

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

**Salil Joshi, Diptesh Kanojia and  
Pushpak Bhattacharyya,**

**IIT Bombay**

**(NAACL 2013, Atlanta, 11 June, 2013)**

The background features a series of smooth, flowing, translucent green and white lines that create a sense of movement and depth. The lines are layered, with some appearing in the foreground and others receding into the background, set against a light, off-white background.

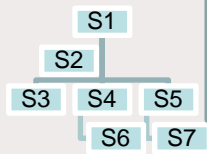
# Motivation

# WSD: strong AI or weak AI?

- Strong AI: mimic human processing
- Weak AI: capture only the functionality of human processing
- Both structurally and functionally
  - Ape whatever the human does
- Has been very successful in many tasks at which humans are good

***However, consistently reported low accuracy for WSD 😞***

Preferred in multilingual multi-domain scenarios



Knowledge Based Approaches

Overlap Based Approaches

- (Lesk 1986)
- (Walker and Amsler 1986)
- (Banerjee and Pedersen 2003)

Similarity Measure Based Approaches

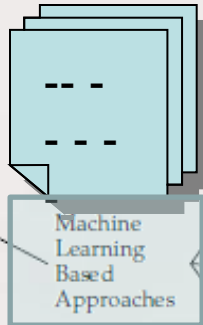
- (Agirre and Rigau 1996)
- (Leacock and Chodorow 1998)
- (Lin 1998)
- (Resnik 1995)
- (Hirst and St-Ogne 1998)
- (Pedersen, Banerjee, and Patwardhan 2005)

**Parameters:**  
similarity with context words

Graph Based Approaches

- (Mihalcea 2005)
- (Agirre and Soroa 2009)

Preferred in scenarios where high accuracy is desired



Monolingual WSD

Machine Learning Based Approaches

Supervised Approaches

- (Yarowsky 1994)
- (Ng and Lee 1996)
- (Escudero, Márquez, and Rigau 2000b)
- (Klein et al. 2002)
- (Lee, Ng, and Chia 2004)

**Parameters:**  
Sense Distributions  
Co-occurrence Statistics

Semi-supervised Approaches

- (Yarowsky 1995)

Unsupervised Approaches

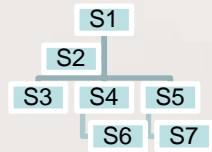
- Corpus Induced Senses
  - (Véronis 2004)
  - (Klapaftis and Manandhar 2008)
- Dictionary Defined Senses
  - (Yarowsky 1992)
  - (Lin 1997)
  - (Agirre et al. 2006)

Hybrid Approaches

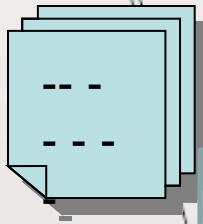
- (Navigli and Velardi 2005)

Web Based Approaches

- (Agirre, Ansa, and Martinez 2001)
- (Mihalcea 2002)



+



**A TAXONOMY OF WSD APPROACHES:**

**ALL WORDS GENERAL PURPOSE ACCURACY HOVERS AROUND 60%**

# Should we change the approach?

- Maybe WSD should be done the way humans disambiguate?
- Study human cognitive process involved in doing sense disambiguation

# Insights from our earlier work\*

Humans need Context for Annotation

Tagging without context is often erroneous, and also a cognitive load due to uncertainty

In supervised WSD, machines rely primarily on prior sense distribution probability

Machines seem to be able to do best with just  $P(S/W)$ ; context per se does not seem important

\*Paper titled "***A Study of the Sense Annotation Process: Man v/s Machine***" published in GWC 2012

# Questions

## Human Cognition in Sense Annotation

- What are the cognitive sub-processes associated with the human sense annotation task?

## Lexicographer's Difficulty

- Which classes of words are more difficult to disambiguate and why?

# Experimental Setup

The background of the slide features a series of smooth, flowing, translucent green and white lines that create a sense of movement and depth. The lines are layered, with some appearing in the foreground and others receding into the background, set against a light, off-white background.



