complex cognitive process and the factors involved in it. The remainder of this paper is organized as follows. Section 2 contains related work. In section 3 we present the experimental setup. Section 4 displays the results. We summarize our findings in section 5. Finally, we conclude the paper in section 6 presenting the future work.

2 Related Work

As mentioned earlier, we used the eye-tracking device to ascertain the fact that contextual evidence is the prime parameter for human sense annotation as quoted by Chatterjee et al. (2012) who used different annotation scenarios to compare human and machine annotation processes. An eye movement experiment was conducted by Vainio et al. (2009) to examine effects of local lexical predictability on fixation durations and fixation locations during sentence reading. Their study indicates that local lexical predictability influences in decisions but not where the initial fixation lands in a word. In another work based on word grouping hypothesis and eye

movements during reading by Drieghe et al. (2008), the distribution of landing positions and durations of first fixations in a region containing a noun preceded by either an article or a high-frequency three-letter word were compared.

Recently, some work is done on the study of sense annotation. A study of sense annotations done on 10 polysemous words was conducted by Passonneau et al. (2010). They opined that the word meanings, contexts of use, and individual differences among annotators gives rise to inter-annotation variations. De Melo et al. (2012) present a study with a focus on MASC (Manually-Annotated SubCorpus) project, involving annotations done using using WordNet sense identifiers as well as FrameNet lexical units.

In our current work we use eye-tracking as a tool to make findings regarding the cognitive processes connected to the human sense disambiguation procedure, and to gain a better understanding of “contextual evidence” which is of paramount importance for human annotation. Unfortunately, our work seems to be a first of its kind, and to the best of our knowledge we do not know of any such work done before in the literature.

3 Experimental Setup

We used a generic domain (*viz.*, News) corpus in Hindi language for experimental purposes. To identify the levels of difficulties associated with human annotation, across various POS categories, we conducted experiments on around 2000 words (including function words and stop words). The analysis was done only for open class words. The statistics pertaining to the our experiment are illustrated in table 1. For statistical significance of our experiments, we collected the data with the help of 3 skilled lexicographers and 3 unskilled lexicographers.

POS	Noun	Verb	Adjective	Adverb
 #(senses)	2.423	3.814	2.602	3.723
 #(tokens)	452	206	96	177

Table 1: Number of words (tokens) and average degree of corpus polysemy (senses) of words per POS category (taken from Hindi News domain) used for experiments

