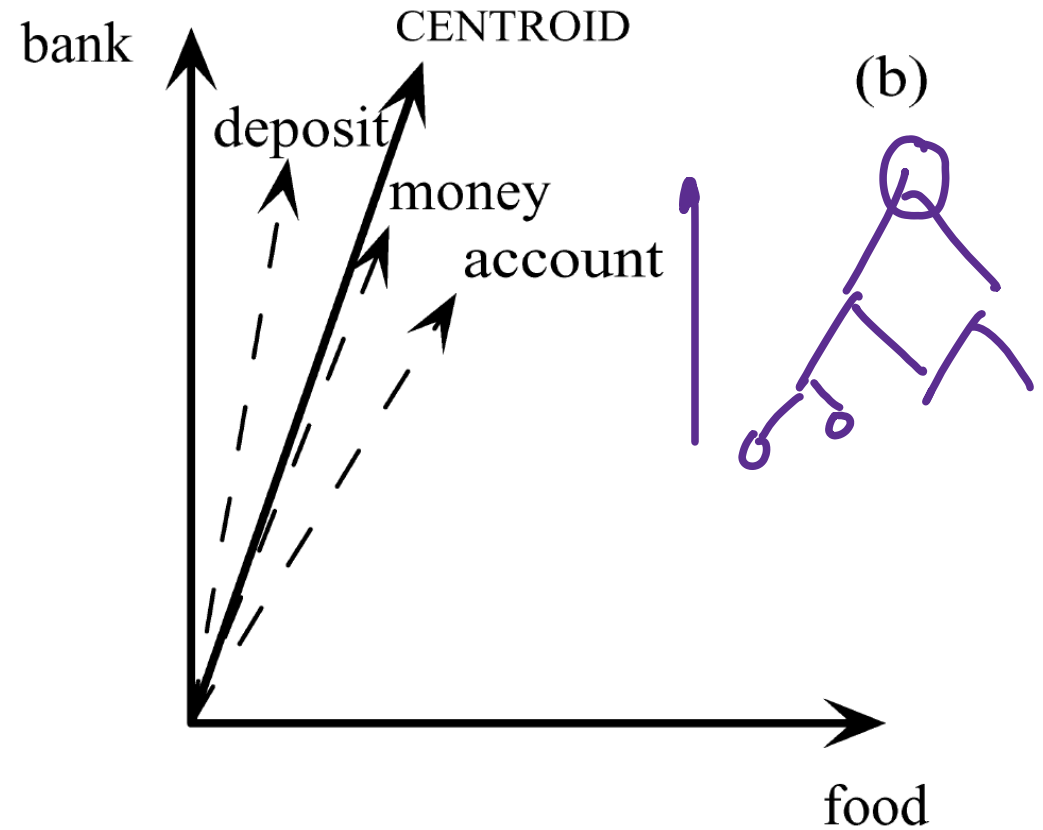
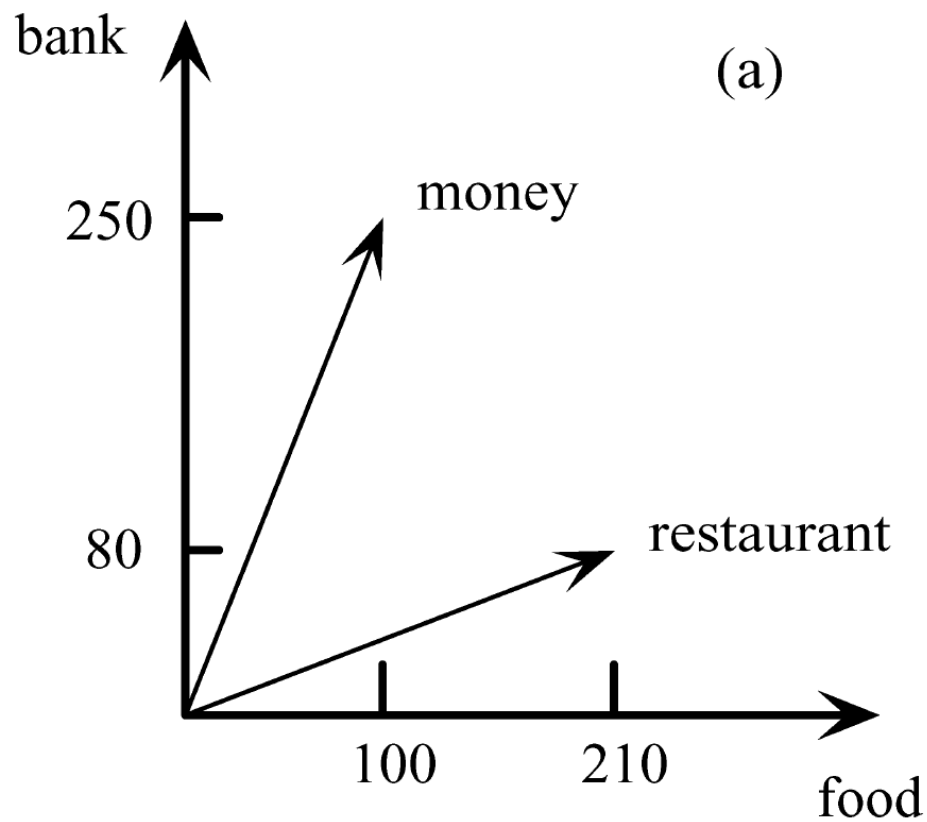