# Eye-tracking

## Fixation

- Eye pause at a certain spot
- First data point
- *Where* someone is focusing, *for how long* and possibly *why*

## Saccades

- Second data point
- Eye gaze movement from one position to another

## Scan Path

- Combination of fixations and saccades

# Techniques for eye-tracking



*Most comfortable technique to measure gaze based on infrared light*



*A bit more complicated way to measure gaze using electric potential around the eye.*



*The eye tracking glasses are used for broad range of mobile eye tracking studies.*



*The ergonomic chin rest eye tracking device for high speed and accurate measurements with a large visual field.*

# Sense marker tool

The screenshot displays the SENSE MAEKER TOOL interface with the following components:

- File Options Help**: Menu bar at the top.
- Format Settings**: Includes checkboxes for Bold, Size (18), Chandas, and HINDI.
- Corpus Pane**: The main text area containing Hindi text with sense IDs. A callout bubble points to the word 'अमरीका' (America) with ID '17867'.
- Sense Tagging of Corpus files**: A callout bubble pointing to the text area.
- Synsets Pane**: A list of synsets for the word 'अमरीका', including 'संयुक्त\_राज्य\_अमरीका' and 'यूनाइटेड\_स्टेट\_ऑफ\_अमेरिका'. A callout bubble points to this pane.
- Use in WSD**: A callout bubble pointing to the synsets pane.
- Assumes one sense per discourse for faster tagging**: A callout bubble pointing to the text area.
- Comments**: A dropdown menu at the bottom.
- Remove ID**: A button at the bottom right.

Marking words with Wordnet sense IDs

# Facts and Figures

- 2000 words used for experimentation
- Analysis done on data for open class words (nouns, verbs, adverbs and adjectives)
- Data from 6 lexicographers (3 skilled, 3 unskilled) collected
- Annotators used Sense-marker tool for tagging the word senses
- Gaze patterns analyzed

# An example of eye movement during sense marking

The screenshot displays the SENSE MAEKER TOOL interface with a Hindi news article. The article text includes:  
:1: पाकिस्तान\_110838 में जून\_17067 की सज्जद\_16211 ने नारंगीना\_12329 की कारवां\_10933 में दो\_49167 पाकिस्तानी\_414212 सैनिकों\_17774 के मारे\_27622 जाने की घटना\_12968 को भयानक\_11240 हादसा\_12860 बताते\_27125 हुए इसके लिए माफी\_12942 मांगी\_17573 है।  
:2: पाकिस्तान\_110838 स्थित\_62770 अमरीका\_413405 दूतावास\_128405 द्वारा जारी\_430147 बयान\_14036 के मुताबिक एक\_4187 संयुक्त जांच\_15560 में पता\_113235 चला है कि 30 सितंबर\_111069 को अमरीका\_413405 हेलाकाटर\_117822 पाकिस्तान\_110838 की सीमा\_19423 में घुस\_23430 आए आतंकवादियों\_13748 का पीछा\_कर\_27861 रहे

Eye movement paths are overlaid on the text, with black dots representing fixations and lines representing saccades. The size of the dots indicates the duration of each fixation. A red arrow points to the first fixation on the word 'अमरीका' in the first line, with the text "America: USA or North America?".

Below the text, two sense entries are shown:

<1> WORDS :: अमरीका, संयुक्त\_राज्य\_अमरीका, संयुक्त\_राज्य\_अमेरिका, यू\_एस\_ए, यू\_एस, यूएसए, यूएस, यूनाइटेड\_स्टेट\_ऑफ\_अमेरिका, संयुक्त\_राष्ट्र  
GLOSS :: उत्तरी अमेरिकी महाद्वीप का एक देश;"कई भारतीय" का में जाकर बस गए है"  
OFFSET :: NOUN[7867] "USA" sense


<2> WORDS :: अमेरिका, अमरीका  
GLOSS :: एक महाद्वीप जिसमें उत्तरी, दक्षिणी आर मध्य अमेरिका है;"अमेरिका दो भागों में विभाजित है, उत्तरी अमेरिका और दक्षिणी अमेरिका"  
OFFSET :: NOUN[10832] "North America" sense

At the bottom of the window, there are fields for "Relations" and "Comments", and a status bar showing "| N : 2 | V : 0 | ADV : 0 | ADJ : 0 |" and a "Remove ID" button.