For our experiments we used a Sense Annotation

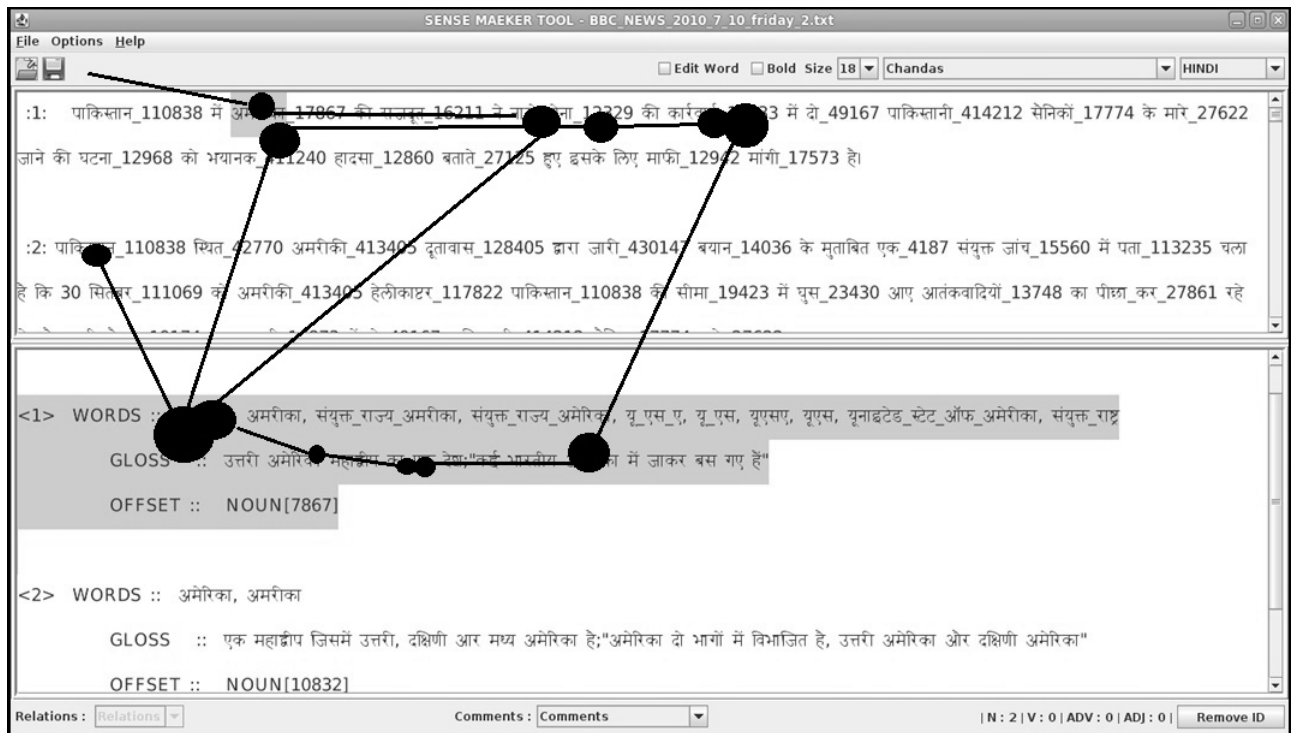


Figure 1: Sense marker tool showing an example Hindi sentence in the *Context Window* and the wordnet synsets of the highlighted word in the *Synset Window* with the black dots and lines indicating the scan path

Tool, designed at IIT Bombay and an eye-tracking device. The details of the tools and their purposes are explained below:

3.1 The Sense Marker Tool

A word may have a number of senses, and the task of identifying and marking which particular sense has been used in the given context, is known as *sense marking*.

The Sense Marker tool¹ is a Graphical User Interface based tool developed using Java, which facilitates the task of manual sense marking. This tool displays the senses of the word as available in the Marathi, Hindi and Princeton (English) WordNets and allows the user to select the correct sense of the word from the candidate senses.

3.2 Eye-Tracking device

An eye tracker is a device for measuring eye positions and eye movement. A *saccade* denotes move-

¹<http://www.cse.iitb.ac.in/~salilj/resources/SenseMarker/SenseMarkerTool.zip>

ment to another position. The resulting series of fixations and saccades is called a *scan path*. Figure 1 shows a sample scan path. In our experiments, we have used an eye tracking device manufactured by SensoMotoric Instruments². We recorded saccades, fixations, length of each fixation and scan paths on the stimulus monitor during the annotation process. A remote eye-tracking device (RED) measures gaze hotspots on a stimulus monitor.

4 Results

In our experiments, each lexicographer performed sense annotation on the stimulus monitor of the eye tracking device. Fixation times, saccades and scan paths were recorded during the sense annotation process. We analyzed this data and the corresponding observations are enumerated below.

Figure 2 shows the annotation time taken by different lexicographers across POS categories. It can be observed that the time taken for disambiguating the verbs is significantly higher than the remaining POS

²<http://www.smivision.com/>

Word	Degree of polysemy	Unskilled Lexicographer (Seconds)				Skilled Lexicographer (Seconds)			
		T_{hypo}	T_{clue}	T_{gloss}	T_{total}	T_{hypo}	T_{clue}	T_{gloss}	T_{total}
लाना (laanaa - to bring)	4	0.63	0.80	5.20	6.63	0.31	1.20	1.82	3.30
करना (karanaa - to do)	22	0.90	1.42	2.20	4.53	0.50	0.64	1.14	2.24
जताना (jataanaa - to express)	4	0.70	2.45	5.93	9.09	0.25	0.39	0.62	1.19