WSD : Part 2

Discrimination Unsupervised WSD



LSA, Clustering using EM or hierarchical clustering

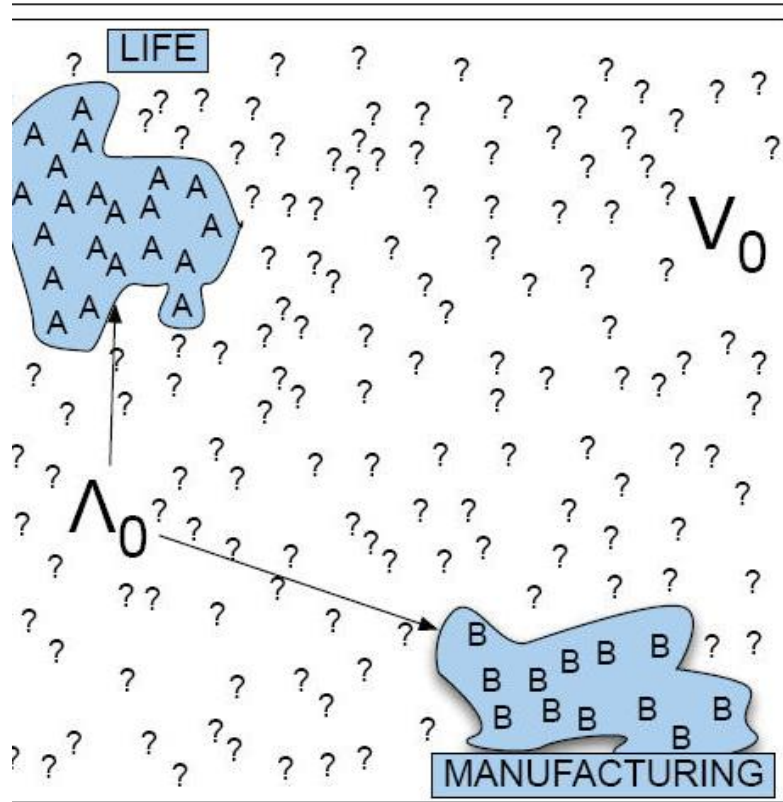
Ref. Fig. 15 in paper

Unsupervised WSD: Word Clustering

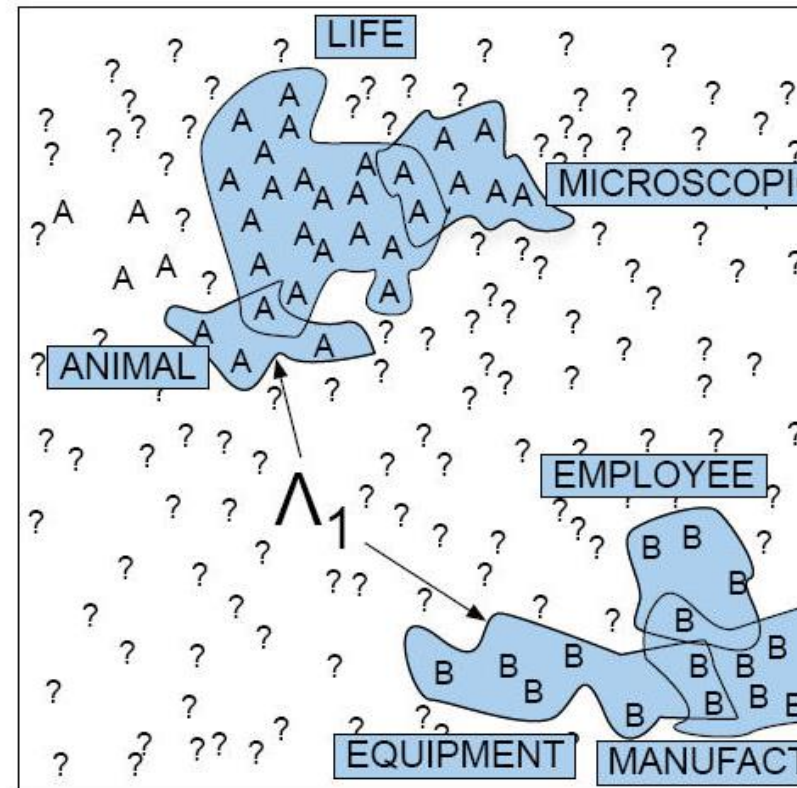
- Based on the idea of word similarity
- Distributional models of word similarity

Semi-supervised WSD

active learning



(a)



(b)

The Web as Corpus

- An entry point into several interesting papers in NLP
- Reference paper (for those interested in NLP research):
<https://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf>

*Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.*

-- (Robert Frost)

(from the book “Web as Corpus” by Maristella Gatto (Bloomsbury Academic, 2014))

Knowledge based WSD: The Lesk algorithm (using glosses)

WordNet Sense Inventory for the First Three Senses of key_n

Sense	Definition and Examples
