Word Sense Disambiguation

Didier SCHWAB
Didier.Schwab@imag.fr

What is Word Sense Disambiguation ?

- Natural languages are ambiguous:

The mouse ate some cheese

What is Word Sense Disambiguation ?

- Natural languages are ambiguous:

*The **mouse** ate some cheese*

What is Word Sense Disambiguation ?

- Natural languages are ambiguous:

The ***mouse*** *ate some cheese*

What is Word Sense Disambiguation ?

- Natural languages are ambiguous:

The mouse ate some cheese

1. ...rodent...	1. ...food...	1. ...food...
2. ...bruise...	2. ...meal...	2. ...cluster...
3. ...person...	3. ...animal...	
4. ...device...	4. ...anxiety...	
	5. ...use up...	
	6. .deteriorate..	

What is Word Sense Disambiguation ?

- Natural languages are ambiguous:

The mouse ate some cheese

1. ...rodents...	1. ...food...	1. ...food...
2. ...bruise...	2. ...meal...	2. ...cluster...
3. ...person...	3. ...animal...	
4. ...device...	4. ...anxiety...	
	5. ...use up...	
	6. .deteriorate..	

What is Word Sense Disambiguation ?

- Most words have several possible meanings
- => Very few have a single meaning
- Monosemic : *'neuroleptic', 'daucus carota'*,
- Polysemic : *'mouse', 'rabbit', 'carot'*
- In English : the 121 most frequent nouns
  - On average 1 word out of five in actual texts
  - ~7.8 meanings per word (in Princetown WordNet)
- What is (often) really easy task for a human is difficult for a computer
- Finding a better sense for a word in a text is called

  Word Sense Disambiguation

What is Word Sense Disambiguation ?

- Aim of WSD: selecting a sense for each word in a text from an inventory (set) of predefined possibilities
- A word sense is the meaning of a word in a given context
- Inventories are produced from dictionnaries, raw texts, …
- How to represent word senses ?
- How to fetch the meanings of a word ?

Sets of Word Senses

- How to fetch the meanings of a word ?
  - With respect to a dictionary, a lexical base...
    - **mouse#1** : any of numerous small rodents...
    - **mouse#2** : a hand-operated electronic device...
  - With respect to the translation in a second language
    - **mouse#1 :** tikus
    - **mouse#2 :** tetikus

Sets of Word Senses

- How to fetch the meanings of a word ?
  - With respect to the context where it occurs...
    - **mouse#1** : „The cat hurt the mouse“ ; “The mouse is eating the cheese“ ; ...
    - **mouse#2** : „The mouse is linked to the computer.“ ; „My mouse is broken.“ ; ...
  - With respect to relations it shares in a semantic network
    - **mouse#1 :** hypernyms (kind-of) : *'rodent*', *'mammal*',... ; related-to : *'mousy'*, *'mousey'*
    - **mouse#2 :** hypernyms : *'electronic device'* ; related-to : *'to mouse'*
  - Others
  - Combinations

Sense Tagging

Sense Tagging

- Given a pre-defined inventory of word senses
- Given a text
- Tag each ambiguous word occurrence with the most likely word sense
- Example :
- 'The cat is eating the mouse'

Sense Tagging

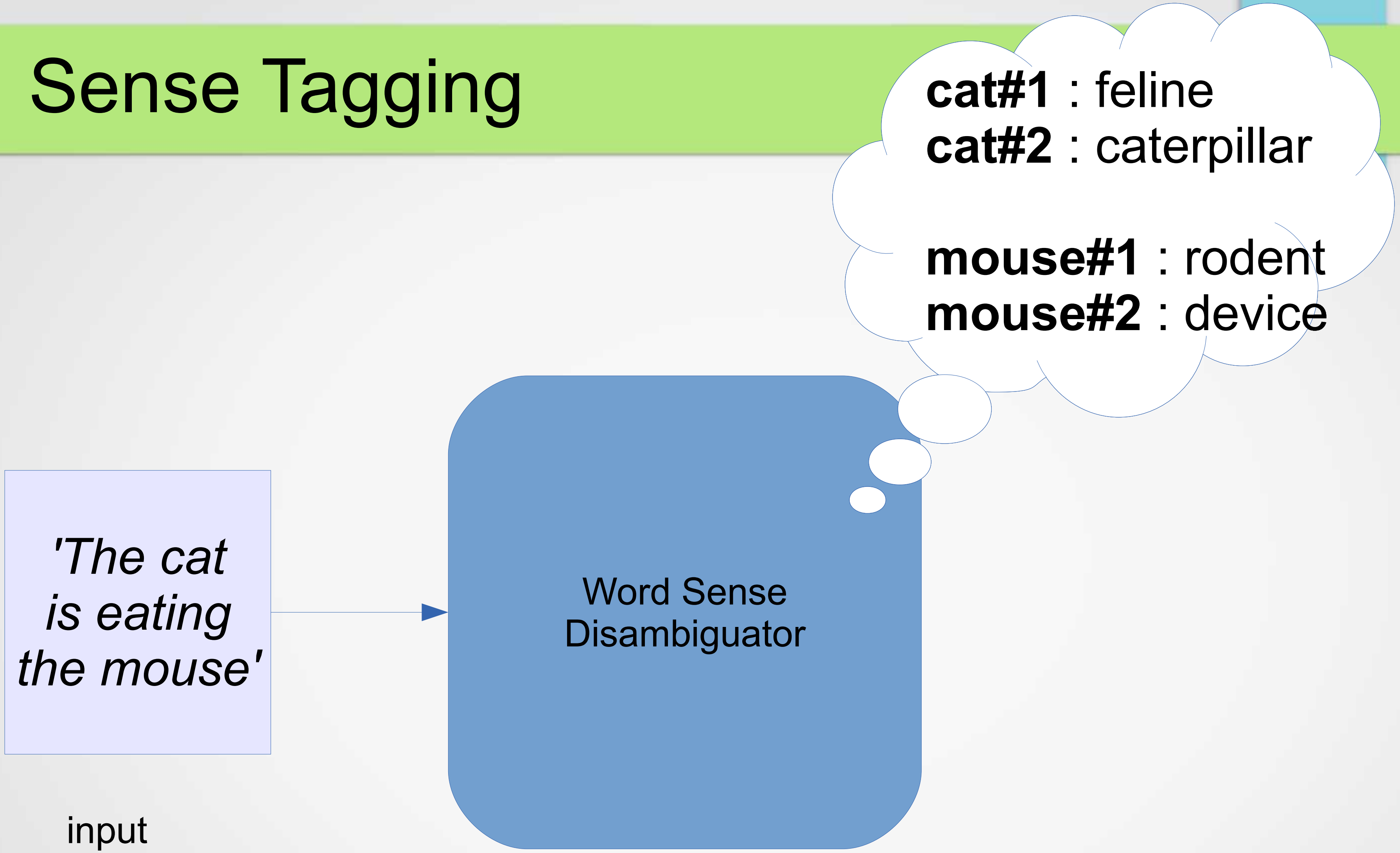
'The cat is eating the mouse'

Sense Tagging

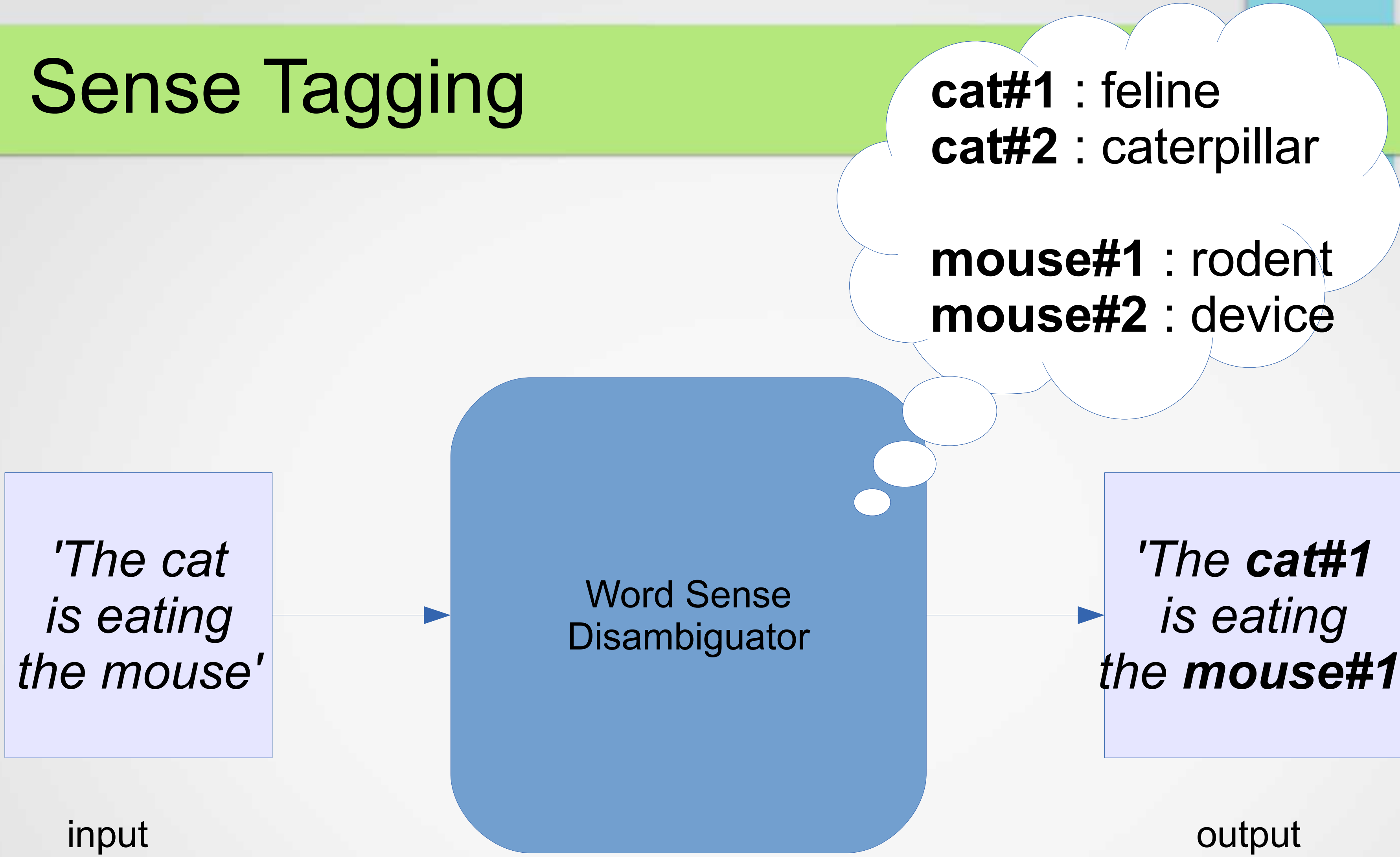
'The cat is eating the mouse'

Word Sense Disambiguator

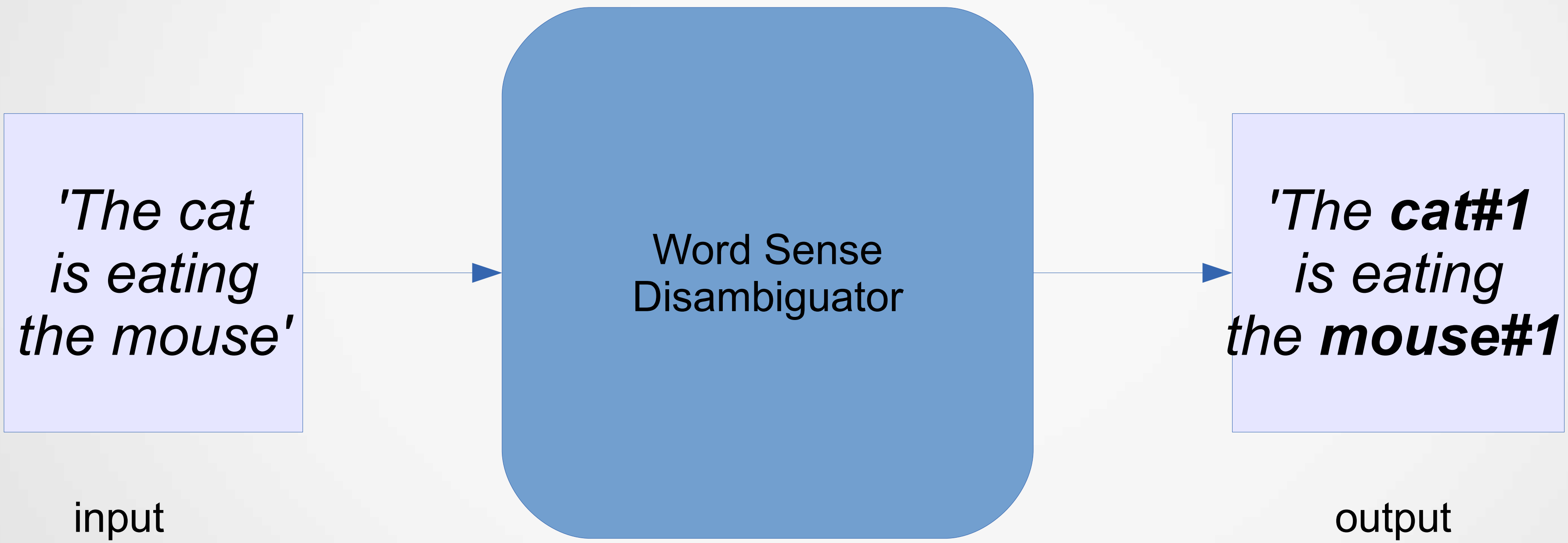
Sense Tagging



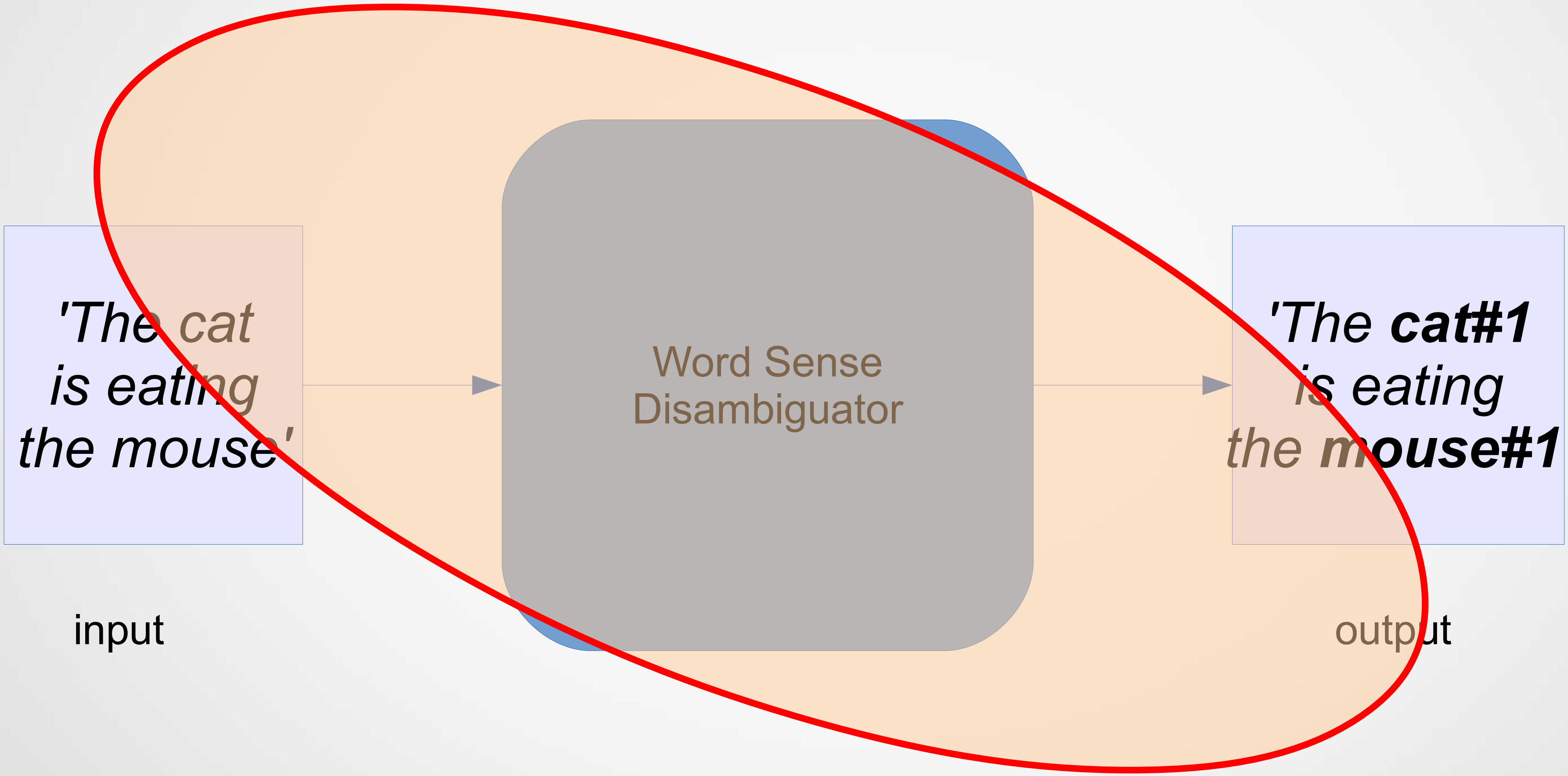
Sense Tagging



Sense Tagging



Sense Tagging



Practical Applications

WSD for machine translation

- Which translation of "mouse" ?

tetikus

tikus

- Which translation of “bank” in French?

Bank → Berge

Bank → Banque

WSD for machine translation

….
bank#1: a financial institution that accepts deposits and channels money into lending activities

bank#2: sloping land (especially the slope beside a body of water)
….

(Malay translations)

bank

tebing

…withdraw money from the **bank**...

sense-tag (WSD)

…withdraw money from the **bank#1**...

select translation word

Malay output

…mengeluarkan wang dari **bank**...

WSD for Information Retrieval

 *mouse*

 *mouse*

 *house*

WSD for Information Retrieval

Query :

house ?

WSD for Information Retrieval

Query :

house ?

WSD for Information Retrieval

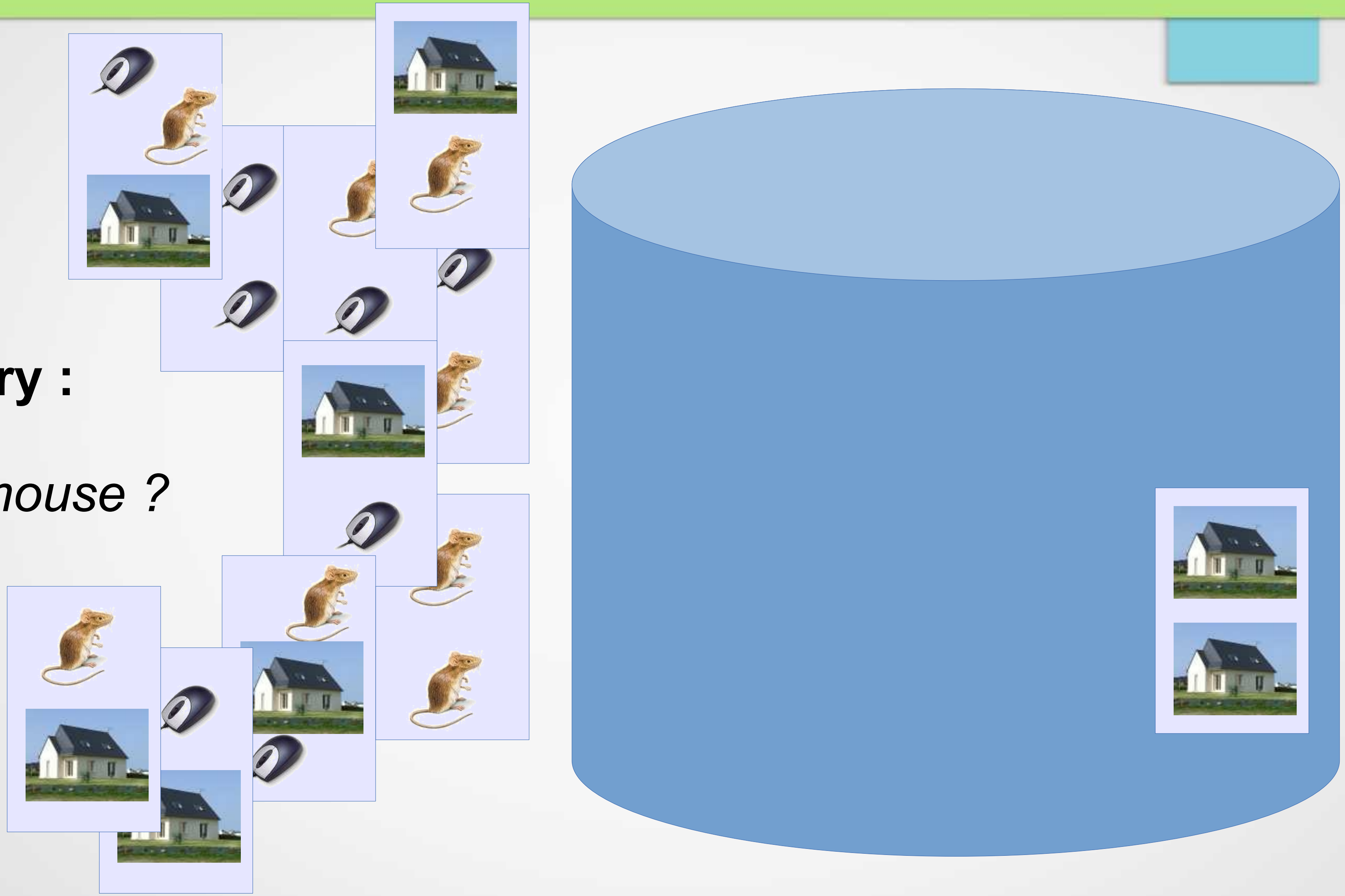
Query :

mouse ?