Note: Ball size indicative of the fixation time; lines are saccades

# Cognitive sub-processes in sense annotation\*

Hypothesis Building: During annotation, the lexicographer makes initial hypotheses regarding meaning and domain of a word  
( $T_{hypo}$ )



Clue-word Searching: Consequently he/she looks for contextual clues around the word to narrow down on 1 or at most 2 of the initial hypotheses  
( $T_{clue}$ )



Gloss Matching: The lexicographer then scans the wordnet candidate senses of the word for synset words and gloss to map their hypothesis to one of the senses  
( $T_{gloss}$ )

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

**\*as discussed with the lexicographers, arguably our Most important contributionn**

# Results

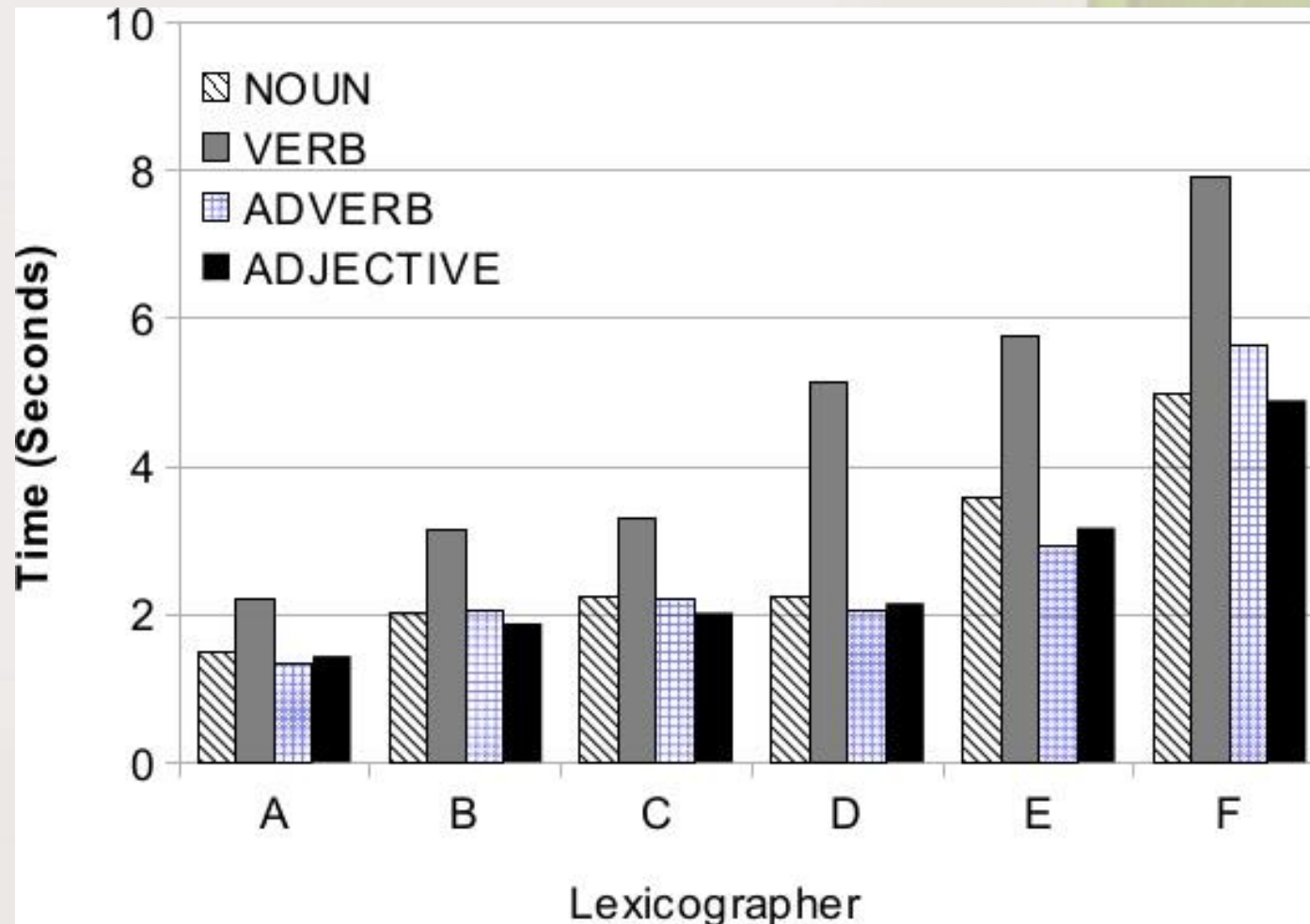
Lexicographer	Time Taken (seconds)			
	$T_{hypo}$	$T_{clue}$	$T_{gloss}$	$T_{total}$
Skilled	0.33	0.74	1.16	2.24
Unskilled	0.74	1.56	4.44	6.75