key_n^1	Metal device shaped in such a way that when it is <i>inserted</i> into the appropriate <i>lock</i> the <i>lock's</i> mechanism can be rotated
key_n^2	Something crucial for explaining; "the key to development is economic integration"
key_n^3	Pitch of the voice; "he spoke in a low key"

$\frac{19,617}{19,617}$ { I inserted the key and locked the door. }

Refer Section 5.1

Knowledge based WSD: Using selectional restrictions

Idea on selectional restrictions discussed in Part 1 of this presentation with example, Refer Section 5.2 in paper

Knowledge based WSD: Using structural methods

Self-study: Refer Section 5.3

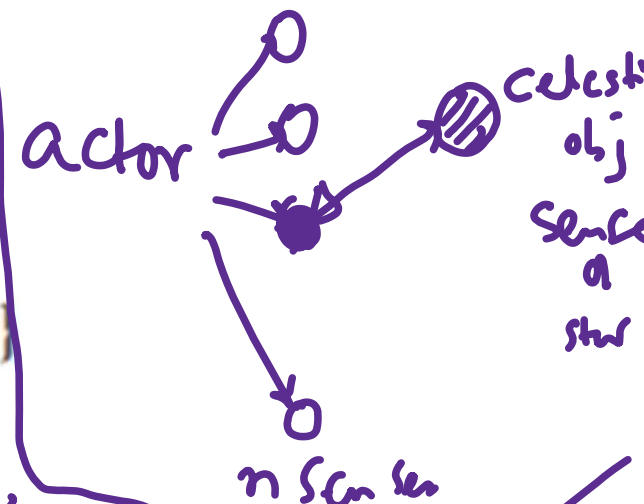
WSD using Dominance

to be disambiguated

$$N(star_n) = \{actor_n, footballer_n, planet_n\}$$

$$\left. \begin{aligned} sim_{dist}(star_n, actor_n) &= 0.22 \\ sim_{dist}(star_n, footballer_n) &= 0.12 \\ sim_{dist}(star_n, planet_n) &= 0.08 \end{aligned} \right\}$$