WSD for Information Retrieval

Query :

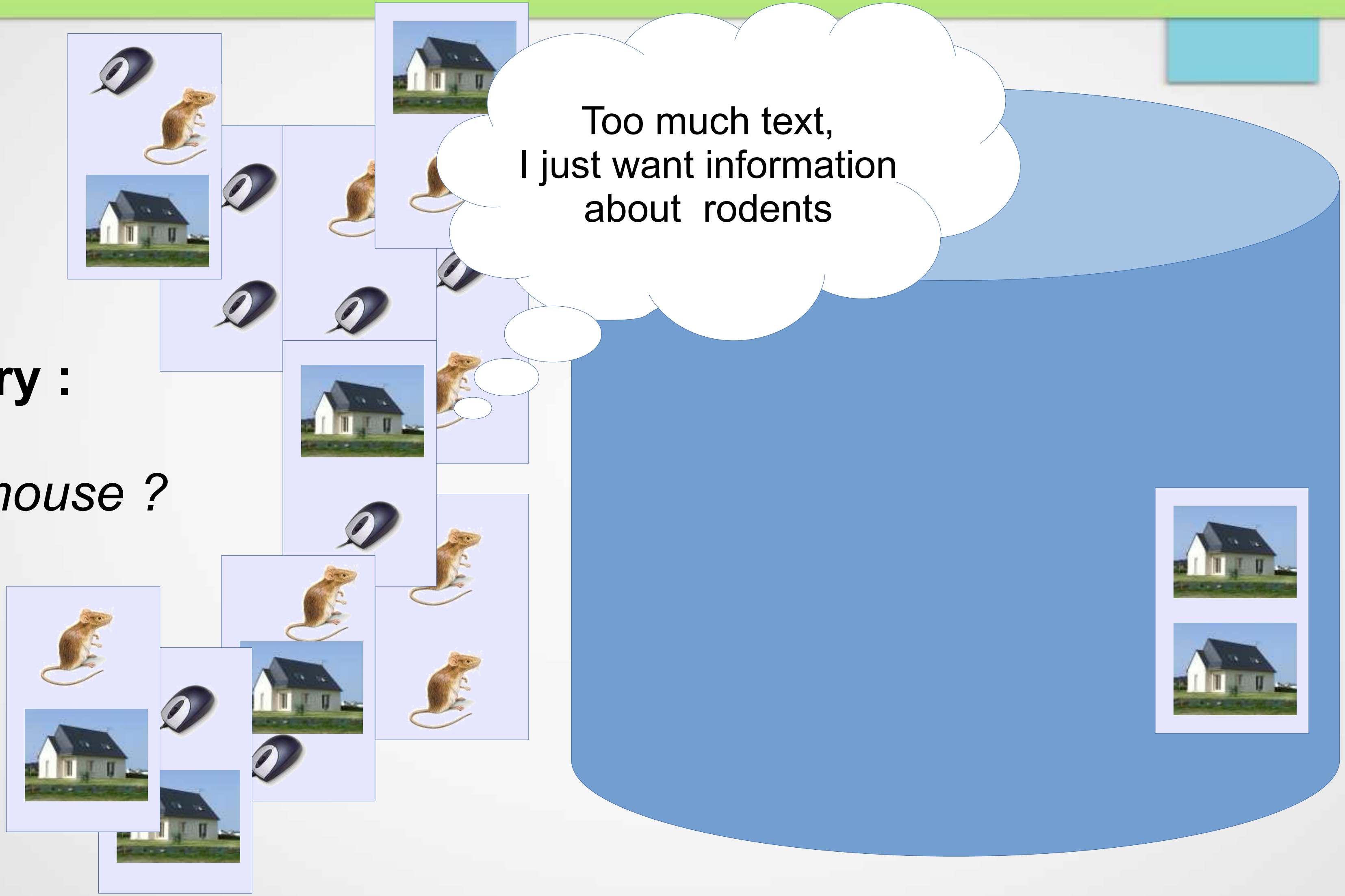
mouse ?



WSD for Information Retrieval

Query :

mouse ?



WSD for Information Retrieval

Query :

mouse
rodent ?

WSD for Information Retrieval

Query :

mouse
rodent ?

WSD for Question Answering

- Systems that automatically answer questions posed by humans in a natural language
- Examples :
  - Where is the Effel Tower ?
  - What time is it ?
  - When did George Bush enter in White House ?

WSD for Question Answering

When did George Bush enter in White House ?

WSD for Question Answering

When did George Bush enter in office?

Which George Bush ?

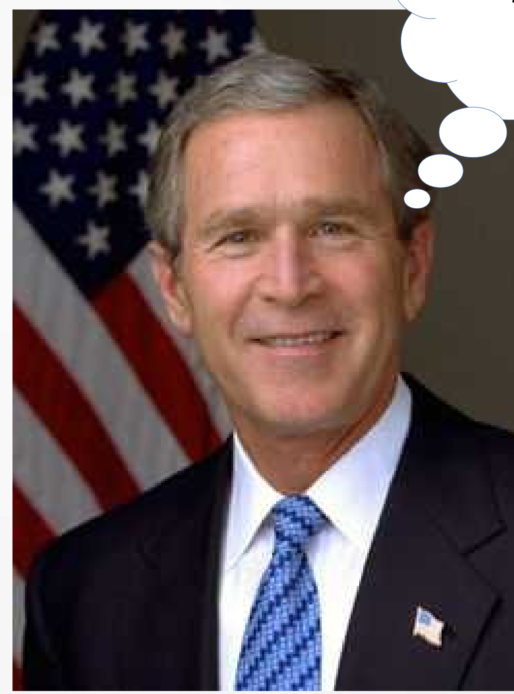
WSD for Question Answering

When did Georc ter in White House ?

1989

Wh e Bush ?

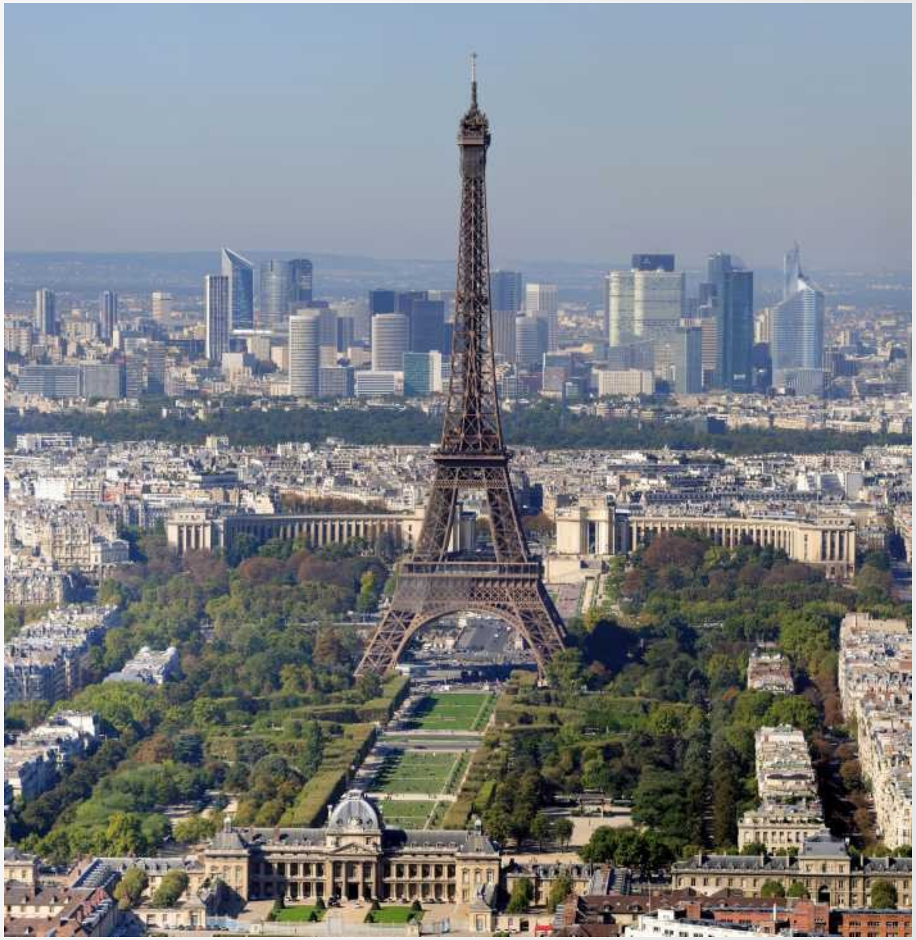
2001



Knowledge Acquisition

The liberation of Paris was in 1944

Kentucky, USA



France

Knowledge Acquisition

Mozart est mort à Vienne

Austria

France

WSD for speech synthesis

- Artificial production of human speech from written text
- Integrated in some operating systems
- Useful for:
  - Blind people
  - Mutes
  - System interaction through phones

WSD for speech synthesis

French : fils (yarn)

[fil]

[fis]

Speech recognition

- Artificial production of text from human speech
- Homophones: Two words that sound the same but have different meanings

night [nIt] knight

Speech recognition

ancre [ancre] encre

Evaluating Word Sense Disambiguation Performance

Evaluation of WSD Systems

- *In vivo* evaluation
  - WSD systems evaluated through their contributions to the overall performance of a particular NLP application
  - The most natural way to evaluate
  - But the harder to set up
- *In vitro* evaluation
  - WSD task defined independently of any particular application
  - Systems evaluated using specially constructed benchmarks

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without annotations

Evaluation of WSD Systems

- *In vivo* evaluation (extrinsic)
    - WSD systems evaluated through their contributions to the overall performance of a particular NLP application
    - The most natural way to evaluate
    - But the most difficult to set up
- *In vitro* evaluation (intrinsic)
    - WSD task defined independently from any particular application
    - Systems evaluated using specifically constructed benchmarks

In Vitro Evaluation

- A benchmark (gold-standard):reference sense-annotated corpus
- The same corpus without annotations

d001 d001.s001.t001 editorial%1:10:00:: !! lemma=editorial#n
d001 d001.s001.t002 ill%3:00:01:: !! lemma=Ill#a
d001 d001.s001.t003 homeless%1:14:00:: !! lemma=Homeless#n
d001 d001.s001.t004 refer%2:42:00:: !! lemma=refer#v
d001 d001.s001.t005 research%1:09:00:: !! lemma=research#n
d001 d001.s001.t006 six%5:00:00:cardinal:00 !! lemma=six#a
d001 d001.s001.t007 colleague%1:18:01:: !! lemma=colleague#n
d001 d001.s001.t008 report%2:32:13:: !! lemma=report#v

In Vitro Evaluation

- A benchmark (gold-standard):reference sense-annotated corpus
- The same corpus without annotations

d001 d001.s001.t001 editorial%1:10:00:: !! lemma=editorial#n
d001 d001.s001.t002 ill%3:00:01:: !! lemma=Ill#a
d001 d001.s001.t003 homeless%1:14:00:: !! lemma=Homeless#n
d001 d001.s001.t004 refer%2:42:00:: !! lemma=refer#v
d001 d001.s001.t005 research%1:09:00:: !! lemma=research#n
d001 d001.s001.t006 six%5:00:00:cardinal:00 !! lemma=six#a
d001 d001.s001.t007 colleague%1:18:01:: !! lemma=colleague#n
d001 d001.s001.t008 report%2:32:13:: !! lemma=report#v

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The sam[illegible] without annotations

First document

d001 d001.s001.t001 editorial%1:10:00:: !! lemma=editorial#n
d001 d001.s001.t002 ill%3:00:01:: !! lemma=Ill#a
d001 d001.s001.t003 homeless%1:14:00:: !! lemma=Homeless#n
d001 d001.s001.t004 refer%2:42:00:: !! lemma=refer#v
d001 d001.s001.t005 research%1:09:00:: !! lemma=research#n
d001 d001.s001.t006 six%5:00:00:cardinal:00 !! lemma=six#a
d001 d001.s001.t007 colleague%1:18:01:: !! lemma=colleague#n
d001 d001.s001.t008 report%2:32:13:: !! lemma=report#v

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The sam… without annotations

First document

d001 d001.s001.t001 editorial%1:10:00:: !! lemma=editorial#n
d001 d001.s001.t002 ill%3:00:01:: !! lemma=Ill#a
d001 d001.s001.t003 homeless%1:14:00:: !! lemma=Homeless#n
d001 d001.s001.t004 refer%2:42:00:: !! lemma=refer#v
d001 d001.s001.t005 research%1:09:00:: !! lemma=research#n
d001 d001.s001.t006 six%5:00:00:cardinal:00 !! lemma=six#a
d001 d001.s001.t007 colleague%1:18:01:: !! lemma=colleague#n
d001 d001.s001.t008 repo… report#v

5th term
of the first sentence
of the first document

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The sam… without annotations

First document

Solution
(best sense)

d001 d001.s001.t001 editorial%1:10:00:: !! lemma=editorial#n
d001 d001.s001.t002 ill%3:00:01:: !! lemma=Ill#a
d001 d001.s001.t003 homeless%1:14:00:: !! lemma=Homeless#n
d001 d001.s001.t004 refer%2:42:00:: !! lemma=refer#v
d001 d001.s001.t005 research%1:09:00:: !! lemma=research#n
d001 d001.s001.t006 six%5:00:00:cardinal:00 !! lemma=six#a
d001 d001.s001.t007 colleague%1:18:01:: !! lemma=colleague#n
d001 d001.s001.t008 repo… report#v

5th term
of the first sentence
of the first document

lemma

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”The Ill Homeless” referred to research by us and six of our colleagues that was reported in the Sept. 8 issue of the Journal of the American Medical Association .

- Texts

<text id="d001">
<sentence id="d001.s001">
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

In Vitro Evaluation

- A benchmark : a reference sense-annotated corpus
- The same corpus without annotations

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”The Ill Homeless” referred to research by us and six of our colleagues that was reported in the Sept. 8 issue of the Journal of the American Medical Association .

- Texts

```
<text id="d001">
<sentence id="d001.s001">
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless"
pos="n">Homeless</instance>
...
```

In Vitro Evaluation

- A benchmark[…] corpus
- The same c[…] annotations

First text

- Raw Texts

Your Oct. 6 editorial "The Ill Homeless" referred to research by us and six of our colleagues that was reported in the Sept. 8 issue of the Journal of the American Medical Association .

- Texts

<text id="d001">
<sentence id="d001.s001">
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”The Ill Homeless” referred to research by us and six of our colleagues that was reported in the Sept. 8 issue of the Journal of the American Medical Association .

- Texts

```
<text id="d001">
<sentence id="d001.s001">
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless"
pos="n">Homeless</instance>
...
```

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same c nnotations

First sentence
of the first text

- Raw Texts

Your Oct. 6 editorial ”The Ill Homeless” referred to research by us and six of our colleagues that was reported in the Sept. 8 issue of the Journal of the American Medical Association .

- Texts

<text id="d001">
<sentence id="d001.s001">
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”The Ill Homeless” referred to research by us and six of our colleagues that was reported in the Sept. 8 issue of the Journal of the American Medical Association .

- Texts

<text id="d001">
<sentence id="d001.s001">
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same … sense-annotations

Unevaluated parts

- Raw Texts

Your Oct. 6 … eless" referred to research by us and six of our colleagues that was … the Sept. 8 issue of the Journal of the American Medical Association .

- Texts

<text id="d00…">
<sentence id="d001.s001">
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”The Ill Homeless” referred to research by us and six of our colleagues that was reported in the Sept. 8 issue of the Journal of the American Medical Association .

- Texts

```
<text id="d001">
<sentence id="d001.s001">
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless"
pos="n">Homeless</instance>
...
```

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”Th … rred to research by us and six of our colleagues that was repor … of the Journal of the American Medical Association .

First term
of the first sentence
of the first document

- Texts

<text id="d001">
<sentence id="d001.s001
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”Th rred t us and six of our colleagues that was repor of th American Medical Association .

First term
of the first sentence
of the first document

lemma

- Texts

<text id="d001">
<sentence id="d001.s001
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”Th rred t us and six of our colleagues that was repor of th American Medical Association .

- Texts

<text id="d001">
<sentence id="d001.s001
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</inst ce>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

First term of the first sentence of the first document

lemma

Part of speech

In Vitro Evaluation

- A benchmark : a sense-annotated corpus
- The same corpus without sense-annotations

- Raw Texts

Your Oct. 6 editorial ”Th … rred t… us and six of our colleagues that was repor… of th… Am… Association .

- Texts

<text id="d001">
<sentence id="d001.s001
Your Oct. 6
<instance id="d001.s001.t001" lemma="editorial" pos="n">editorial</instance>
``The
<instance id="d001.s001.t002" lemma="Ill" pos="a">Ill</instance>
<instance id="d001.s001.t003" lemma="Homeless" pos="n">Homeless</instance>
...

First term of the first sentence of the first document

lemma

word

Part of speech

In Vitro Evaluation : metrics

$$precision = \frac{words\ correctly\ tagged}{tagged\ words}$$

$$recall = \frac{words\ correctly\ tagged}{words}$$

$$F-measure = \frac{2 \times precision \times recall}{precision + recall}$$

If all words are tagged

$$P = R \rightarrow F-measure = \frac{2 \times P \times P}{P + P} = \frac{2 \times P^2}{2 \times P} = P$$

In Vitro Evaluation : metrics

$$precision = \frac{words\ correctly\ tagged}{tagged\ words}$$

$$recall = \frac{words\ correctly\ tagged}{words}$$

$$F-measure = \frac{2 \times precision \times recall}{precision + recall}$$

If all words are tagged

$$P = R = F-measure$$

In Vitro Evaluation : example

- Example :
  - 100 words to tag
  - The system tags 75 words
  - 50 are correctly tagged
  - Precision : 50/75 = 66%
  - Recall : 50/100 = 50%
  - F-measure ≈ 56.9%

Bounds of performance

- Evaluating performance of an algorithm relative to the difficulty of the benchmark
- Lower bound (baseline)
  - random assignement: average score obtained when a random sense is chosen for each words in the text

$$\text{random baseline} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|senses(w_i)|}$$

  - most frequent sense: score when the most frequent sense in the language is chosen for each word in the text
- Upper bound
  - Highest performance reasonably attainable
  - Average human interannotator agreement : Around 90%

Example: Semeval 2007 task 7

- All-words task: sense labelling task over all parts-of-speech (nouns, verbs, adjectives, adverbs)
- 2269 words over 5 texts: journalism, book review, travel, computer science, biography
- Disambiguated reference tagged with WordNet senses Evaluation in terms of Precision, Recall, F1 score
- Currently the most recent general English All-words disambiguation task available.

Example: semeval 2007 task 7

- Coarse-grained evaluation : close senses are counted as equivalent (e.g. snow/precipitation and snow/cover)
- Two ways to use this benchmark
  - *A Posteriori*
    - Input: fine-grained (WordNet Senses)
    - Random baseline: 61,27%
    - First sense baseline: 78,89%
  - *A priori*
    - Input: coarse-grained
    - Random baseline: 52,57%
    - First sense baseline: 78,89%

General Overview
of
Word Sense Disambiguation Systems

Word Sense Disambiguation Process

- Composed of 3 steps
  - Build/select raw lexical material(s)
  - Build an elaborate resource
  - Use that resource to lexically disambiguate a text

Build/Select of Raw Lexical Material(s)

- One or more of several types of materials can be used:
  - Dictionaries, encyclopedias, lexical databases
  - Unnanotated corpora, Sense-annotated corpora
- Among existing material, some:
  - Are generated/built automatically
  - Require significant human effort and supervision

Build an elaborate resource

- Computational representation of an inventory of possible word senses
- Two ways of obtainig inventories of word senses:
  - Induction from word contexts
    - When only non-annotated corpora are available
  - Human experts
    - e.g. Dictionaries, Structured Lexical Resources
- Many undelying computational representations:
  - Semantic Networks (graphs)
  - Bags of words & n-gram models
  - Vector spaces

Use the resource to disambiguate

- The Word Sense Disambiguation algorithm
  - More or less complex
  - SVMs, Naive Bayes, Deep Neural Network, etc.
  - PageRank, Ant Colony algorithms, genetic algorithms,etc.
- Several common parameters are involved:
  - Context : window, phrase, sentence, text,...
  - Depth in a graph

Resources

- In WSD, we consider two kinds of resources
  - Knowledge
    - Machine readable dictionaries
    - Lexical Databases
    - Encyclopedias
  - Corpus
    - Non-sense-annotated corpus
    - Sense-annotated corpus

Resources : knowledge

- Machine readable dictionaries
  - Longman, Oxford Advanced Learner's dictionary,...
  - Until the 1990's for English
- Lexical Databases
  - WordNet from the 1990's [Miller]
  - BabelNet [Navigli, 2012]
- Encyclopedias
  - Wikipedia from 2007 [Mihalcea, 2007]

Resources: non-sense-annotated corpora

- A set of texts
- Covers one or more domains
- One or more languages
- Up to dozens of millions of words
- Can be lemmatized and tagged with part of speech information
- Various sources :
  - Newspapers, books, encyclopedias, Web,...

Resources: sense-annotated corpora

- SemCor [Miller et al., 1993]
- Subset of the Brown Corpus (1961)
  - 700,000 words
  - 30,000 words manually tagged with Wordnet synsets
  - 352 texts
    - For 186 texts, nouns, verbs, adjectives, and adverbs tagged : 192,639 words
    - For 166, only verbs are tagged : 41,497 words

Resources: sense-annotated corpora

- The Defense Science Organisation corpus [Ng & Lee, 1996]
  - Non-freely available sense- annotated English corpus
  - 192800 word occurences manually tagged with WordNet synsets
  - Annotations cover
    - 121 nouns (113,000 occurences)
    - 70 verbs (79,800 occurences)
  - The most frequent, as ambiguous possible.
  - Coverage corresponding to 20% of verb and noun occurences in English texts

Resources: Sense-annotated corpora

- Corpora from evaluation campaigns
  - Most of them in English
  - But also on Japanese, Spanish, Chinese
  - Uncommonly beyond 5000 tagged words
- Other languages:
  - Dutch SemCor [Vossen et al., 2012]
    - 250,000 manually tagged words
  - Basque SemCor [Agirre, 2006]

Sense-annotated corpora : limitations

- Really difficult task compared to other annotation tasks
- Penn Treebank [Taylor et al., 2003]
  - Part of speech tagged corpus
  - Only 45 possible tags
  - 3000 annotations per hour
- WordNet synset-annotated corpus
  - 117,000 possible tags
  - Example for the Defense Science Organisation corpus
    - 191 different nouns, 1800 possible tags
    - 1 man-year for 192000 word occurrences 150-250 annotations per hour

Sense-annotated corpora : limitations

- Have to be repeated for
  - each sense inventory;
  - each language;
  - each domain.