Time variations between skilled and unskilled lexicographers

Word	Degree of polysemy	Unskilled lexicographers (seconds)				Skilled lexicographers (seconds)			
		$T_{hypo}$	$T_{clue}$	$T_{gloss}$	$T_{total}$	$T_{hypo}$	$T_{clue}$	$T_{gloss}$	$T_{total}$
लाना (laana - to bring)	4	0.63	0.8	5.2	6.63	0.31	1.2	1.82	3.3
करना (karanaa - to do)	22	0.9	1.42	2.2	4.53	0.5	0.64	1.14	2.24
जताना (jataanaa - to express)	4	0.7	2.45	5.93	9.09	0.25	0.39	0.62	1.19

Time taken for verbs by lexicographers (examples)

# Results : time taken for different POS categories



Time taken for different POS categories for skilled (A-C) and unskilled (D-F) lexicographers



# Ontological statistics (verbs)

Ontology	Average of Time Taken	No. of words
घटनासूचक (Event)	1.870816444	11
अनैच्छिक क्रिया (Verbs of Non-volition)	2.59201	1
अवस्थासूचक क्रिया(Verb of State)	4.403871355	77
शारीरिक कार्यसूचक bodily action	4.97281795	40
कर्मसूचक क्रिया (Verb of Action)	5.376058091	11
प्रेरणार्थक क्रिया (causative verb)	5.635743	5
संप्रेषणसूचक (Communication)	5.895843818	11
अधिकारसूचक (Possession)	6.00231725	9
परिवर्तनसूचक (Change)	6.517663706	17
विनाशसूचक (Destruction)	8.7992645	3
होना क्रिया (Verb of Occur)	12.06406657	7
भौतिक अवस्थासूचक (Physical State)	13.47733335	2
निरंतरतासूचक क्रिया (Verbs of Continuity)	17.896006	2
कार्यसूचक (Act)	20.2321495	2
मानसिक अवस्थासूचक (Mental State)	74.698983	1
<b>Grand Total</b>	<b>5.896812948</b>	<b>199</b>

# Discussions

## Cognitive sub-processes for Sense Annotation

- Three stages: Hypothesis building, clue-word searching and gloss matching

## Skilled v/s unskilled lexicographers

- Unskilled  $T_{gloss} \gg T_{clue}$
- Skilled  $T_{gloss} \sim T_{clue}$ ; latch on to the POS quickly

## Maximum annotation time for verbs

- High degree of polysemy
- Senses are fine-grained
- In some cases the hypothesis does not match the candidate senses

## Adverbs and Adjectives

- Annotation time comparable to nouns
- Adjective and adverbs' proximity to the noun helps

# Conclusions

- ✓ Sense annotation process can be divided into 3 stages: Hypothesis building ( $T_{\text{hypo}}$ ), Clue-word searching ( $T_{\text{clue}}$ ) and gloss matching ( $T_{\text{gloss}}$ )
- ✓ The theory can be verified by analyzing the gaze patterns
- ✓ Skilled lexicographers annotate the words faster
  - ✓ have knowledge about the senses of a word (significantly reducing the time  $T_{\text{gloss}}$ )
- ✓ Verbs take the highest time among the POS categories given the high degree of polysemy and lack of exact senses
- ✓ Adverbs and adjectives are easier to annotate given their position near a verb or a noun
- ✓ Automating the process of identifying the clue-words from the gaze patterns can lead to building a rich *discrimination-net*

# References

- E. Agirre, O.L. De Lacalle, A. Soroa, and I. Fakultatea. 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 8<sup>th</sup> international conference on language resources and evaluation (LREC12). Istanbul.
- 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 Bhattacharyya. 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.
- R.J. Passonneau, A. Salleb-Aouissi, V. Bhardwaj, and N. Ide. 2010.** *Word sense annotation of polysemous words by multiple annotators.* Proceedings of LREC-7, Valletta, 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.
- 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.

The background features a series of smooth, flowing, translucent green and white lines that create a sense of movement and depth. The lines are layered, with some appearing more prominent than others, giving the overall effect a soft, ethereal quality. The colors transition from a pale, almost white green to a slightly darker, more vibrant green, especially in the lower right quadrant where the lines are more densely packed.

**Thank You!**

The background features a light green and white color palette with soft, flowing, and layered lines that create a sense of movement and depth. The lines are most prominent on the right side, curving downwards and then back up, while the left side is more open and white.

Backup slides

# Eye-tracking

Saccades

Fixations