Table 2: Comparison of time taken across different cognitive stages of sense annotation by lexicographers for verbs

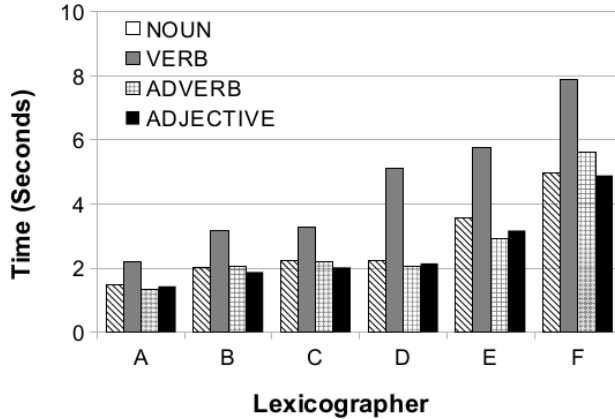


Figure 2: Histogram showing time taken (in seconds) by each lexicographer across POS categories for sense annotation

categories. This behavior can be consistently seen in the timings recorded for all the six lexicographers.

Table 2 presents the comparison of time taken across different cognitive stages of sense annotation by lexicographers for some of the most frequently occurring verbs.

To know if the results gathered from all the lexicographers are consistent, we present the correlation between each pair of lexicographers in table 3. The table also shows the value of the t-test statistic generated for each pair of lexicographers.

5 Discussion

The data obtained from the eye-tracking device and corresponding analysis of the fixation times, saccades and scan paths of the lexicographers' eyes reveal that sense annotation is a complex cognitive process. From the videos of the scan paths obtained from the eye-tracking device and from detailed discussion with lexicographers it can be inferred that

this cognitive process can be broken down into 3 stages:

1. When a lexicographer sees a word, he/she makes a hypothesis about the domain and consequently about the correct sense of the word, mentally. In cases of highly polysemous words, the hypothesis may narrow down to multiple senses. We denote the time required for this phase as T_{hypo} .
2. Next the lexicographer searches for clues to support this hypothesis and in some cases to eliminate false hypotheses, when the word is polysemous. These clues are available in the form of neighboring words around the target word. We denote the time required for this activity as T_{clue} .
3. The clue words aid the lexicographer to decide which one of the initial hypotheses was true. To narrow down the candidate synsets, the lexicographers use synonyms of the words in a synset to check if the sentence retains its meaning.

From the scan paths and fixation times obtained from the eye-tracking experiment, it is evident that stages 1, 2 and 3 are chronological stages in the human cognitive process associated with sense disambiguation. In cases of highly polysemous words and instances where senses are fine-grained, stages 2 and 3 get interleaved. It is also clear that each stage takes up separate proportions of the sense disambiguation time for humans. Hence time taken to disambiguate a word using the Sense Marker Tool (as explained in Section 3.1) can be factored as follows:

$$T_{total} = T_{hypo} + T_{clue} + T_{gloss}$$

Where:

T_{total} = Total time for sense disambiguation

Lexicographer	Correlation value					T-test statistic				
	B	C	D	E	F	B	C	D	E	F
A	0.933	0.976	0.996	0.996	0.769	0.007	0.123	0.185	0.036	0.006
B		0.987	0.960	0.915	0.945		0.009	0.028	0.084	0.026
C			0.989	0.968	0.879			0.483	0.088	0.067
D				0.988	0.820				0.367	0.709
E					0.734					0.418

Table 3: Pairwise correlation between annotation time taken by lexicographers

T_{hypo} = Time for hypothesis building

T_{clue} = Clue word searching time

T_{gloss} = Gloss Matching time and winner sense selection time.