Mitigating the limitations

- Improving annotation speeds
  - [Mihalcea & Chklovski, 2003] WSD algorithm on corpus -> Then human verification
  - Not much improvment
- Usage of new kinds of sense-annotated corpora
  - E.g. Wikipedia and its internal links [Mihalcea, 2007]
  - A page can be considered as a sense
- More languages
  - BabelCor

UFSAC: Unification of Sense Annotated Corpora and Tools [Vial et al., 2018]

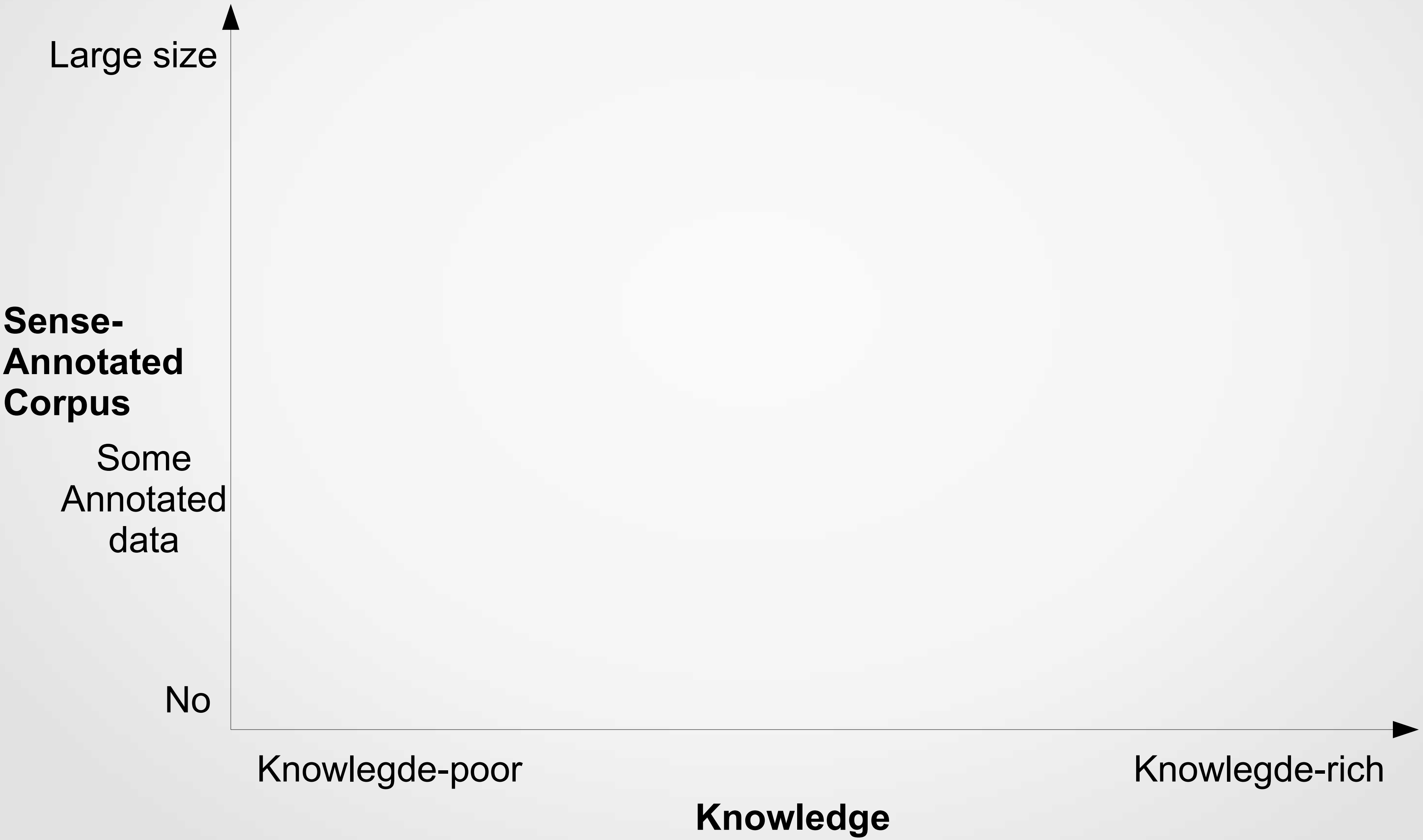
- In English, there are a dozen of manually annotated sense annotated corpora, but their file formats are very different from one another.
- Unification of these corpora in a format
  - easy to use
  - Easy to understand
- Facilitate
  - the creation of new WSD systems
  - the evaluation of existing ones

https://github.com/getalp/UFSAC

UFSAC: Unification of Sense Annotated Corpora and Tools [Vial et al., 2018]

Corpus	Sentences	Words		Annotated parts of speech			
		Total	Annotated	Nouns	Verbs	Adj.	Adv.
SemCor [7]	37176	778587	229517	87581	89037	33751	19148
DSO [11]	178119	5317184	176915	105925	70990	0	0
WordNet GlossTag [6]	117659	1634691	496776	232319	62211	84233	19445
MASC [4]	34217	596333	114950	49263	40325	25016	0
OMSTI [14]	820557	35843024	920794	476944	253644	190206	0
Ontonotes [3]	21938	435340	52263	9220	43042	0	0
Senseval 2 [2]	238	5589	2301	1061	541	422	277
Senseval 3 task 1 [13]	300	5511	1957	886	723	336	12
SemEval 2007 task 07 [10]	245	5637	2261	1108	591	356	206
SemEval 2007 task 17 [12]	120	3395	455	159	296	0	0
SemEval 2013 task 12 [9]	306	8142	1644	1644	0	0	0
SemEval 2015 task 13 [8]	138	2638	1053	554	251	166	82

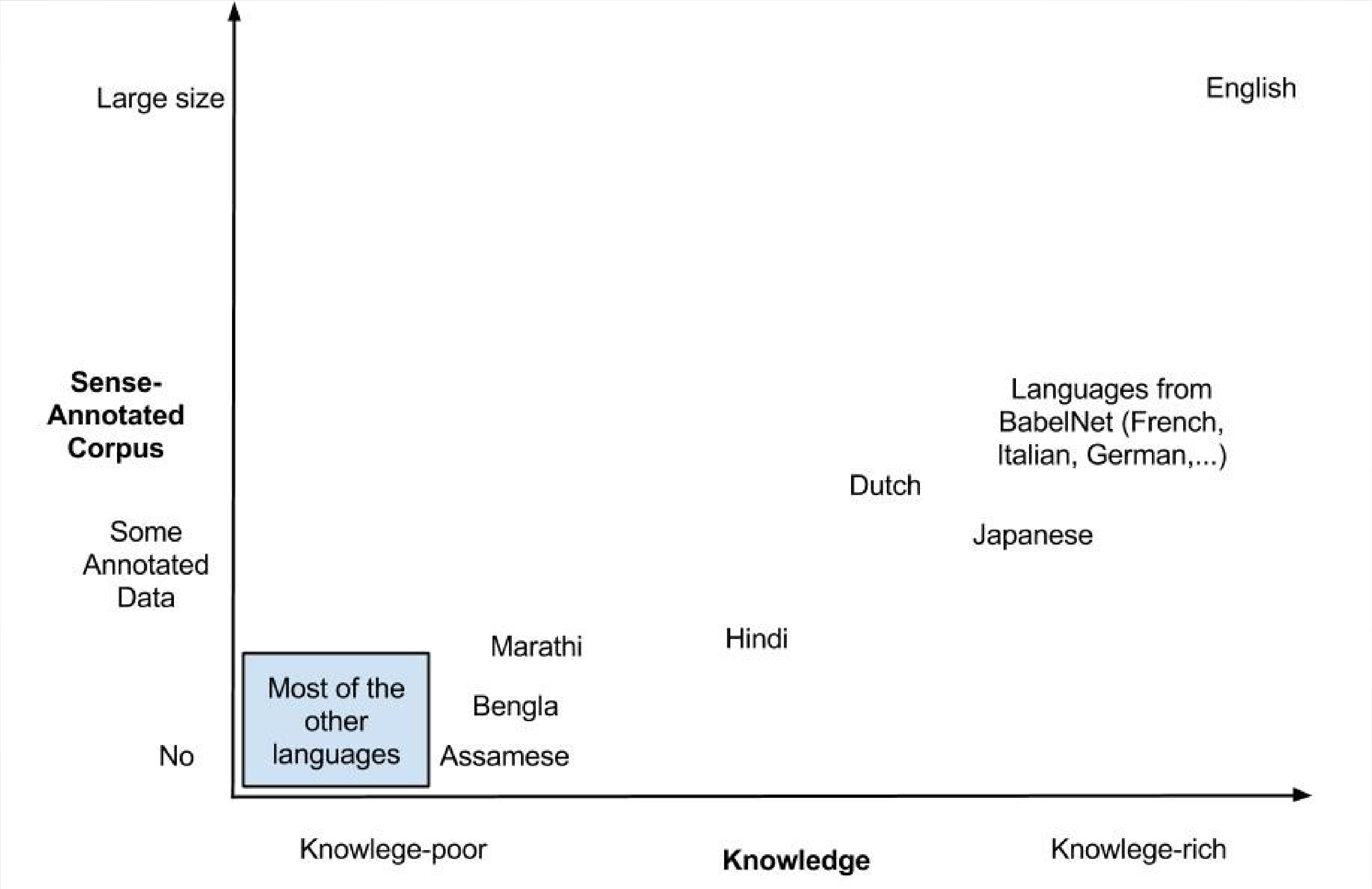
Analysis of resources for WSD



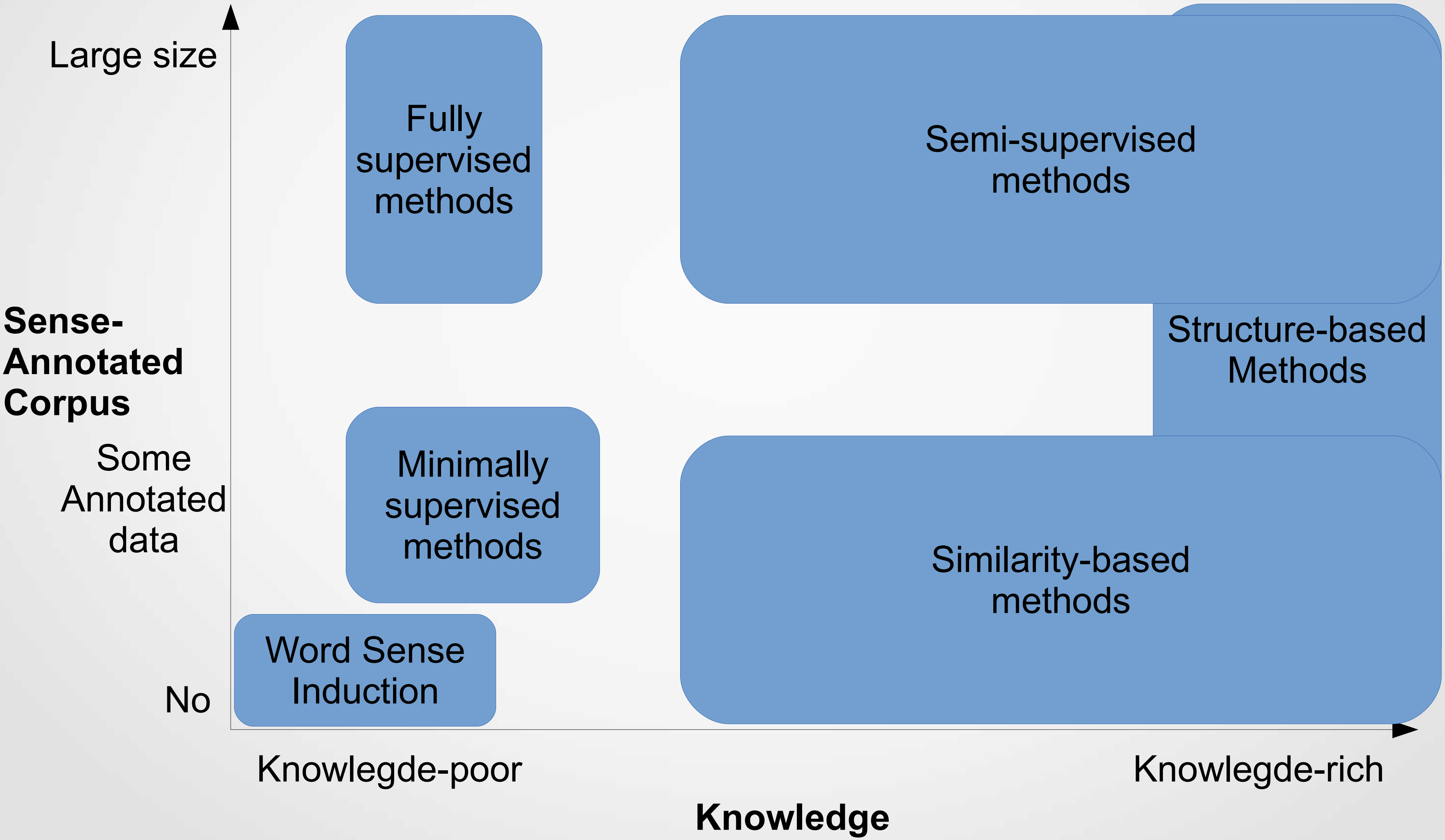
Analysis of resources for WSD

Analyse of resources for WSD

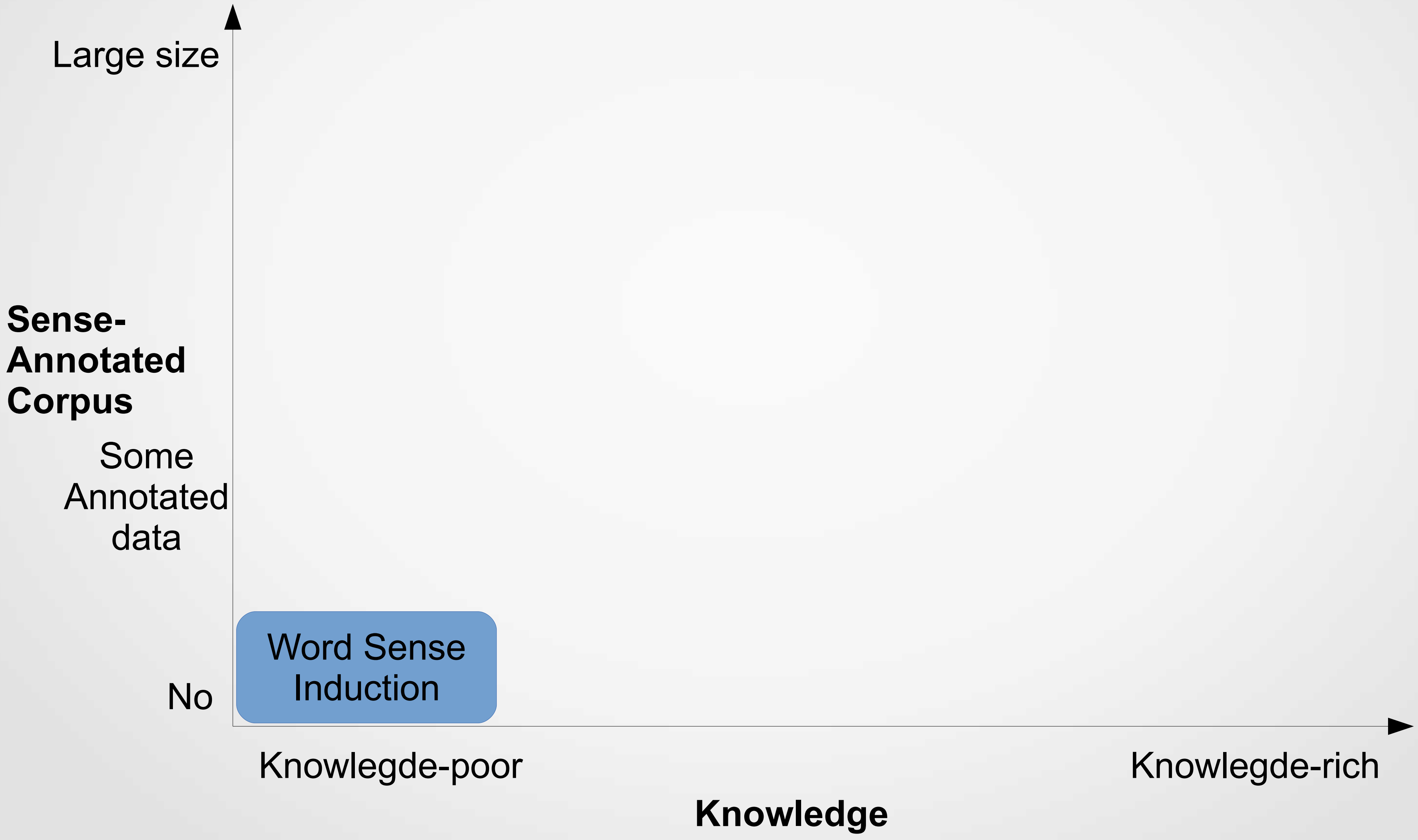
Languages Resources Available



WSD Approaches



WSD Approaches



Word Sense Disambiguation Process

- Composed of 3 steps
  - Build/select of raw lexical material(s)
  - Build an elaborate resource
  - Use that resource to lexically disambiguate a text

Word Sense induction (WSI)

- Word Sense induction (or discrimination)
- Build/select raw lexical material(s)
  - Only raw (no sense annotations) corpora
- Build an elaborate resource
  - Induce word senses from contexts
- Use that resource to lexically disambiguate a text
  - Open

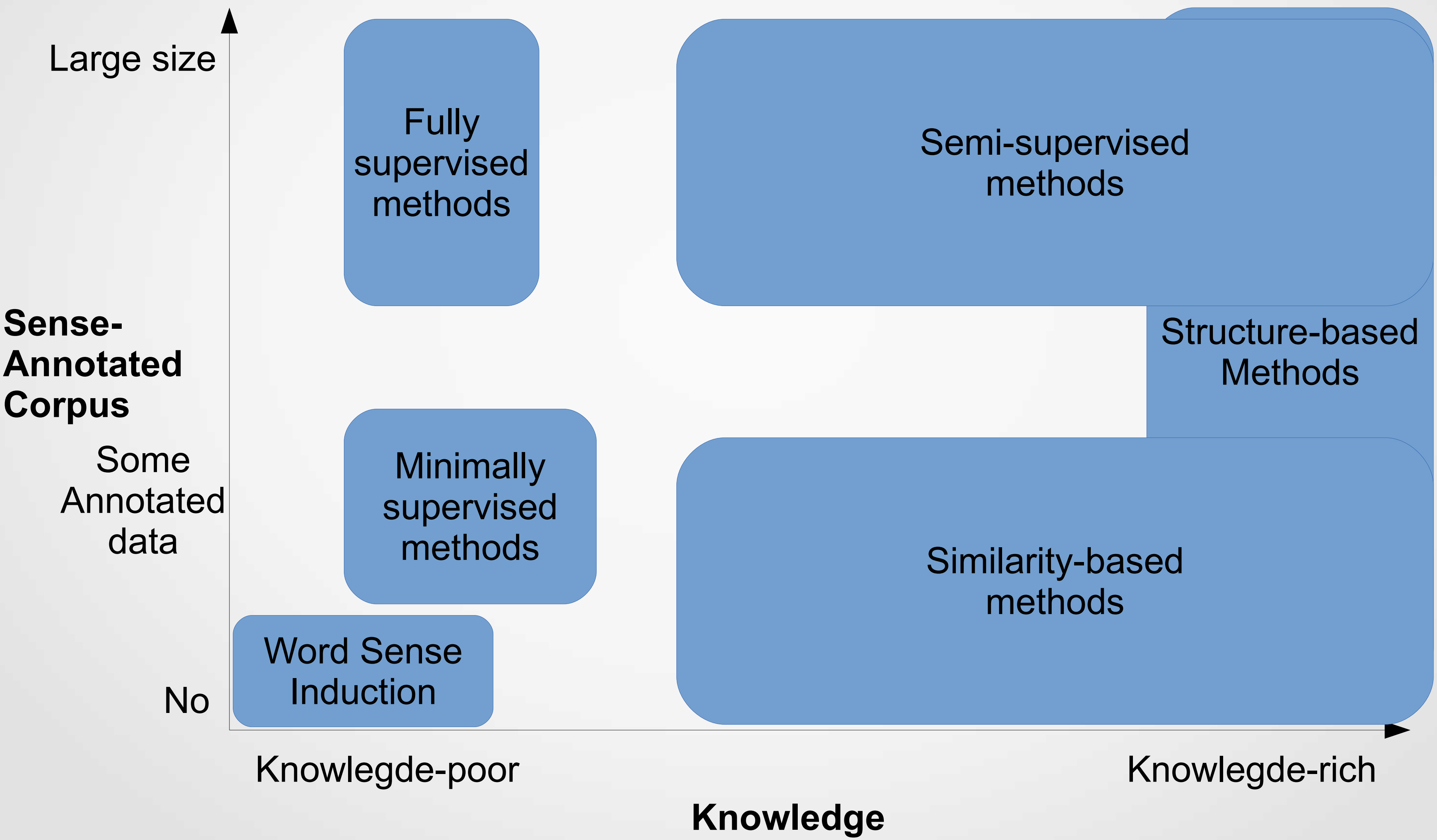
WSI : Build an elaborate resource

- Use only raw corpora
- Induce word senses from contexts
- Harris' (1954) Distributional semantics principle -
  - Hypothesis : the meaning of a word comes from its context
- Example:
  - „The mouse is eating cheese“, „The cat is hunting a mouse“
  - „The mouse is linked to the computer“,“my mouse is broken“

WSI : Build an elaborate resource

- Induce word senses from input text by clustering word occurrences
- Computational representation:
  - Vectors, Bag of words
- Clustering algorithms : Kmean,...
- Graphs: each node is a word and edges are coocurences, senses are given by identification of hubs (clusters)

WSD Approaches



Useful heuristics

- Based on observations
- One sense per discourse [Gale *et al.*, 1991]
- One sense per collocation [Yarowsky, 1993]