KL divergence



Party 1

Celestial object sense

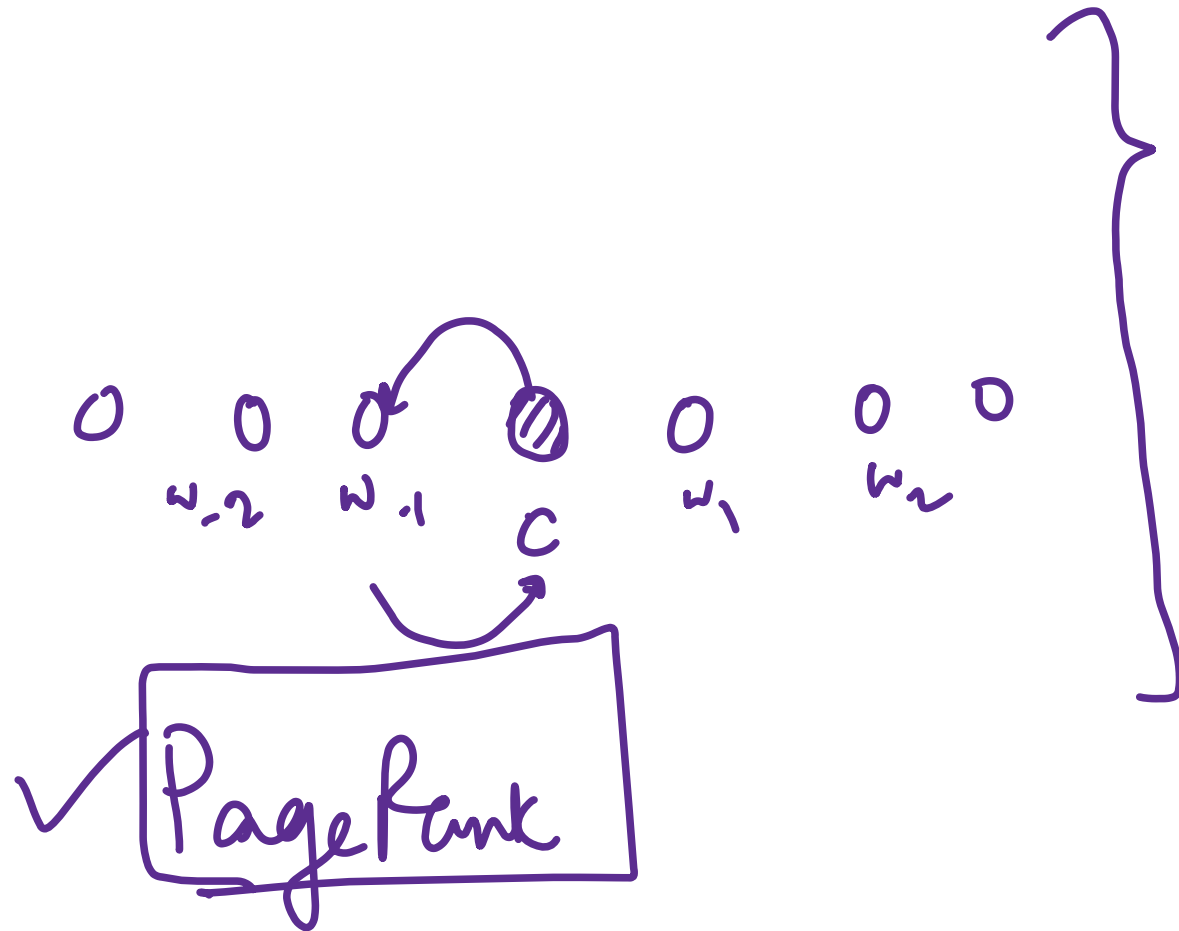
Celebrity sense

$$\begin{aligned} score_{prev}(star_n^1) &= 0.22 \cdot \frac{0.01}{0.48} + 0.12 \cdot \frac{0.01}{0.57} + 0.08 \cdot \frac{0.68}{0.93} = 0.068, \checkmark \\ score_{prev}(star_n^5) &= 0.22 \cdot \frac{0.42}{0.48} + 0.12 \cdot \frac{0.53}{0.57} + 0.08 \cdot \frac{0.02}{0.93} = 0.314, \checkmark \end{aligned}$$