The results in table 2 reveal the different ratios of time invested during each of the above stages. T_{hypo} takes the minimum amount of time among the different sub-processes. $T_{gloss} > T_{clue}$ in all cases.

- For unskilled lexicographers: $T_{gloss} \gg T_{clue}$ because of errors in the initial hypothesis.
- For skilled lexicographers: $T_{gloss} \sim T_{clue}$, as they can identify the POS category of the word and their hypothesis thus formed is pruned. Hence during selection of the winner sense, they do not browse through other POS categories, which unskilled lexicographers do.

The results shown in figure 2 reveal that verbs take the maximum disambiguation time. In fact the average time taken by verbs is around 75% more than the time taken by other POS categories. This supports the fact that verbs are the most difficult to disambiguate.

The analysis of the scan paths and fixation times available from the eye-tracking experiments in case of verbs show that the T_{gloss} covers around 66% of T_{total} , as shown in table 2. This means that the lexicographer takes more time in selecting a winner sense from the list of wordnet senses. This happens chiefly because of following reasons:

1. Higher degree of polysemy of verbs compared to other POS categories (as shown in tables 1 and 2).
2. In several cases the senses are fine-grained.

3. Sometimes the hypothesis of the lexicographers may not match any of the wordnet senses. The lexicographer then selects the wordnet sense closest to their hypothesis.

Adverbs and adjectives show higher degree of polysemy than nouns (as shown in table 1), but take similar disambiguation time as nouns (as shown in figure 2). In case of adverbs and adjectives, the lexicographer is helped by their position around a verb or noun respectively. So, T_{clue} only involves searching for the nearby verbs or nouns, as the case may be, hence reducing total disambiguation time T_{total} .

6 Conclusion and Future Work

In this paper we examined the cognitive process that enables the human sense disambiguation task. We have also laid down our findings regarding the varying levels of difficulty in sense annotation across different POS categories. These experiments are just a stepping stone for going deeper into finding the meaning and manner of usage of contextual evidence which is fundamental to the human sense annotation process.

In the future we aim to perform an in-depth analysis of clue words that aid humans in sense disambiguation. The distance of clue words from the target word and their and pattern of occurrence could give us significant insights into building a ‘Discrimination Net’.

References

- E. Agirre, O.L. De Lacalle, A. Soroa, and I. Farkutatea. 2009. Knowledge-based wsd on specific domains: performing better than generic supervised wsd. *Proceedings of IJCAI*, pages 1501–1506.
- Arindam Chatterjee, Salil Joshi, Pushpak Bhattacharyya, Diptesh Kanojia, and Akhlesh Meena. 2012. A

- study of the sense annotation process: Man v/s machine. In *Proceedings of 6th International Conference on Global Wordnets*, January.
- G. De Melo, C.F. Baker, N. Ide, R.J. Passonneau, and C. Fellbaum. 2012. Empirical comparisons of masc word sense annotations. In *Proceedings of the 8th international conference on language resources and evaluation (LREC12)*. Istanbul: European Language Resources Association (ELRA).
- D. Drieghe, A. Pollatsek, A. Staub, and K. Rayner. 2008. The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6):1552.
- Mitesh M. Khapra, Salil Joshi, and Pushpak Bhat-tacharyya. 2011. It takes two to tango: A bilingual unsupervised approach for estimating sense distributions using expectation maximization. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 695–704, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.
- R.J. Passonneau, A. Salieb-Aouissi, V. Bhardwaj, and N. Ide. 2010. Word sense annotation of polysemous words by multiple annotators. *Proceedings of LREC-7, Valleta, Malta*.
- S.P. Ponzetto and R. Navigli. 2010. Knowledge-rich word sense disambiguation rivaling supervised systems. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 1522–1531. Association for Computational Linguistics.
- S. Vainio, J. Hyönä, and A. Pajunen. 2009. Lexical predictability exerts robust effects on fixation duration, but not on initial landing position during reading. *Experimental psychology*, 56(1):66.