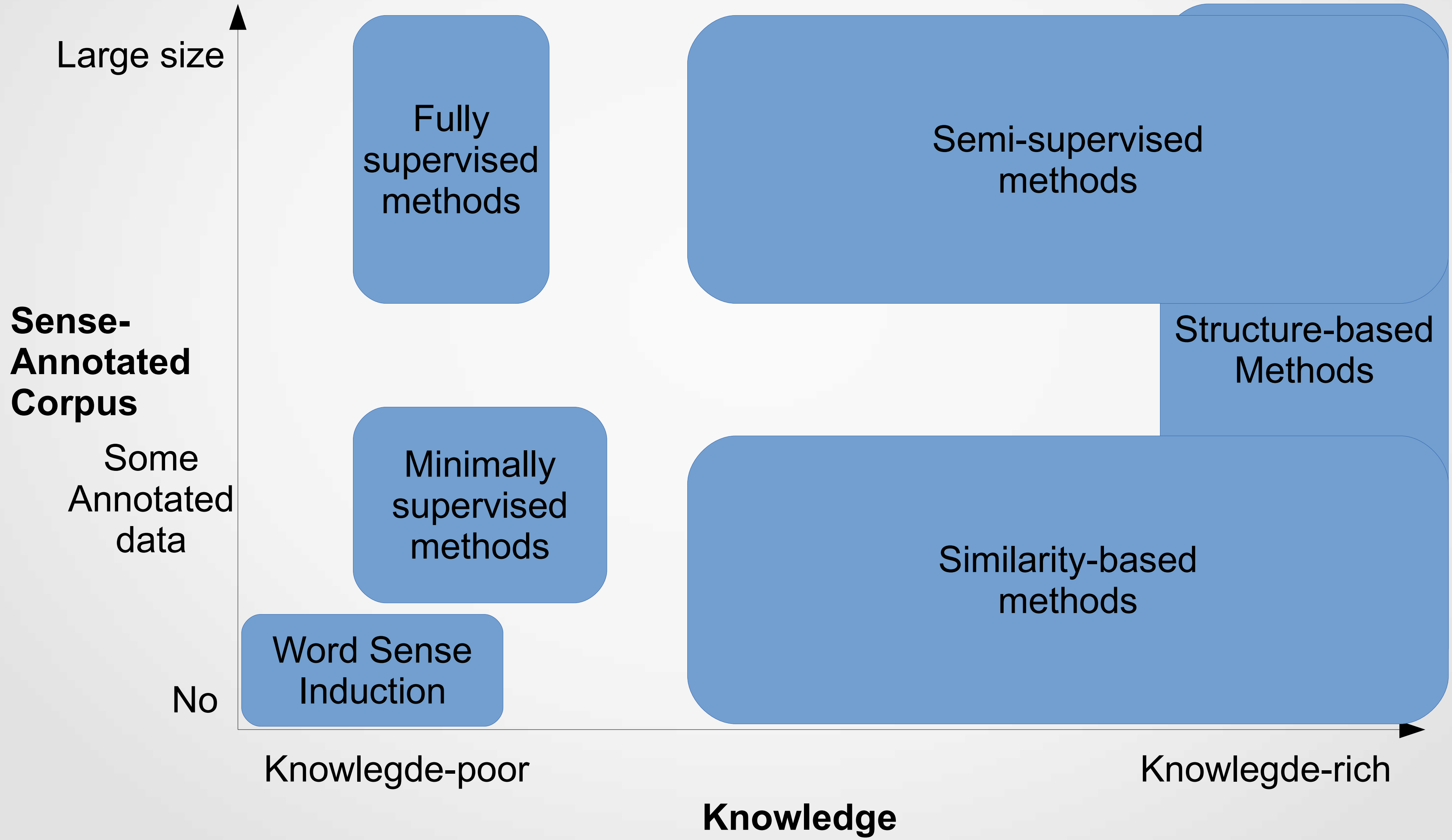
One sense per discourse [Gale et al., 1991]

- Random sample of 108 nouns
- 300 articles studied
- 3 judges
- Only 6 articles judged to contain multiple senses of one of the test words
- Tendency to share senses in the same discourse extremely strong: 98%

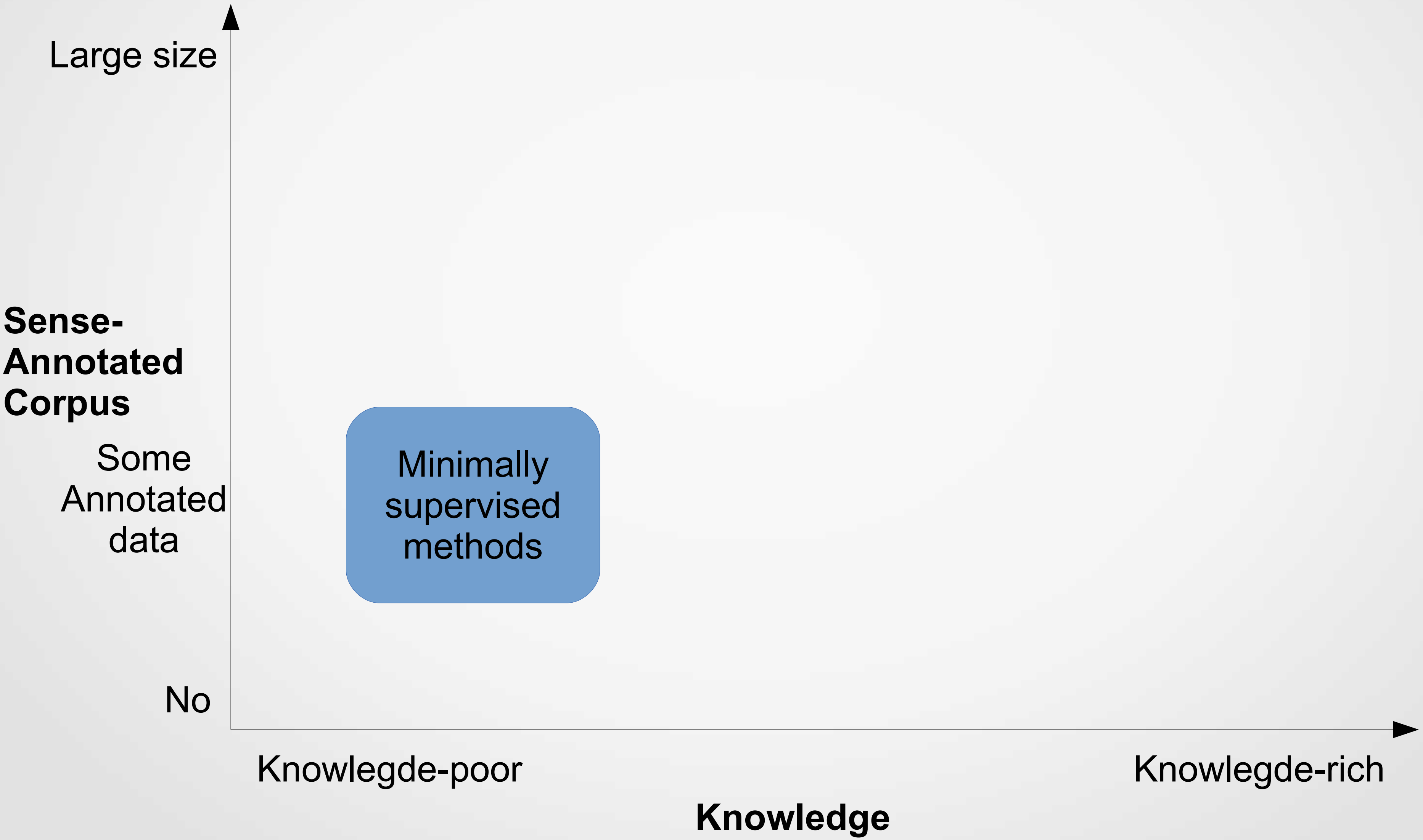
One Sense per Collocation [Yarowsky, 1993]

- Collocation : sequence of words or terms that co-occur more often than would be expected by chance
- Types of collocations:
  - adjective+noun : *peur bleue, strong fever*
  - noun+noun (such as collective nouns): meute de loups, douzaine d'œufs, wolf pack, dozen egg
  - verb+noun: prendre une gifle, prendre l'escalier, chair a meeting, conduct an experiment
- 90% to 99% for an average of 95% share senses in texts

WSD Approaches



WSD Approaches



Word Sense Disambiguation Process

- Composed of 3 steps
  - Build/select of raw lexical material(s)
  - Build an elaborate resource
  - Use that resource to lexically disambiguate a text

Minimally-Supervised WSD

- Build/select raw lexical material(s)
  - Some sense-annotated data
  - Raw corpora
- Build an elaborate resource
  - Induce word senses from evidence in texts
  - Learn one classifier per word
- Use that resource to lexically disambiguate a text
  - Use classifiers to find the best sense for each word in texts

Decision List Algorithm

- Decision list [Rivest, 1987]
- Based on the *one sense per collocation* heuristic
- Collect a large set of collocations for ambiguous words
- Calculate the word-sense probability distributions for all such collocations
- Calculate the log-likelihood ratio

$$\log\left(\frac{P\left(Sense-A \mid Collocation_i\right)}{P\left(Sense-B \mid Collocation_i\right)}\right)$$

- Higher log-likelihood $\Rightarrow$ more predictive evidence
- Collocations ordered in a decision list, with most predictive collocations ranked highest

Decision List Algorithm

Sense	Training Examples (Keyword in Context)
A	used to strain microscopic *plant* **life** from the ...
A	... zonal distribution of *plant* **life** . ...
A	close-up studies of *plant* **life** and natural ...
A	too rapid growth of aquatic *plant* **life** in water ...
A	... the proliferation of *plant* and animal **life** ...
A	establishment phase of the *plant* virus **life** cycle ...
A	... that divide **life** into *plant* and animal kingdom
A	... many dangers to *plant* and animal **life** ...
A	mammals . Animal and *plant* **life** are delicately
A	beds too salty to support *plant* **life** . River ...
A	heavy seas, damage , and *plant* **life** growing on ...
A	... ...
B	... ...
B	automated **manufacturing** *plant* in Fremont ...
B	... vast **manufacturing** *plant* and distribution ...
B	chemical **manufacturing** *plant* , producing viscose
B	... keep a **manufacturing** *plant* profitable without
B	computer **manufacturing** *plant* and adjacent ...
B	discovered at a St. Louis *plant* **manufacturing**
B	... copper **manufacturing** *plant* found that they
B	copper wire **manufacturing** *plant* , for example ...
B	's cement **manufacturing** *plant* in Alpena ...
B	polystyrene **manufacturing** *plant* at its Dow ...
B	company **manufacturing** *plant* is in Orlando ...

Decision List Algorithm

Sense	Training Examples (Keyword in Context)
A	used to strain microscopic *plant* life from the ...
A	... zonal distribution of *plant* **life** . ...
A	close-up studies of *plant* **life** and natural ...
A	too rapid growth of aquatic *plant* **life** in water ...
A	... the proliferation of *plant* and animal **life** ...
A	establishment phase of the *plant* virus **life** cycle ...
A	... that divide **life** into *plant* and animal kingdom
A	... many dangers to *plant* and animal **life** ...
A	mammals . Animal and *plant* **life** are delicately
A	beds too salty to support *plant* **life** . River ...
A	heavy seas, damage , and *plant* **life** growing on ...
A	... ...
B	... ...
B	automated **manufacturing** *plant* in Fremont ...
B	... vast **manufacturing** *plant* and distribution ...
B	chemical **manufacturing** *plant* , producing viscose
B	... keep a **manufacturing** *plant* profitable without
B	computer **manufacturing** *plant* and adjacent ...
B	discovered at a St. Louis *plant* **manufacturing**
B	... copper **manufacturing** *plant* found that they
B	copper wire **manufacturing** *plant* , for example ...
B	's cement **manufacturing** *plant* in Alpena ...
B	polystyrene **manufacturing** *plant* at its Dow ...
B	company **manufacturing** *plant* is in Orlando ...

Initial decision list for *plant* (abbreviated)

LogL	Collocation	Sense
8.10	*plant* **life**	$\Rightarrow$ A
7.58	**manufacturing** *plant*	$\Rightarrow$ B
7.39	**life** (within $\pm$2-10 words)	$\Rightarrow$ A
7.20	**manufacturing** (in $\pm$2-10 words)	$\Rightarrow$ B
6.27	animal (within $\pm$2-10 words)	$\Rightarrow$ A
4.70	equipment (within $\pm$2-10 words)	$\Rightarrow$ B
4.39	employee (within $\pm$2-10 words)	$\Rightarrow$ B
4.30	assembly *plant*	$\Rightarrow$ B
4.10	*plant* closure	$\Rightarrow$ B
3.52	*plant* species	$\Rightarrow$ A
3.48	automate (within $\pm$2-10 words)	$\Rightarrow$ B
3.45	microscopic *plant*	$\Rightarrow$ A
	...	

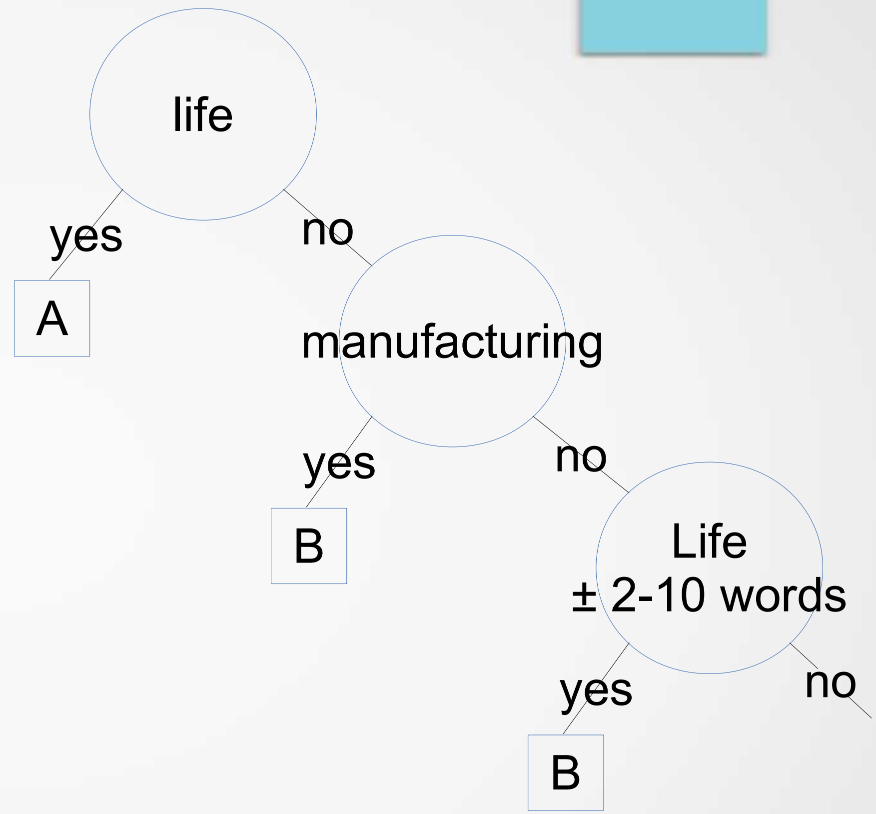
Decision List Algorithm

Initial decision list for *plant* (abbreviated)		
LogL	Collocation	Sense
8.10	***plant* life**	⇒ A
7.58	**manufacturing *plant***	⇒ B
7.39	**life** (within ±2-10 words)	⇒ A
7.20	**manufacturing** (in ±2-10 words)	⇒ B
6.27	animal (within ±2-10 words)	⇒ A
4.70	equipment (within ±2-10 words)	⇒ B
4.39	employee (within ±2-10 words)	⇒ B
4.30	assembly ***plant***	⇒ B
4.10	***plant*** closure	⇒ B
3.52	***plant*** species	⇒ A
3.48	automate (within ±2-10 words)	⇒ B
3.45	microscopic ***plant***	⇒ A
	...	

Decision List Algorithm

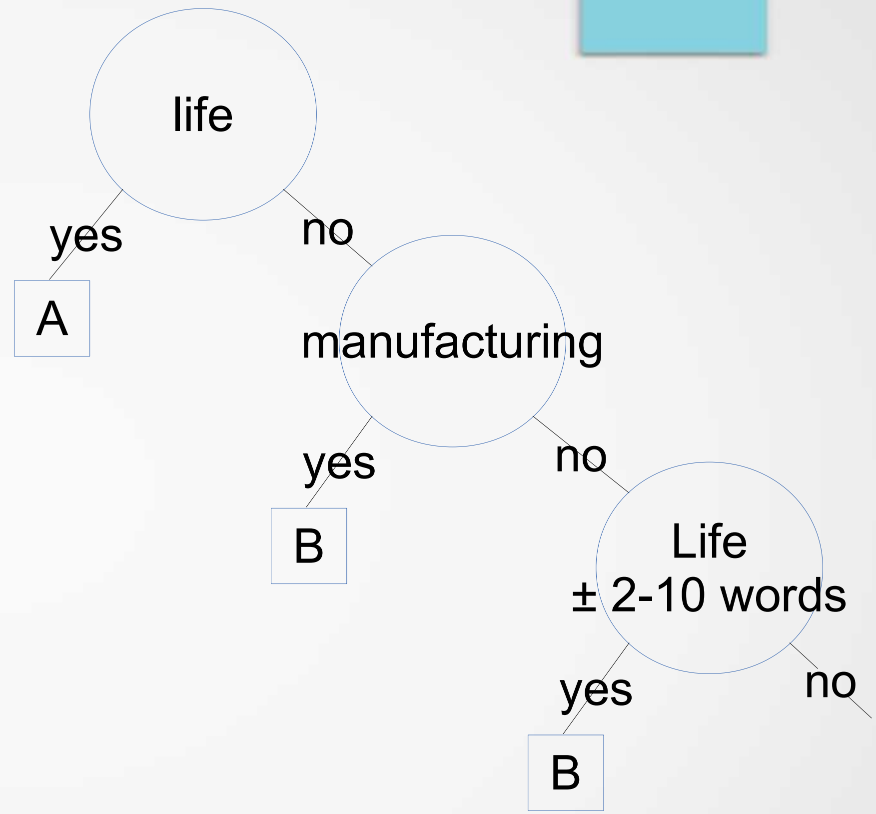
Initial decision list for *plant* (abbreviated)

LogL	Collocation	Sense
8.10	*plant* life	⇒ A
7.58	manufacturing *plant*	⇒ B
7.39	life (within ±2-10 words)	⇒ A
7.20	manufacturing (in ±2-10 words)	⇒ B
6.27	animal (within ±2-10 words)	⇒ A
4.70	equipment (within ±2-10 words)	⇒ B
4.39	employee (within ±2-10 words)	⇒ B
4.30	assembly *plant*	⇒ B
4.10	*plant* closure	⇒ B
3.52	*plant* species	⇒ A
3.48	automate (within ±2-10 words)	⇒ B
3.45	microscopic *plant*	⇒ A
	...	



Decision List Algorithm

Initial decision list for *plant* (abbreviated)		
LogL	Collocation	Sense
8.10	*plant* life	⇒ A
7.58	manufacturing *plant*	⇒ B
7.39	life (within ±2-10 words)	⇒ A
7.20	manufacturing (in ±2-10 words)	⇒ B
6.27	animal (within ±2-10 words)	⇒ A
4.70	equipment (within ±2-10 words)	⇒ B
4.39	employee (within ±2-10 words)	⇒ B
4.30	assembly *plant*	⇒ B
4.10	*plant* closure	⇒ B
3.52	*plant* species	⇒ A
3.48	automate (within ±2-10 words)	⇒ B
3.45	microscopic *plant*	⇒ A
	...	

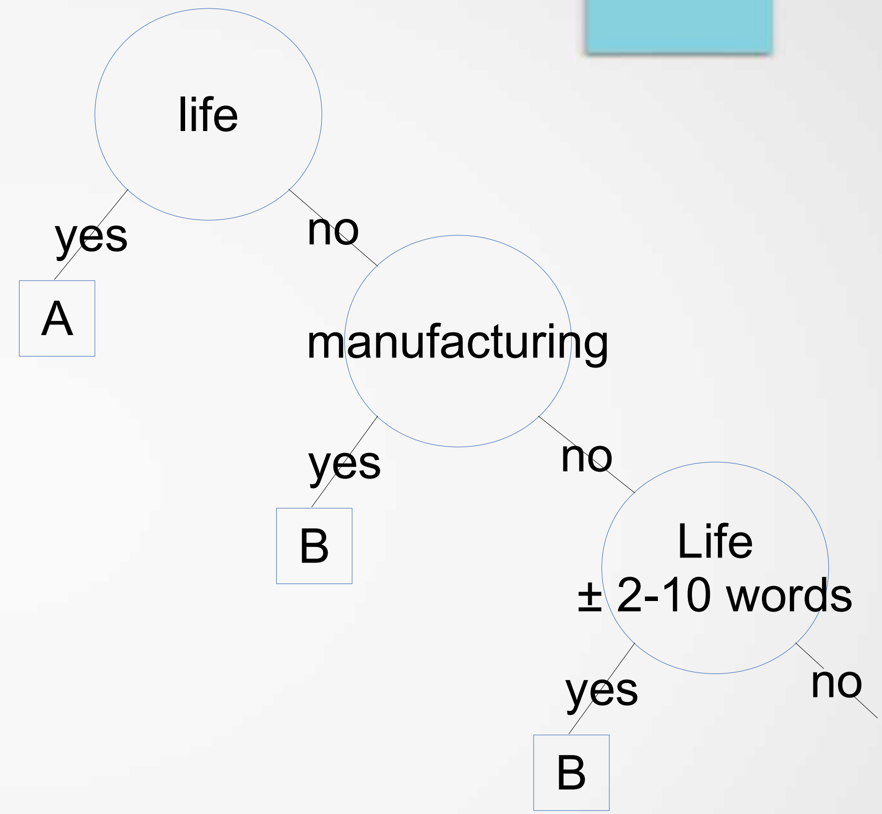


Numerous animals have coevolved with plants. Many animals pollinate flowers in exchange for food in the form of pollen or nectar.

Decision List Algorithm

Initial decision list for *plant* (abbreviated)

LogL	Collocation	Sense
8.10	***plant* life**	⇒ A
7.58	**manufacturing *plant***	⇒ B
7.39	life (within ±2-10 words)	⇒ A
7.20	manufacturing (in ±2-10 words)	⇒ B
6.27	animal (within ±2-10 words)	⇒ A
4.70	equipment (within ±2-10 words)	⇒ B
4.39	employee (within ±2-10 words)	⇒ B
4.30	assembly *plant*	⇒ B
4.10	*plant* closure	⇒ B
3.52	*plant* species	⇒ A
3.48	automate (within ±2-10 words)	⇒ B
3.45	microscopic *plant*	⇒ A
	...	



Numerous animals have coevolved with plants. Many animals pollinate flowers in exchange for food in the form of pollen or nectar.