Democracy analogy

Refer Section 6.1

Circularity in WSD



PageRank for WSD

The **church** bells no longer **rung** on **Sundays**.

church

- 1: one of the groups of Christians who have their own beliefs and forms of worship
- 2: a place for public (especially Christian) worship ✓
- 3: a service conducted in a church

bell

- 1: a hollow device made of metal that makes a ringing sound when struck
- 2: a push button at an outer door that gives a ringing or buzzing signal when pushed ✓
- 3: the sound of a bell

ring

- 1: make a ringing sound
- 2: ring or echo with sound
- 3: make (bells) ring, often for the purposes of musical edification ✓

Sunday

- 1: first day of the week; observed as a day of rest and worship by most Christians
-

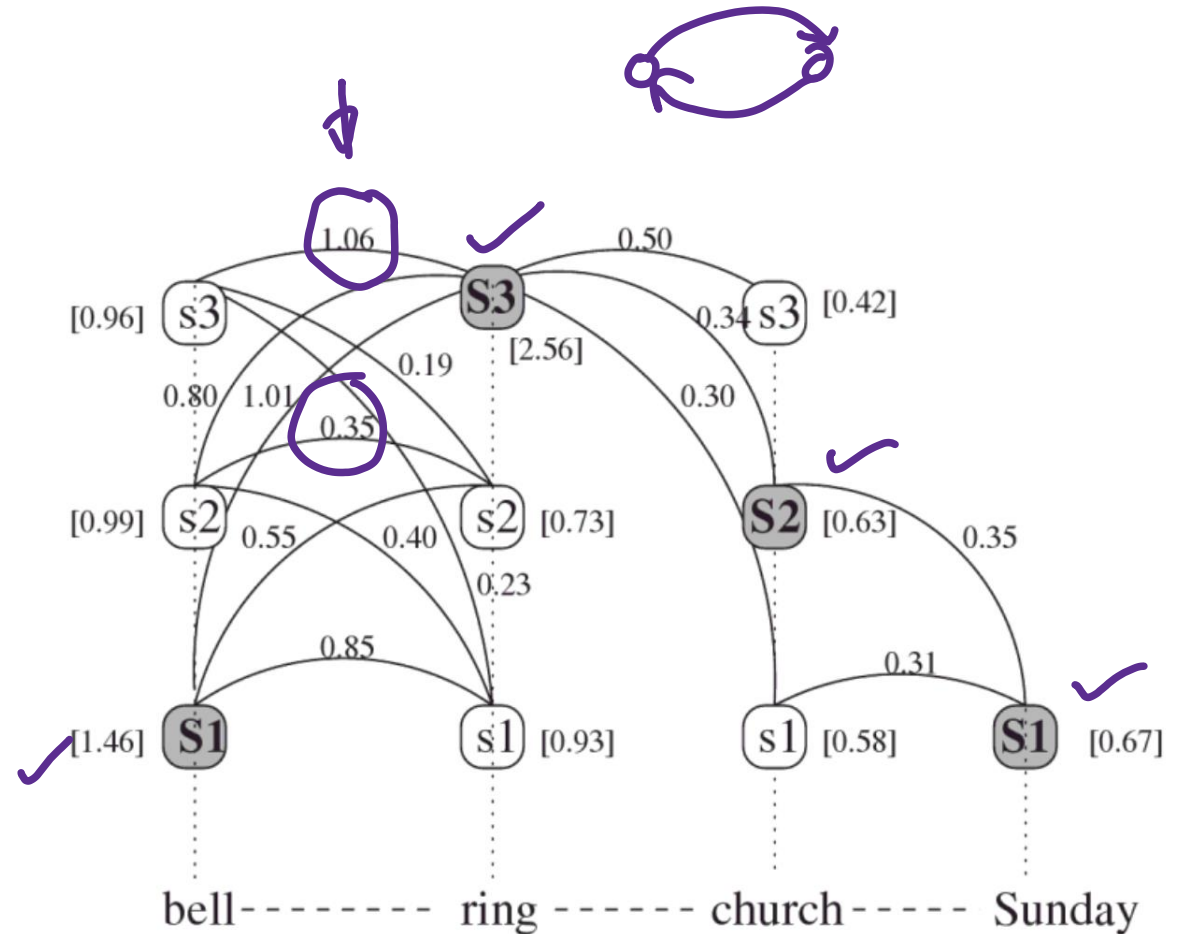


Fig.: Recent Advances in Natural Language Processing V: Selected Papers from RANLP

edited by Nicolas Nicolov, Galia Angelova, Ruslan Mitkov

PageRank: <https://www.aclweb.org/anthology/E09-1005.pdf>

In summarization: TextRank: <https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf>

Lexical Chains

... the **book** that he had just taken out of the drawer. **It** was a peculiarly beautiful **book**. **Its** smooth creamy **paper** was of a kind that had not been manufactured for at least forty years past. He could guess, however, that the **book** was much older than that. He had seen **it** lying in the window of a frowsy little **junk-shop** and had been stricken immediately by an overwhelming desire to **possess** **it**. Party members were supposed not to go into