Minimally-Supervised WSD

- Build/select raw lexical material(s)
  - A sense inventory from MRD, Lexical bases, encyclopedias
  - Some sense-annotated data
  - Raw corpora
- Build an elaborate resource
  - Induce word senses from evidence in text
  - Learn one classifier per word
- Use that resource to lexically disambiguate a text
  - Use classifiers to find the best sense for each word in text

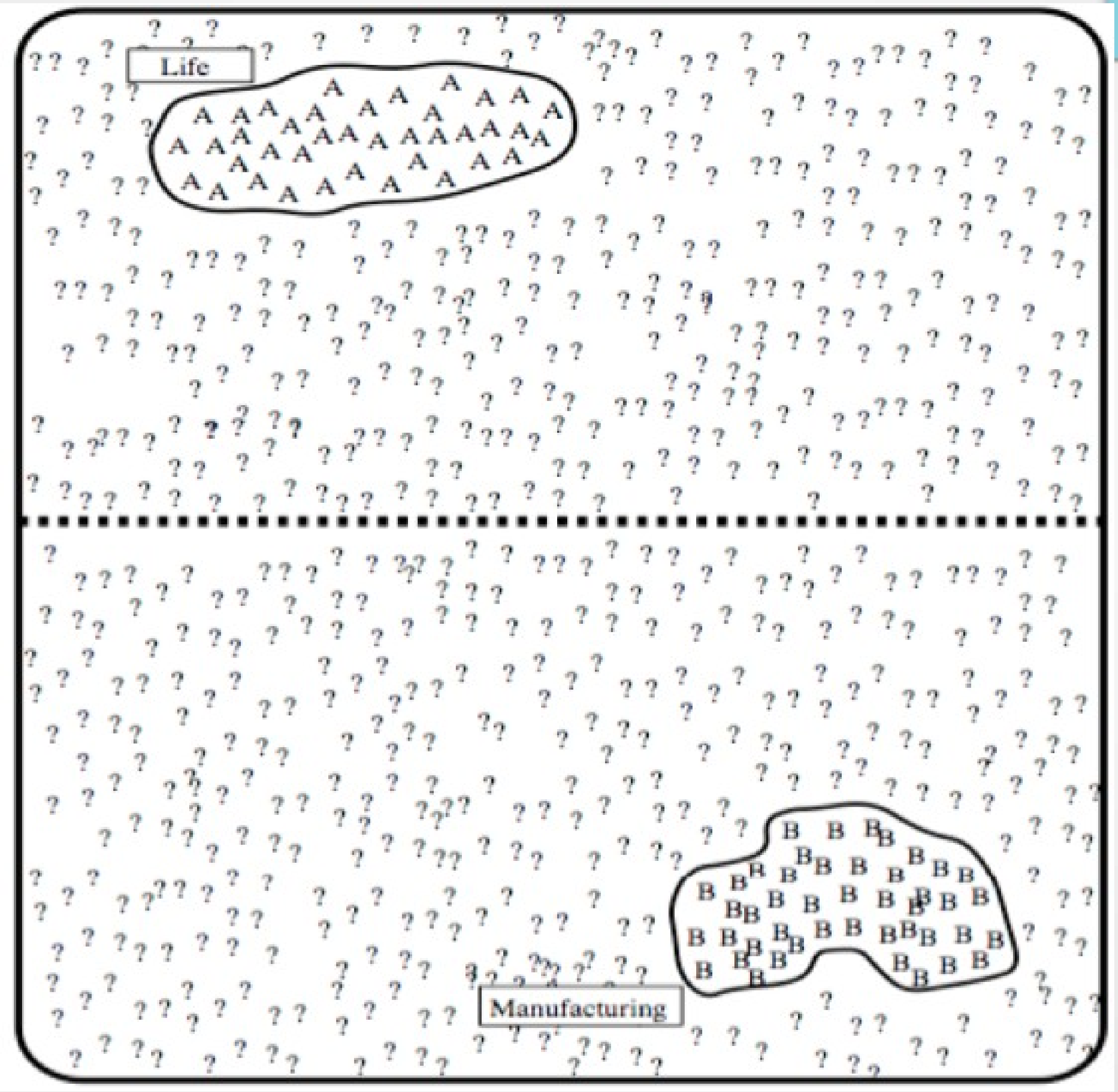
Yarowsky's method [Yarowsky, 1995]

- Annotations are expensive
- Bootstrapping
  - Small annotated corpus
  - Build the (small) corresponding decision list (high precision, low recall)
- One sense per collocation heuristic
- One sense per discourse heuristic
- Repeat
  - Label unannotated data with the decision list
  - Build a new decision list
  - Until decision list doesn't increase
- Lower precision, higher recall

Yarowsky's Method: example

- Disambiguating plant (industrial sense) vs. plant (living thing sense)
- 7538 occurences of *plant* in the 460 million-word corpus
- Annotation of seed features for each sense
  - 'Industrial sense': co-occurring with ‘manufacturing’ (1.1%)
  - 'Living thing' sense: co-occurring with ‘life’ (1.4%)
- Use *one sense per collocation* to build initial decision list classifier

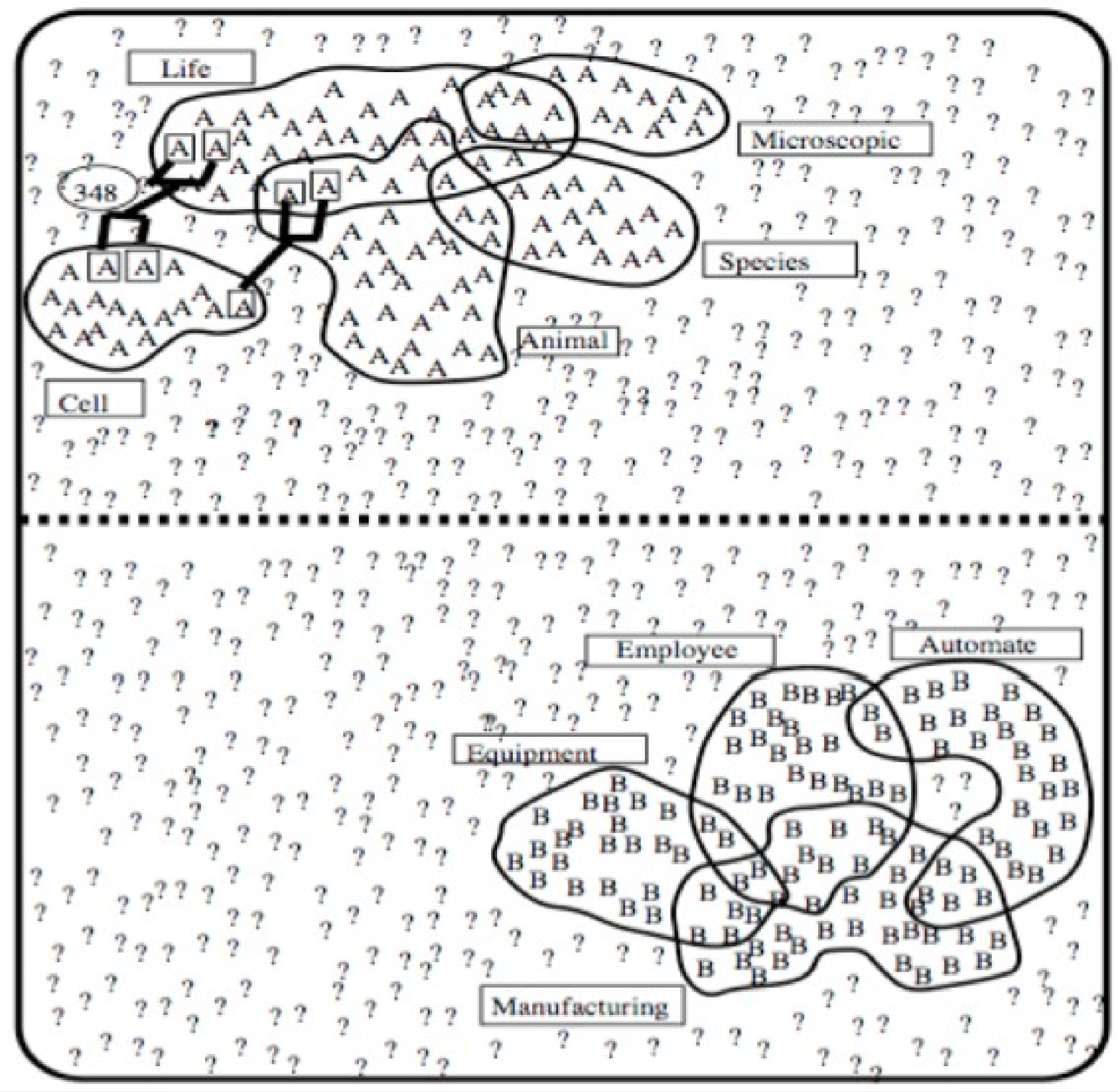
Yarowsky's Method: example



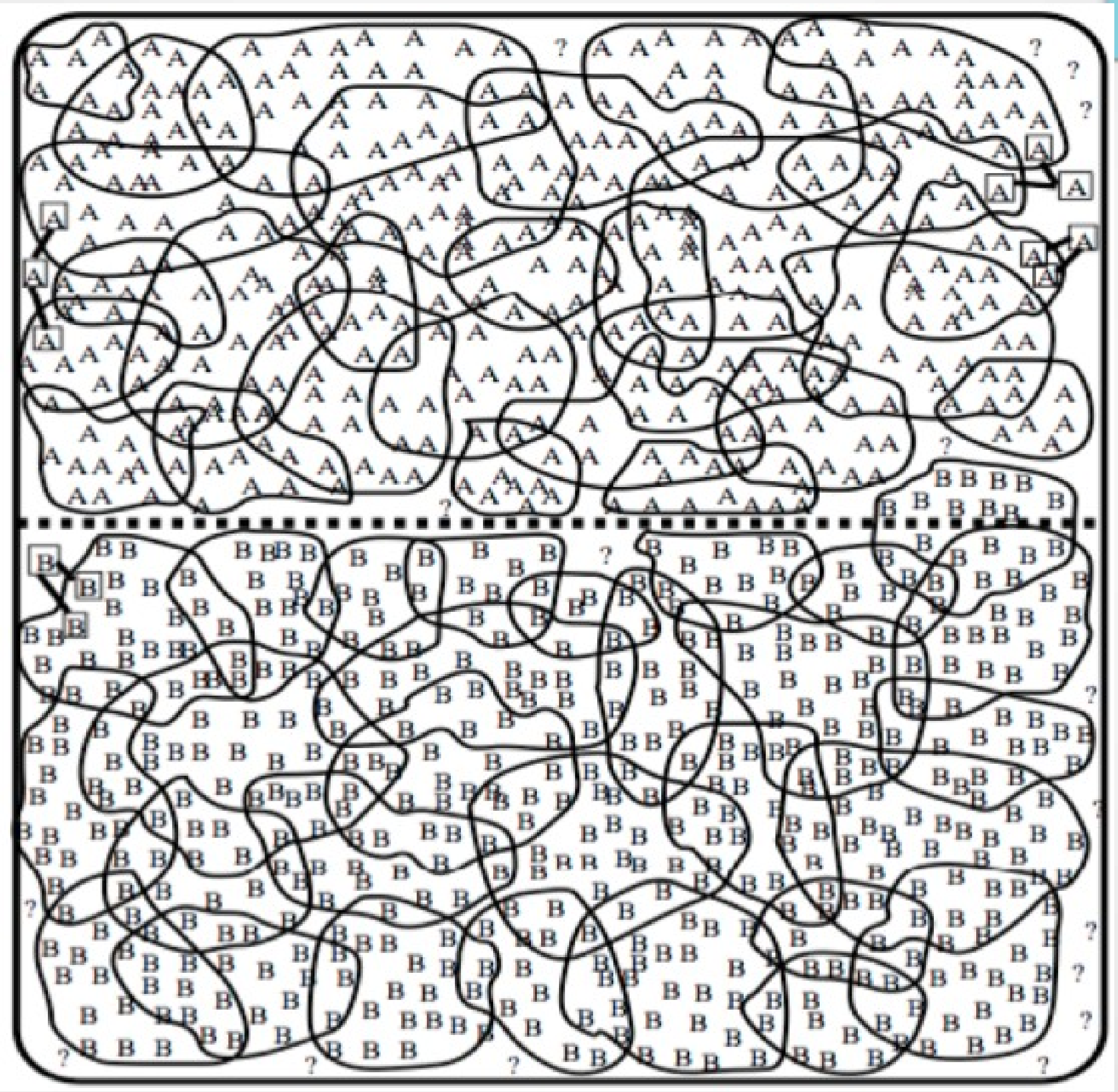
Yarowsky's Method: example

- Disambiguating plant (industrial sense) vs. plant (living thing sense)
- 7538 occurences of *plant* in the 460 million-word corpus
- Annotation of seed features for each sense
  - 'Industrial' sense: co-occurring with ‘manufacturing’ (1.1%)
  - 'Living thing' sense: co-occurring with ‘life’ (1.4%)
- Use *one sense per collocation* to build initial decision list classifier
- Treat results (having high probability) as annotated data, train new decision list classifier, iterate

Yarowsky's Method: example



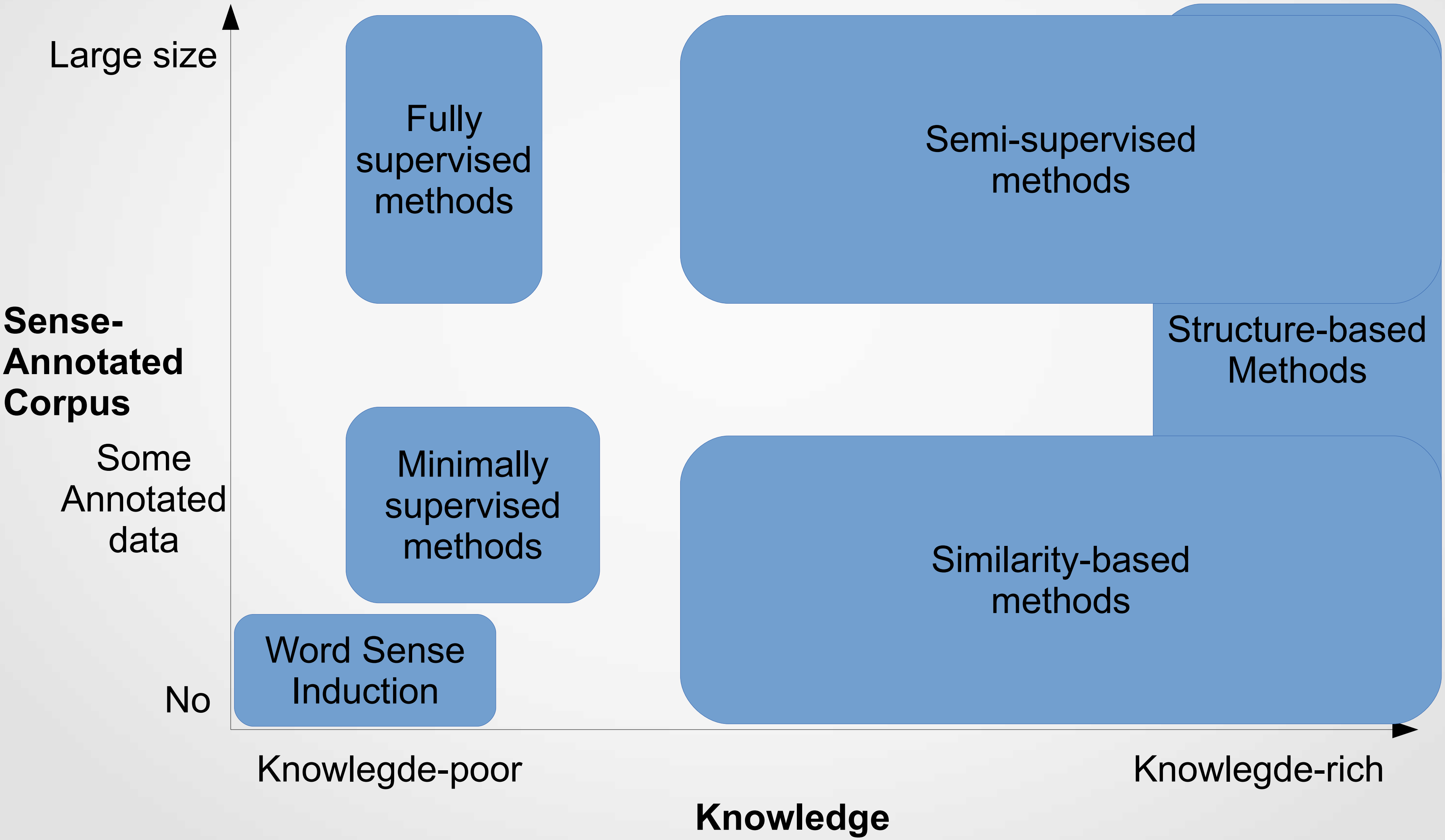
Yarowsky's Method: example



Performance of Yarowsky algorithm

- In 1995
  - Outperforms Schülze's unsupervised algorithm (1992)
  - Nearly same performance as supervised algorithms
- In 2009
  - Sánchez-de-Madariaga & Fernández-del-Castillo
  - Roughly homogeneous corpus: 95% F1
  - general text: about 70% F1 due to domain fluctuations

WSD Approaches



WSD Approaches

Word Sense Disambiguation Process

- Composed of 3 steps
  - Build/select raw lexical material(s)
  - Build an elaborate resource
  - Use that resource to lexically disambiguate a text

Structure-based WSD

- Build/select of raw lexical material(s)
  - Machine readable dictionaries, Lexical bases, Encyclopedias, Sense-annotated corpora,...
- Build an elaborate resource
  - Build a graph from senses and implicit links in raw lexical material
- Use that resource to lexically disambiguate a text
  - Use graph properties to disambiguate

Roberto Navigli's approaches

- Build/select raw lexical material(s)
  - WordNet
  - BabelNet
  - Extended WordNet, SemCor,...
- Build an elaborate resource
- Use that resource to lexically disambiguate a text

Roberto Navigli's approaches

Sense-annotated definitions of senses

- Build/select raw
  - WordNet
  - BabelNet
  - Extended WordNet, SemCor,...
- Build an elaborate resource
- Use that resource to lexically disambiguate a text

Roberto Navigli's approaches

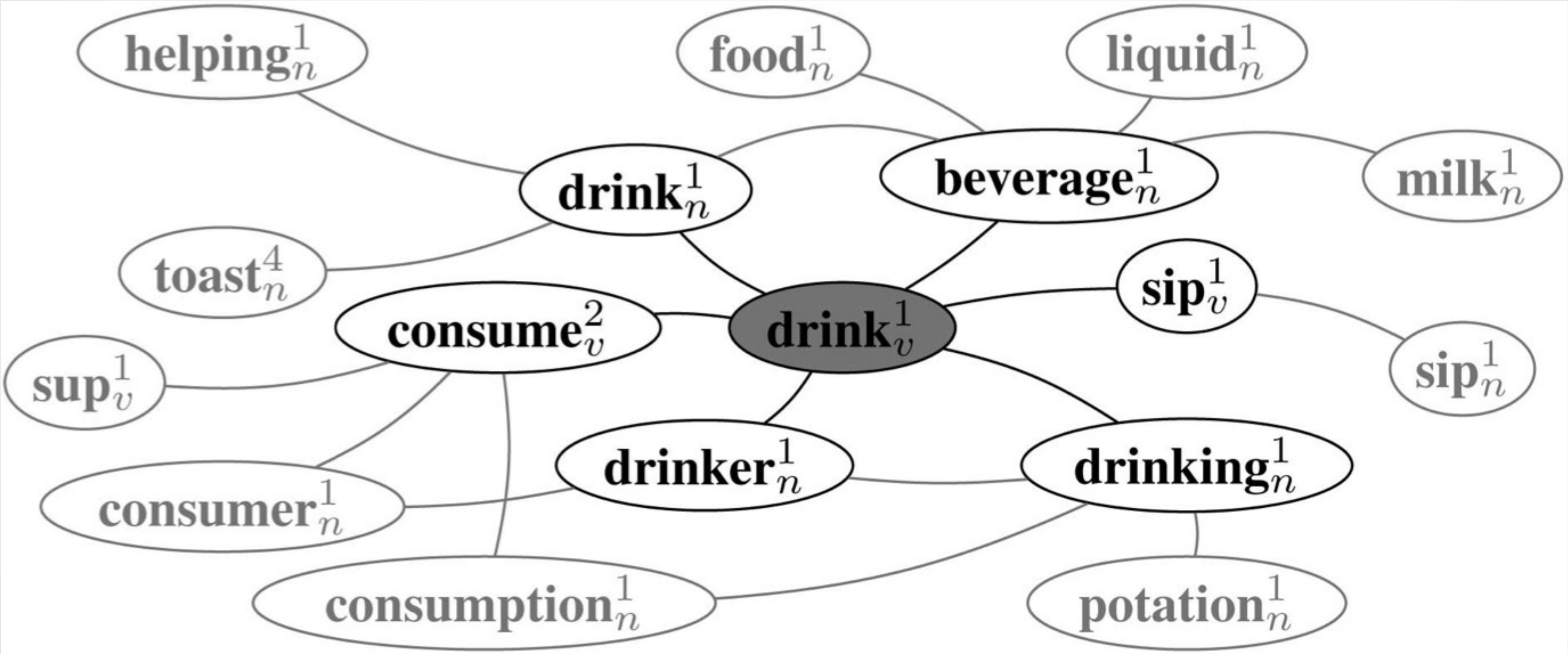
- Build/select raw lexical material(s)
  - WordNet
  - BabelNet
  - Extended WordNet, SemCor,...
- Build an elaborate resource
  - Each sense is a node
  - Add an edge between corresponding nodes if they share a relation
  - Add an edge between corresponding nodes for senses used in definitions
- Use that resource to lexically disambiguate a text

First Sense of the Noun *drink* in WordNet

Noun

- S: (n) **drink** (a single serving of a beverage) *"I asked for a hot drink"; "likes a drink before dinner"*
    - *direct hyponym* / ***full hyponym***
        - S: (n) chaser (a drink to follow immediately after another drink)
        - S: (n) draft, draught, potation, tipple (a serving of drink (usually alcoholic) drawn from a keg) *"they served beer on draft"*
            - S: (n) quaff (a hearty draft)
        - S: (n) pledge, toast (a drink in honor of or to the health of a person or event)
        - S: (n) libation ((facetious) a serving of an alcoholic beverage)
        - S: (n) eye opener (an alcoholic drink intended to wake one up early in the morning)
        - S: (n) nightcap (an alcoholic drink taken at bedtime; often alcoholic)
        - S: (n) hair of the dog (an alcoholic drink supposed to cure a hangover)
        - S: (n) shandygaff, shandy (a drink made of beer and lemonade)

Excerpt of an Elaborate Ressource



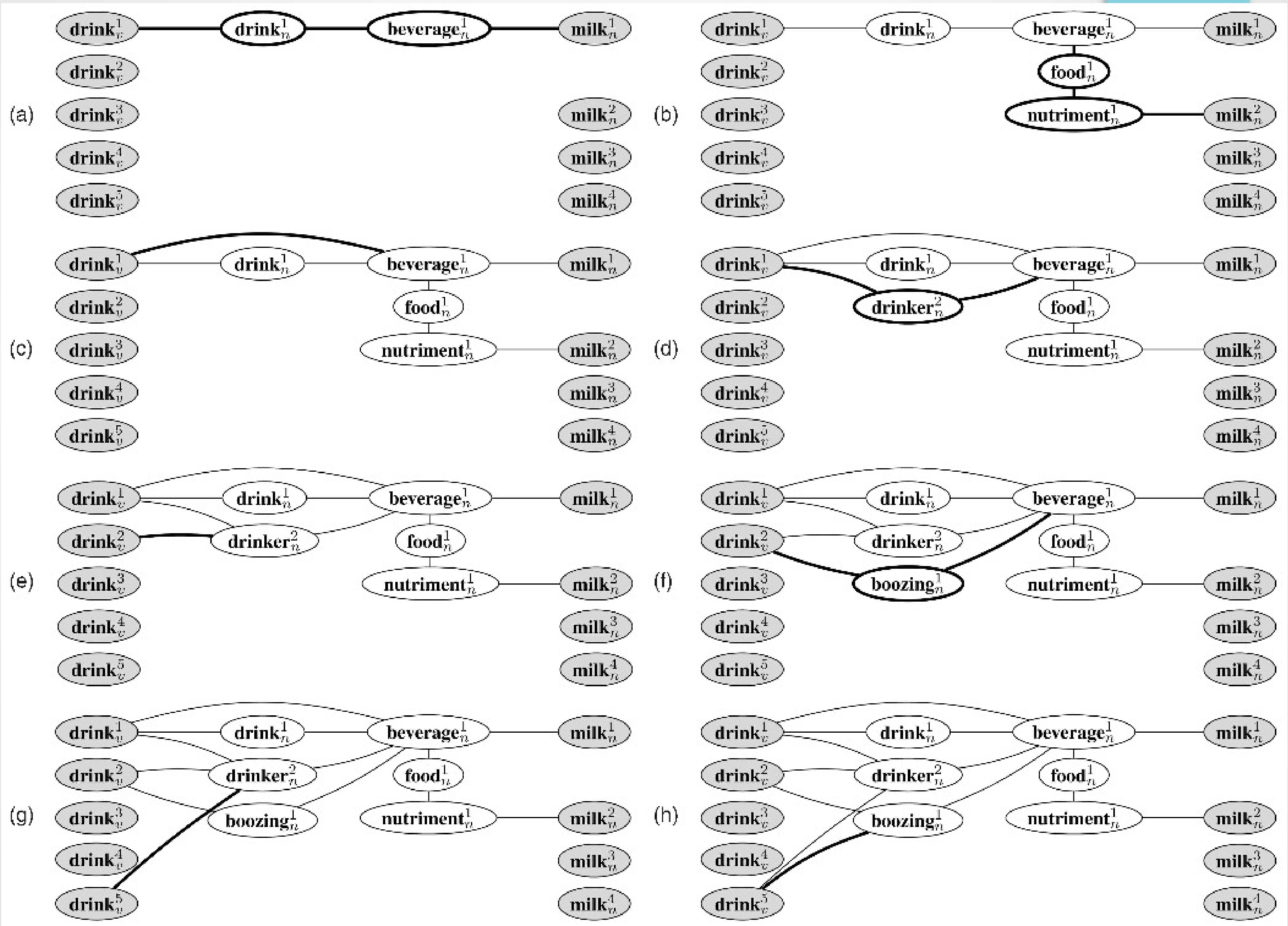
Roberto Navigli's approaches

- Build/select raw lexical material(s)
  - WordNet
  - BabelNet
  - Extended WordNet, SemCor,...
- Build an elaborate resource
  - Each sense is a node
  - Add an edge between corresponding nodes if they share a relation
  - Add an edge between corresponding nodes for senses used in definitions
- Use that resource to lexically disambiguate a text
  - Creation of a disambiguation Environment

Creation of a Disambiguation Environment

- *One sense per discourse* heuristic
- Subgraph of the elaborate ressource that includes all nodes in paths of length ≤ L connecting pairs of senses of words in context
- Principle:
  - Start from possible nodes
  - Perform a deep-search first in the elaborate resource
  - Until Level L+1
- Example : *„She drank some milk“ (drink$_v$, milk$_n$)*

Creation of a Disambiguation Environment



Roberto Navigli's approaches

- Build/select raw lexical material(s)
  - WordNet
  - BabelNet
  - Extended WordNet, SemCor,...
- Build an elaborate resource
  - Each sense is a node
  - Add an edge between corresponding nodes if they share a relation
  - Add an edge between corresponding nodes for senses used in definitions
- Use that resource to lexically disambiguate a text
  - Creation of disambiguation Environment
  - 4 different algorithms possible

Degree centrality (Degree)

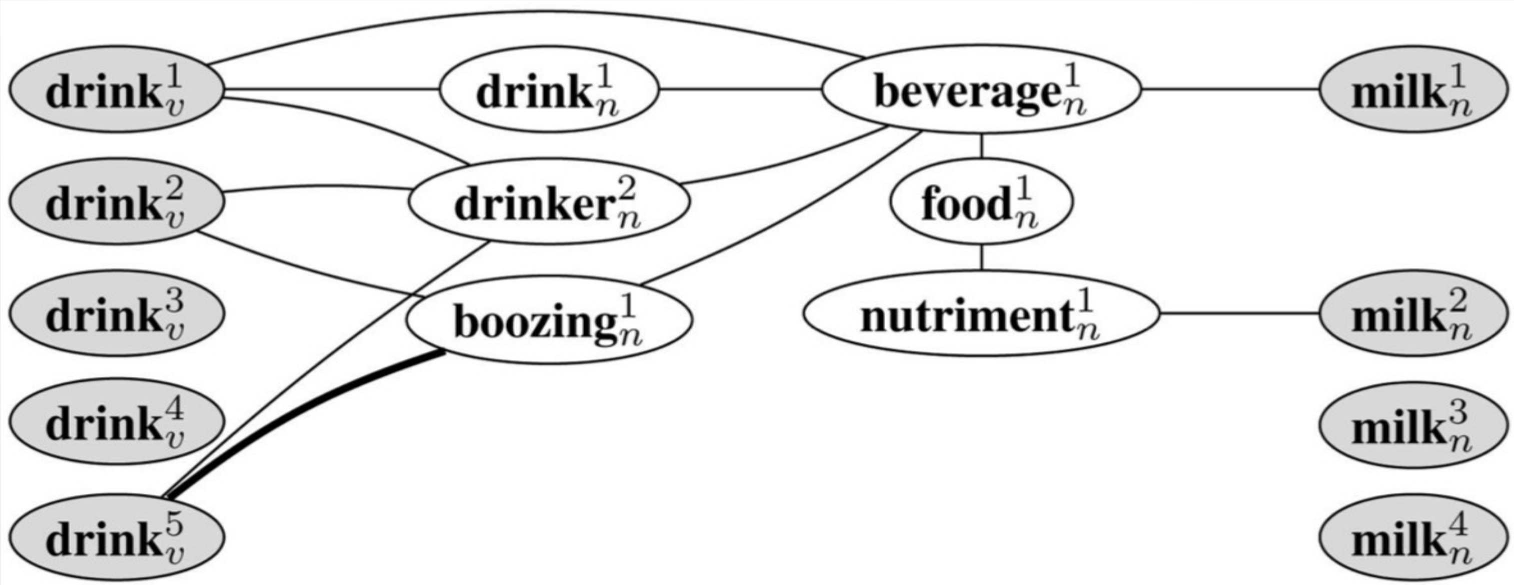
- Relies on the notion of vertex degree
- Score of a sense given by the number of their outgoing edges

$$score(s) = |\{(s, v) \in E : v \in V\}|$$

Degree centrality (Degree)

- Relies on the notion of vertex degree
- Score of a sense given by the number of their outgoing edges

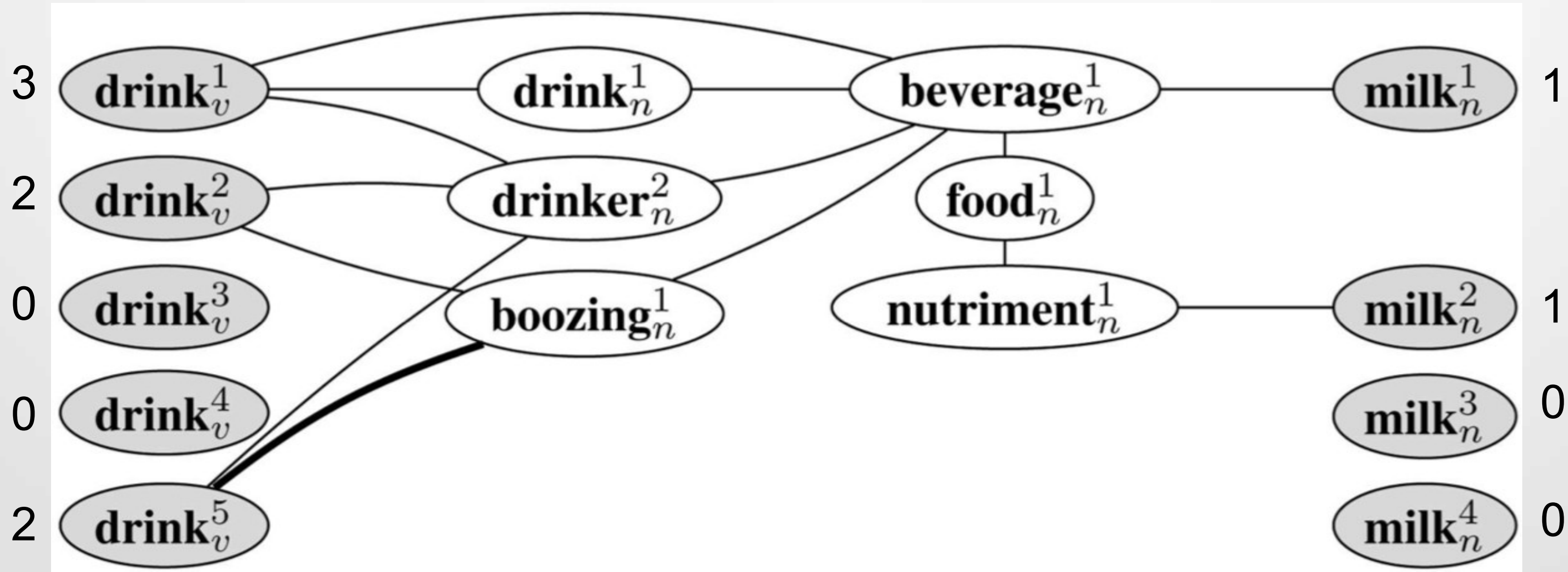
$$score(s) = \left| \{ (s, v) \in E : v \in V \} \right|$$



Degree centrality (Degree)

- Relies on the notion of vertex degree
- Score of a sense given by the number of their outgoing edges

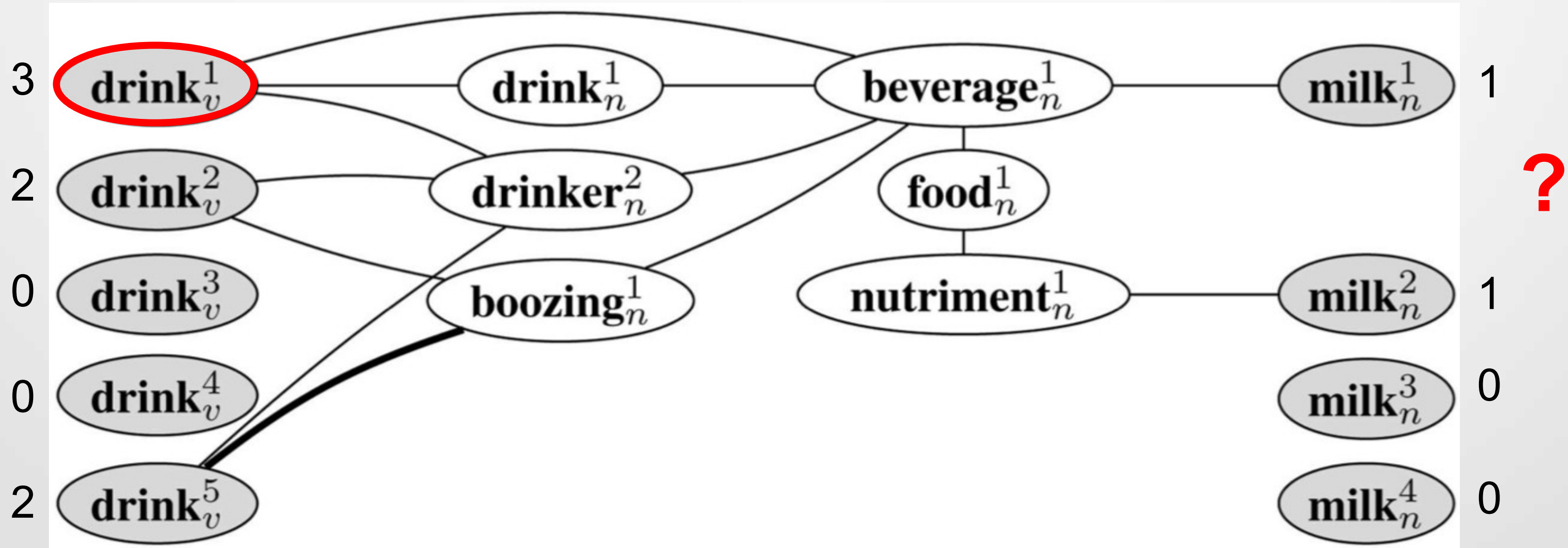
$$score(s) = |\{(s, v) \in E : v \in V\}|$$



Degree centrality (Degree)

- Relies on the notion of vertex degree
- Score of a sense given by the number of their outgoing edges

$$score(s) = |\{(s,v) \in E : v \in V\}|$$



Inverse path length sum (PLength)

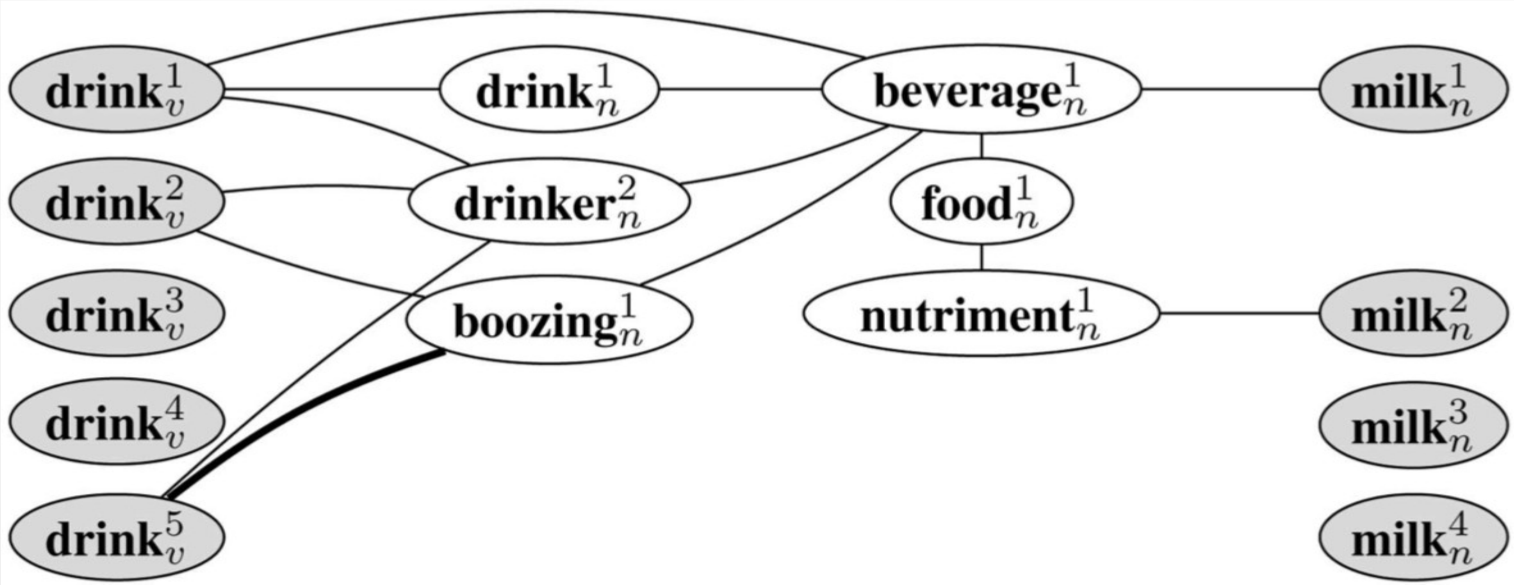
- Relies on fully connecting paths
- Score of a sense given by the length of all paths to other senses in the graph

$$score(s) = \sum_{p \in path(s)} \frac{1}{e^{length(p)-1}}$$

Inverse path length sum (PLength)

- Relies on fully connecting paths
- Score of a sense given by the length of all paths to other senses in the graph

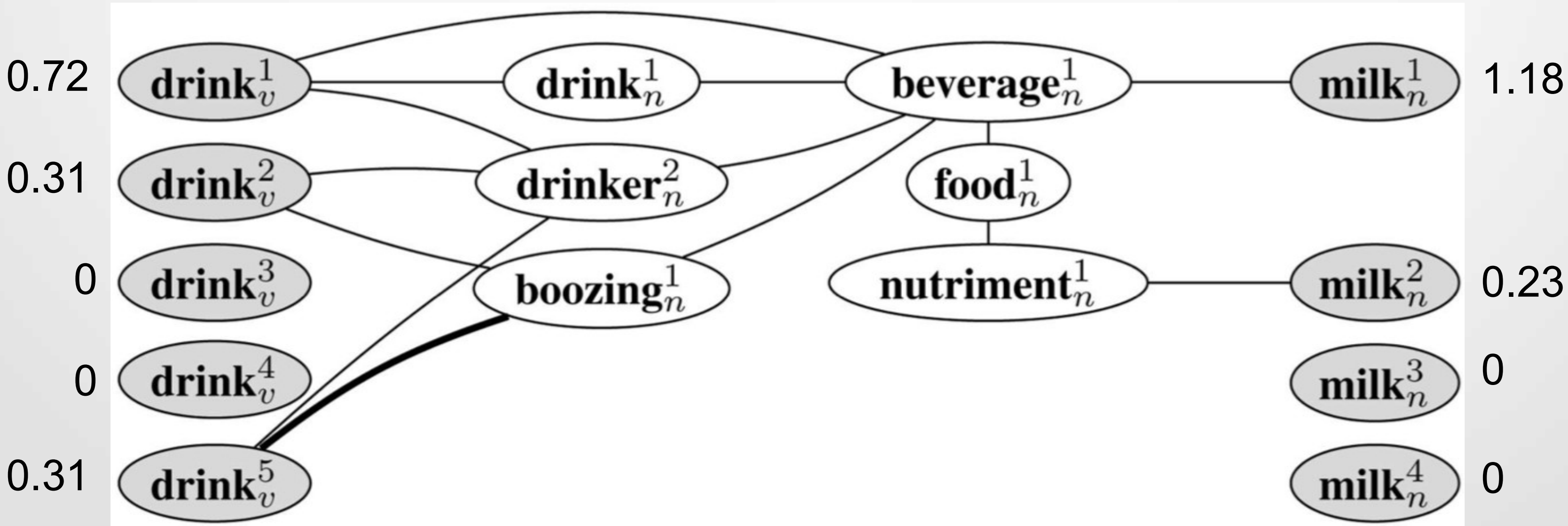
$$score(s) = \sum_{p \in path(s)} \frac{1}{e^{length(p)-1}}$$



Inverse path length sum (PLength)

- Relies on fully connecting paths
- Score of a sense given by the length of all paths to other senses in the graph

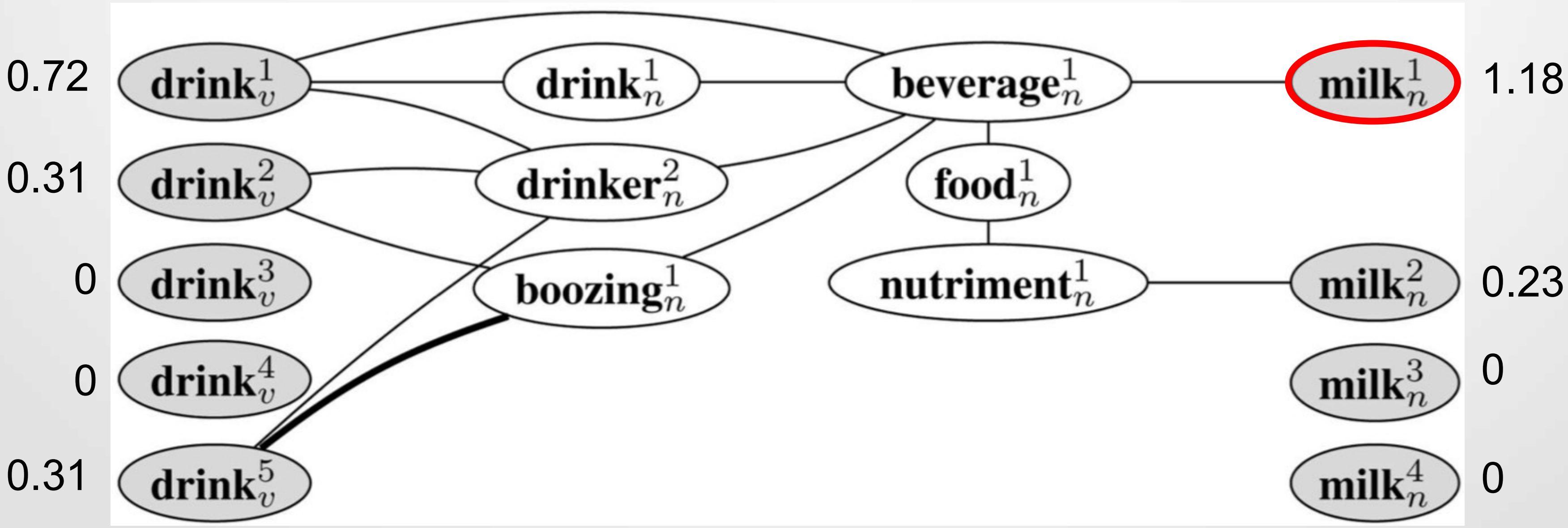
$$score(s) = \sum_{p \in path(s)} \frac{1}{e^{length(p)-1}}$$



Inverse path length sum (PLength)