ordinary **shops**. He had slipped inside and **bought** the **book** for two **dollars** fifty. Even with nothing **written** in **it**, **it** was a compromising **possession**. ¶ The thing that he was about to do was to open a **diary**. Winston fitted a **nib** into the **penholder** and sucked **it** to get the grease off. The **pen** was an archaic **instrument**, seldom used even for **signatures**, and he had **procured** one, furtively and with some difficulty, simply because of a feeling that the beautiful creamy **paper** deserved to be **written** on with a real **nib** instead of being scratched with an **ink-pencil**. Actually he was not used to **writing** by hand. He dipped the **pen** into the **ink**.

{ Cohesion /
Coherence

20%

Colouring a document



Refer Section 5.3.2

WSD Evaluation

Coverage

$$C = \frac{\frac{|A|}{|W|}}{\frac{\# \text{ answers provided}}{\# \text{ total answers to provide}}}$$

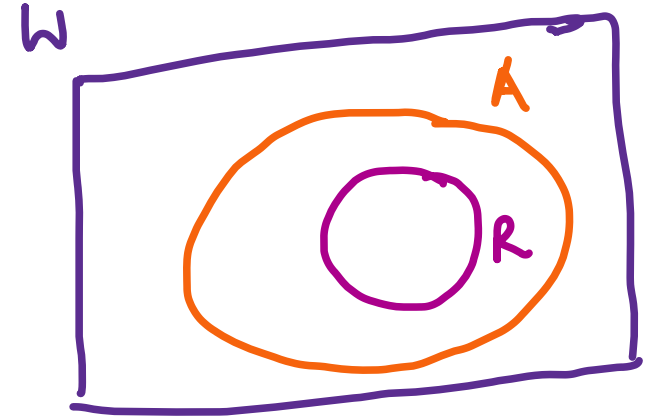
Precision

$$P = \frac{\frac{|R|}{|A|}}{\frac{\# \text{ correct answers provided}}{\# \text{ answers provided}}}$$

Recall

$$R = \frac{\frac{|R|}{|W|}}{\frac{\# \text{ correct answers provided}}{\# \text{ total answers to provide}}}$$

$$P \geq R \quad \checkmark$$



W: all words to be disambiguated

A: words that were disambiguated

R: words that are correctly disambiguated

Refer Section 7.1

Baselines, Lower and Upper Bounds

- Upper Bound : not 100% but Inter-annotator agreement

IAA

94%

↗

Applications of WSD and Open Problems

- Self study (gloss over and make note of important points)