- Relies on fully connecting paths
- Score of a sense given by the length of all paths to other senses in the graph

$$score(s) = \sum_{p \in path(s)} \frac{1}{e^{length(p)-1}}$$



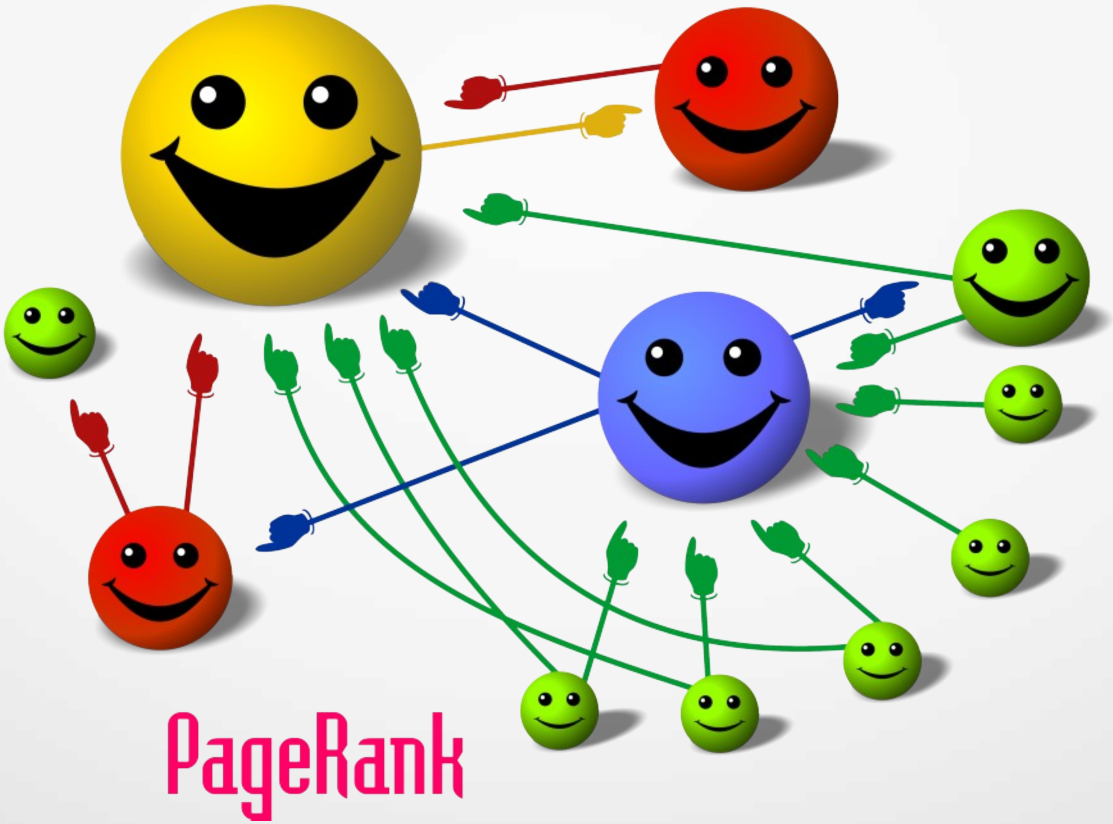
Path probability sum (SProbability)

- Alternative measure for scoring paths
- Sensitive to the weights of each single edge
- Assumes that edges are independent
- Doesn't work with WordNet (no weigths in WN)

$$score(s) = \sum_{p \in paths(s)} \prod_{(u,v) \in p} w(u,v)$$

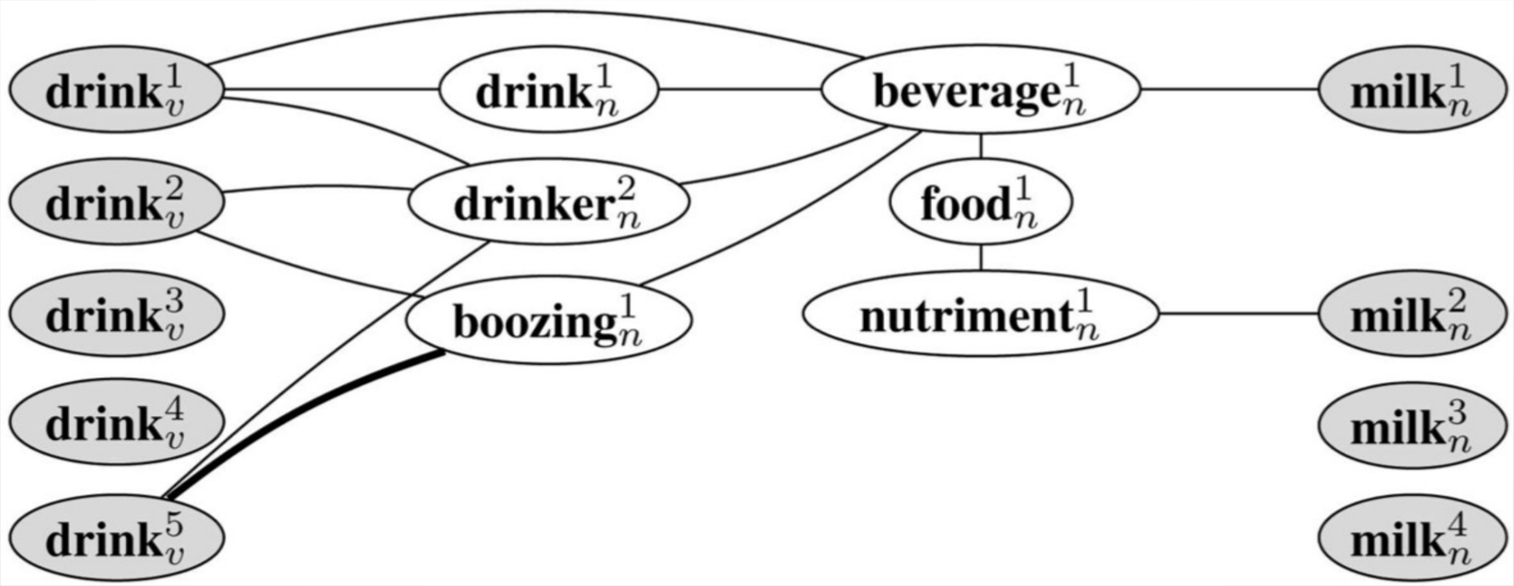
PageRank

- Assign one sense to each node
- Iterate the PageRank algorithm



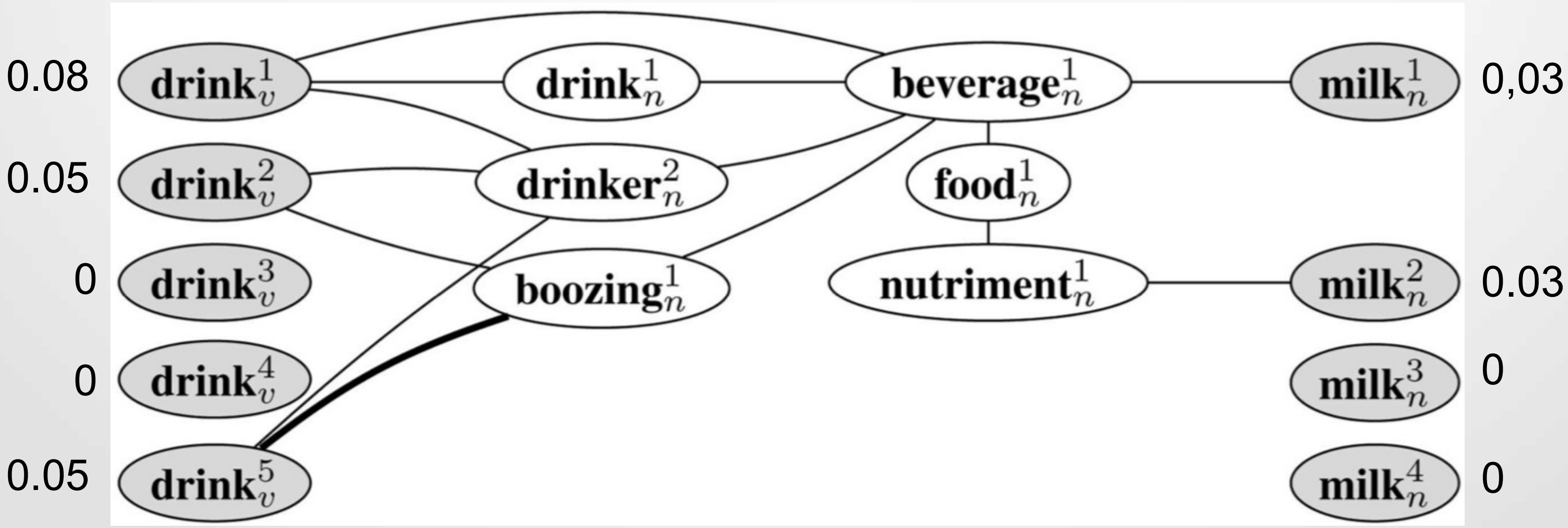
PageRank

- Assign one sense to each node
- Iterate the PageRank algorithm



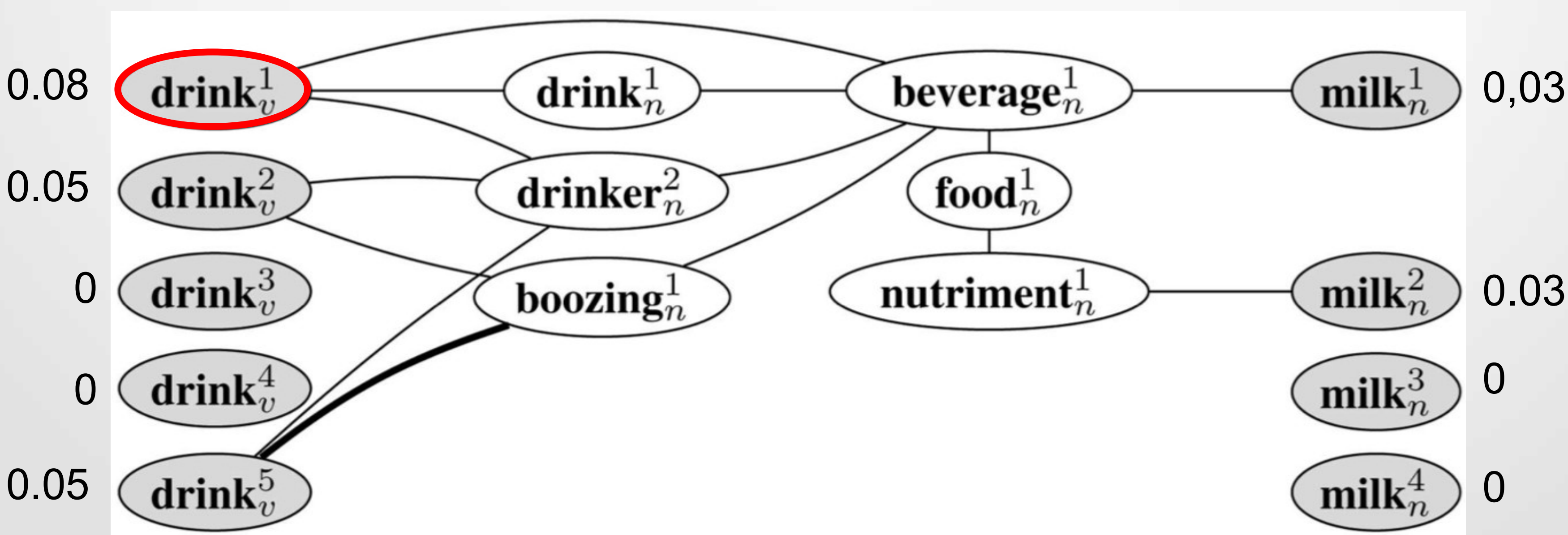
PageRank

- Assign one sense to each node
- Iterate the PageRank algorithm



PageRank

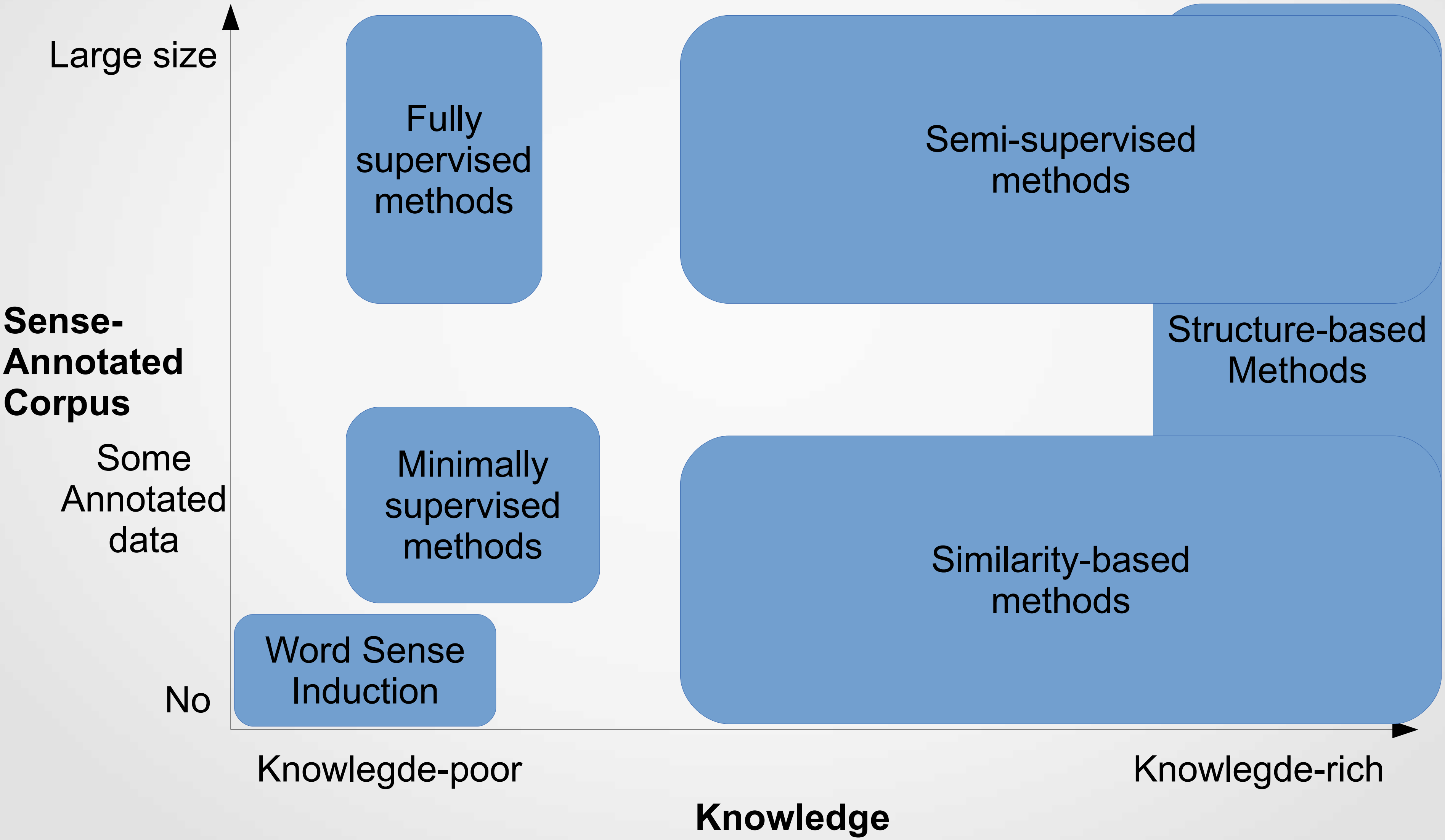
- Assign one sense to each node
- Iterate the PageRank algorithm



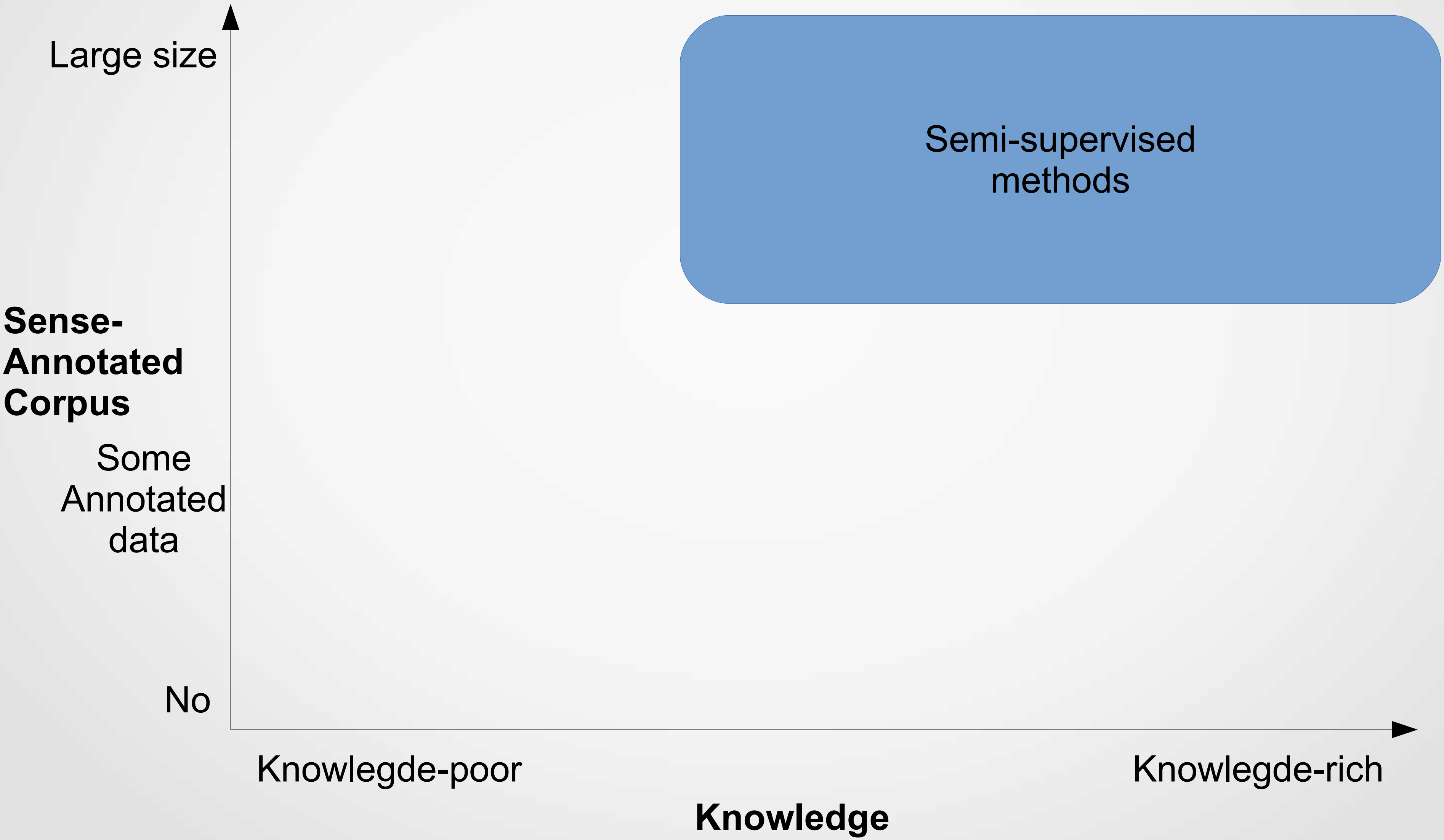
Performance on Semeval 2007

Resource	Algorithm	Nouns only			All words		
		P	R	F_1	P	R	F_1
WordNet	Degree	81.1	67.3	73.6	**79.6**	61.0	69.1
	PLength	81.7	67.9	74.2	78.9	60.6	68.5
	SProbability	79.1	65.7	71.8	77.7	59.6	67.4
	PageRank	80.5	66.5	72.9	79.1	56.2	65.7
BabelNet	Degree	**83.3**	**81.7**	**82.5**	79.4	74.8	77.1
	PLength	82.8	81.1	82.0	77.8	73.3	75.5
	SProbability	82.0	80.3	81.1	77.6	73.2	75.3
	PageRank	81.6	79.9	80.7	78.5	67.6	72.6
	MFS BL	77.4	77.4	77.4	78.9	**78.9**	**78.9**
	Random BL	63.5	63.5	63.5	62.7	62.7	62.7

WSD Approaches



WSD Approaches



Back-off strategies

- Many systems don't tag all words
  - Several solutions with the same evaluation
  - Combinatorial explosion
- How to choose?
  - Randomly
  - Main sense: first sense in WordNet (From SemCor)
  - Other algorithms: often/always supervised

Back-off strategies

- Many systems doesn't tag all words
  - Several solutions with the same evaluation
  - Combinatorial explosion
- How to choose?
  - Randomly
  - Main sense: first sense in WordNet (From SemCor)
  - Other algorithm : often/always supervised

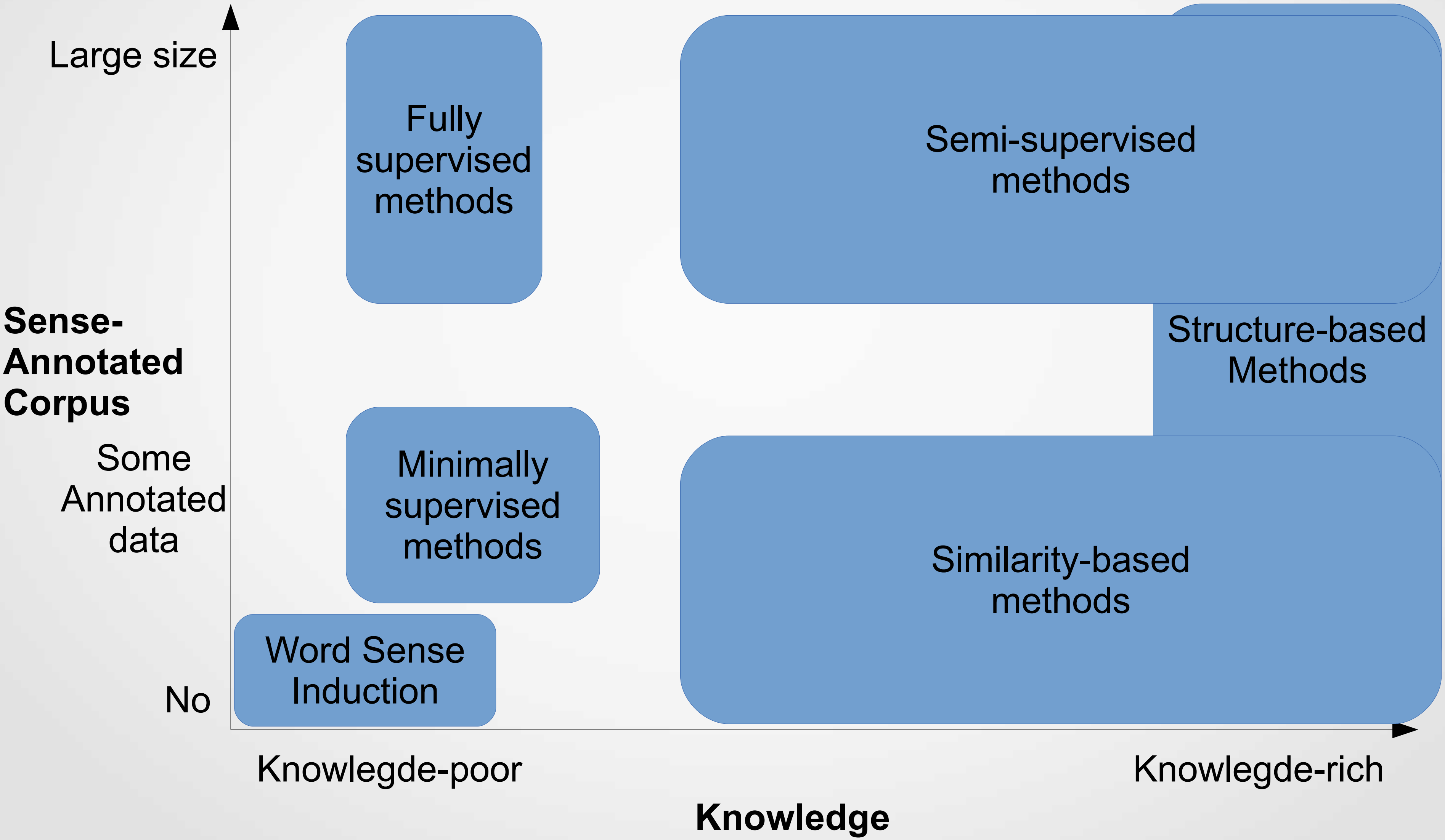
Navigli Approach with Backoff

- Main sense

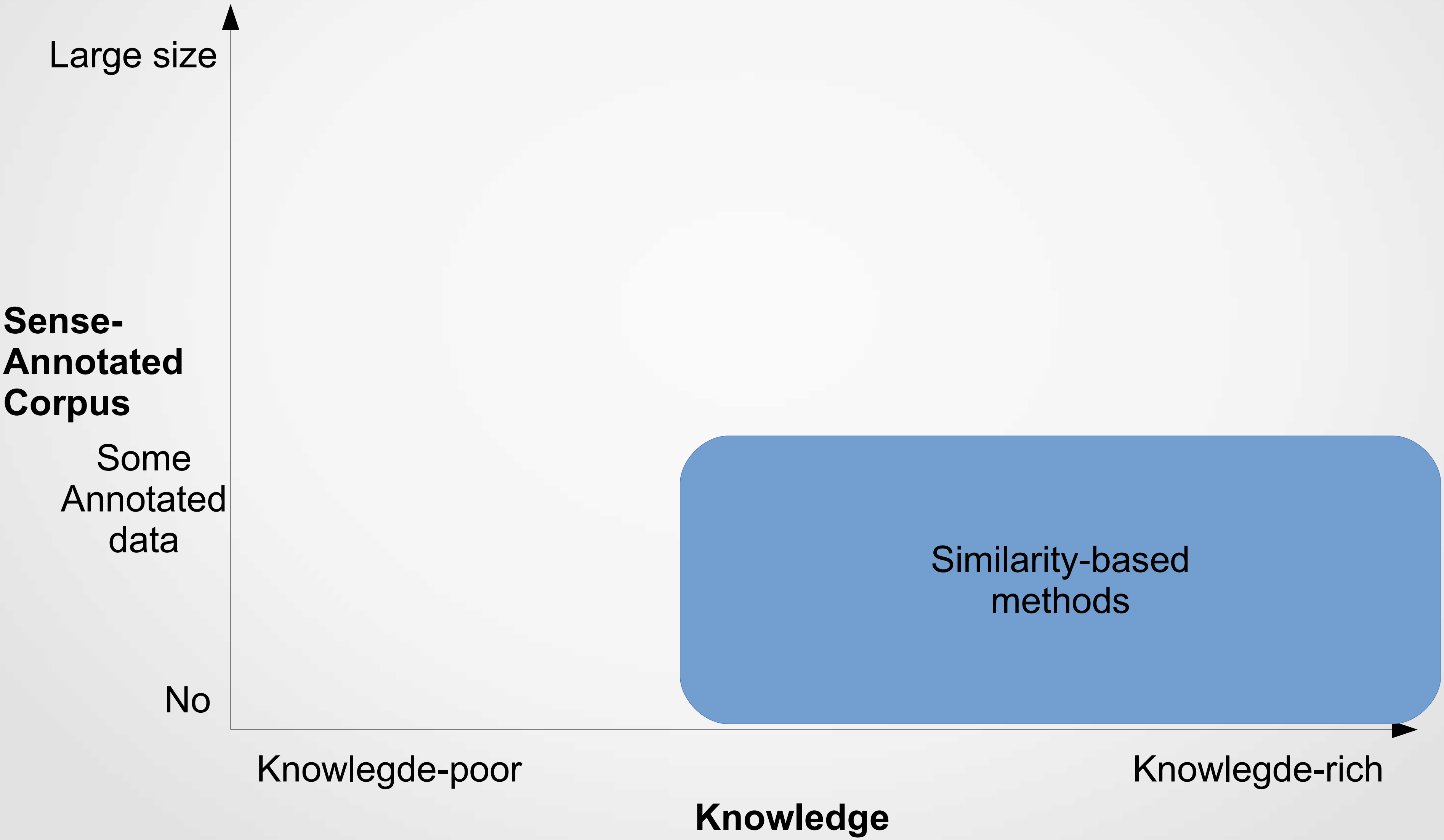
Resource	Algorithm	Nouns only P/R/F_1	All words P/R/F_1
WordNet	Degree	80.1	79.7
	PLength	80.3	79.8
	SProbability	79.5	79.3
	PageRank	79.7	79.4
BabelNet	Degree	**84.7**	**82.3**
	PLength	**85.4**	**82.7**
	SProbability	**84.6**	**82.1**
	PageRank	82.1	80.1

Resource	Algorithm	Nouns only			All words		
		P	R	F_1	P	R	F_1
WordNet	Degree	81.1	67.3	73.6	**79.6**	61.0	69.1
	PLength	81.7	67.9	74.2	78.9	60.6	68.5
	SProbability	79.1	65.7	71.8	77.7	59.6	67.4
	PageRank	80.5	66.5	72.9	79.1	56.2	65.7
BabelNet	Degree	**83.3**	**81.7**	**82.5**	79.4	74.8	77.1
	PLength	82.8	81.1	82.0	77.8	73.3	75.5
	SProbability	82.0	80.3	81.1	77.6	73.2	75.3
	PageRank	81.6	79.9	80.7	78.5	67.6	72.6
	MFS BL	77.4	77.4	77.4	78.9	**78.9**	**78.9**
	Random BL	63.5	63.5	63.5	62.7	62.7	62.7

WSD Approaches



WSD Approaches



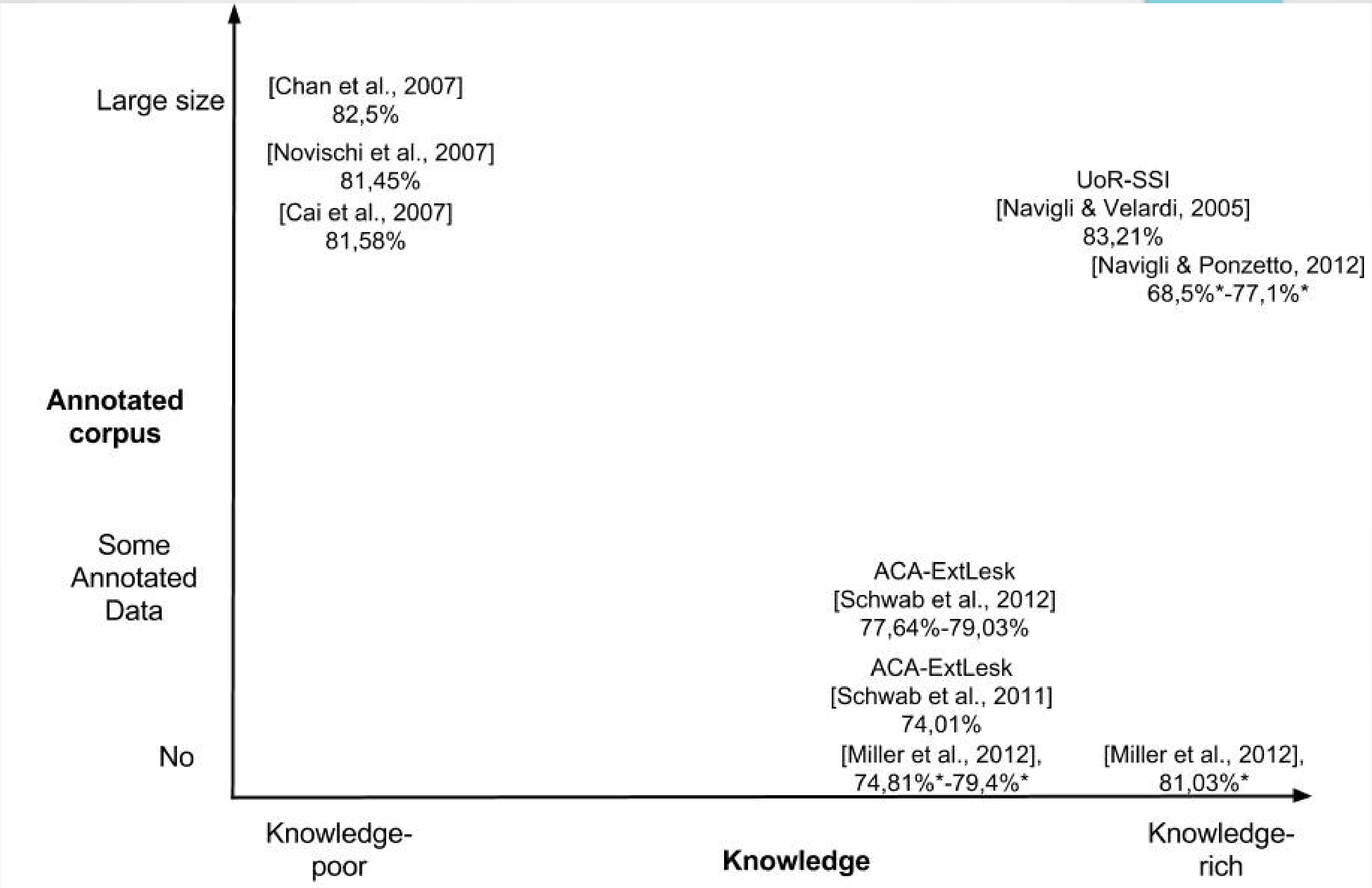
Word Sense Disambiguation Process

- Composed of 3 steps
  - Build/select raw lexical material(s)
  - Build an elaborate resource
  - Use that resource to lexically disambiguate a text

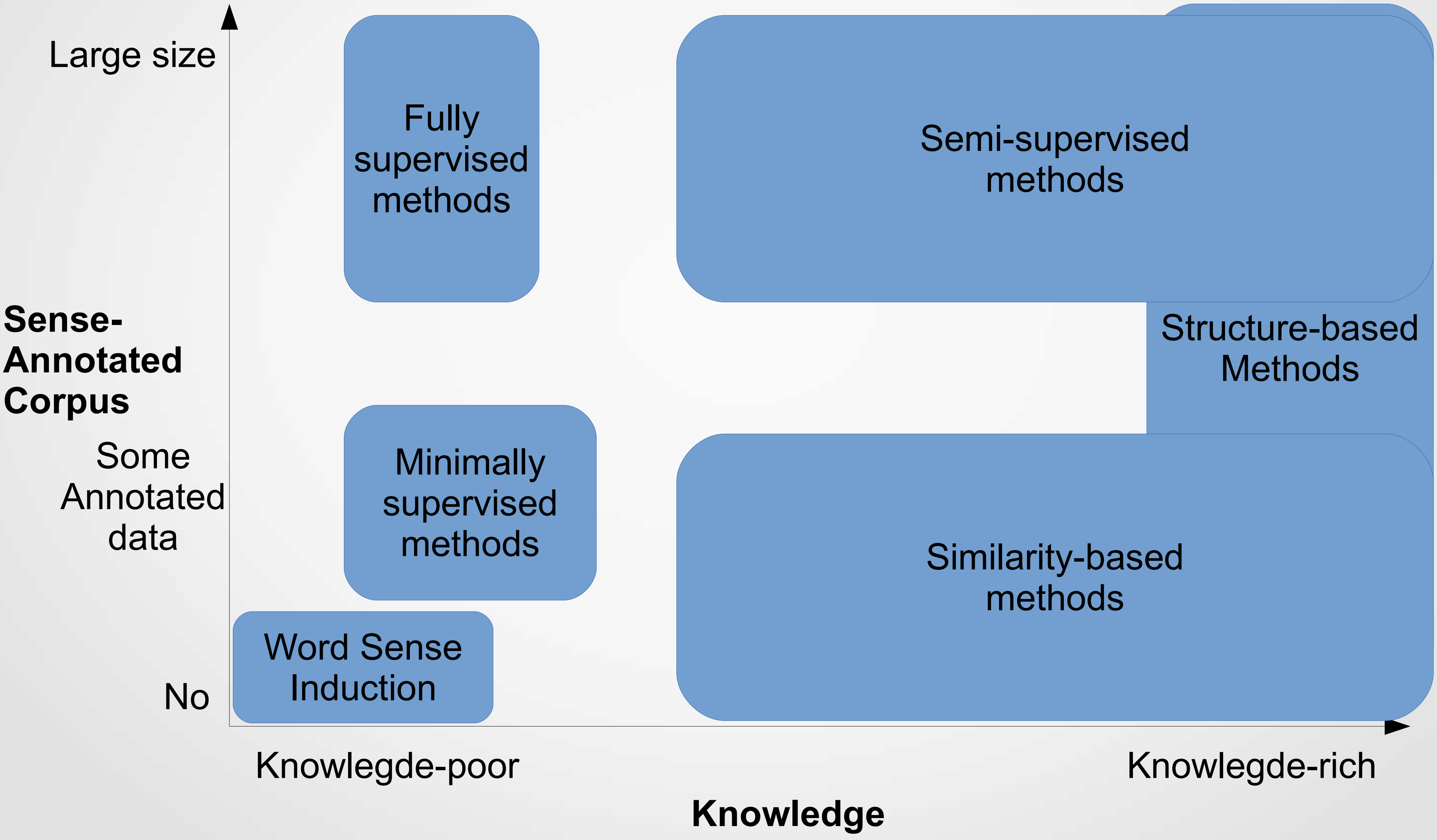
Word Sense Disambiguation Process

- Composed of 3 steps
  - Build/select of raw lexical material(s)
    - Mandatory: MRD or Lexical Base
    - Optional: corpus (sense-annotated or not)
  - Build an elaborate resource
    - Various ways to construct
  - Use that resource to lexically disambiguate a text
    - Local algorithm : semantic relatedness between senses
    - Global algorithm : Various

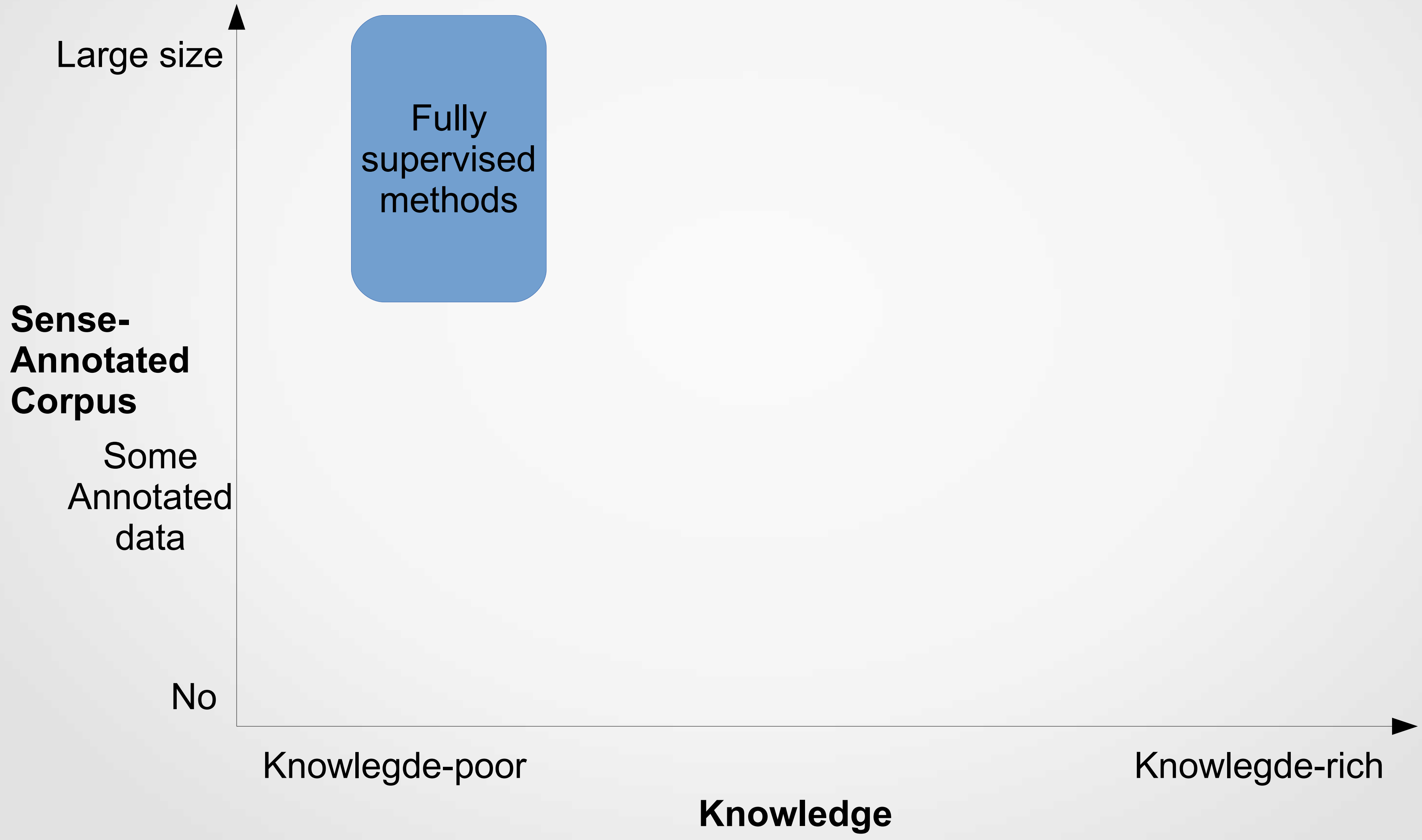
Semeval 2007 map



WSD Approaches



WSD Approaches



Word Sense Disambiguation Process

- Composed of 3 steps
  - Build/select raw lexical material(s)
  - Build an elaborate resource
  - Use that resource to lexically disambiguate a text

Supervised WSD

- Build/select raw lexical material(s)
  - Only using sense annotated corpus/corpora
- Build an elaborate resource
  - Learn one classifier per word
- Use that resource to lexically disambiguate a text
  - Use classifiers to find the best sense for each word in texts

Supervised Word Sense Disambiguation

- Machine Learning techniques
- Learn classical classifiers on sense-tagged corpora
  - Support Vector Machines NUS-PT, (Chan et al., 2007)
  - Naïve Bayes NUS-ML, (Cai et al., 2007)
  - Maximum Entropy / Support Vector Machines LCC-WSD, (Novischi et al., 2007)
- One classifier per word

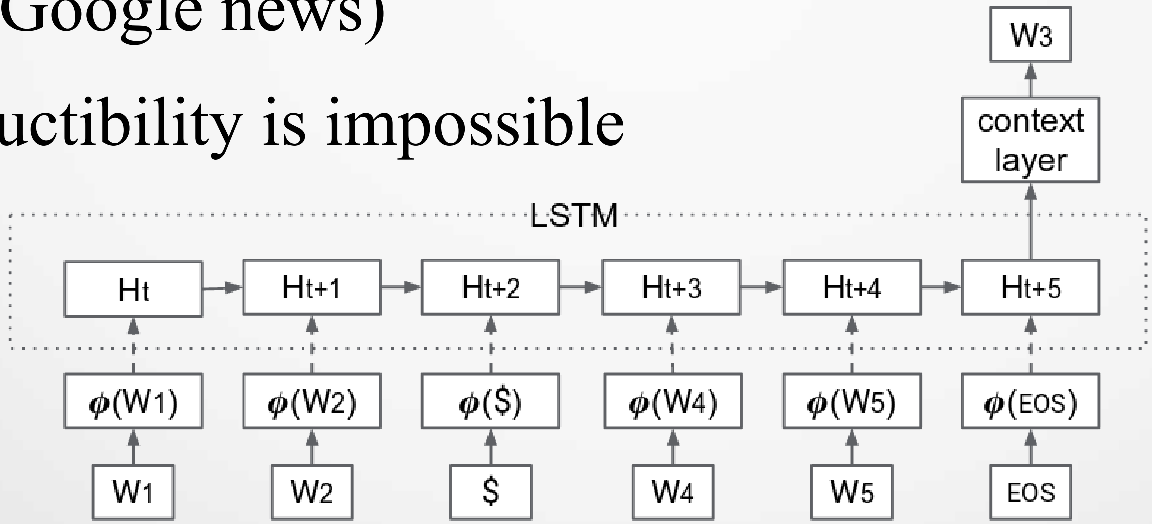
=> **state of the art on WSD 2007 -> 2016**

Deep Neural Networks

- 2016 → …
- [Yuan et al., 2016]
- [Raganato et al., 2017]
- [Vial et al., 2018]
- [Vial et al., 2019]

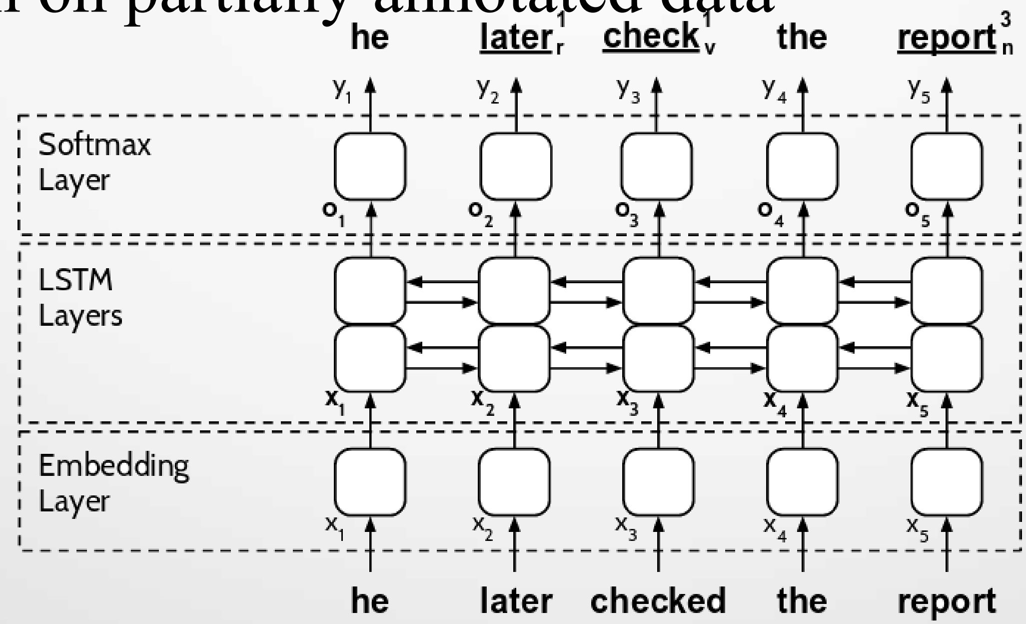
[Yuan et al., 2016]

- LSTM language Model (Long Short-Term Memory)
- Give a prediction for a target word (classification)
- Closest sense is assigned
- Language model learned on a private corpus of 100 billions words (Google news)
- Reproductibility is impossible



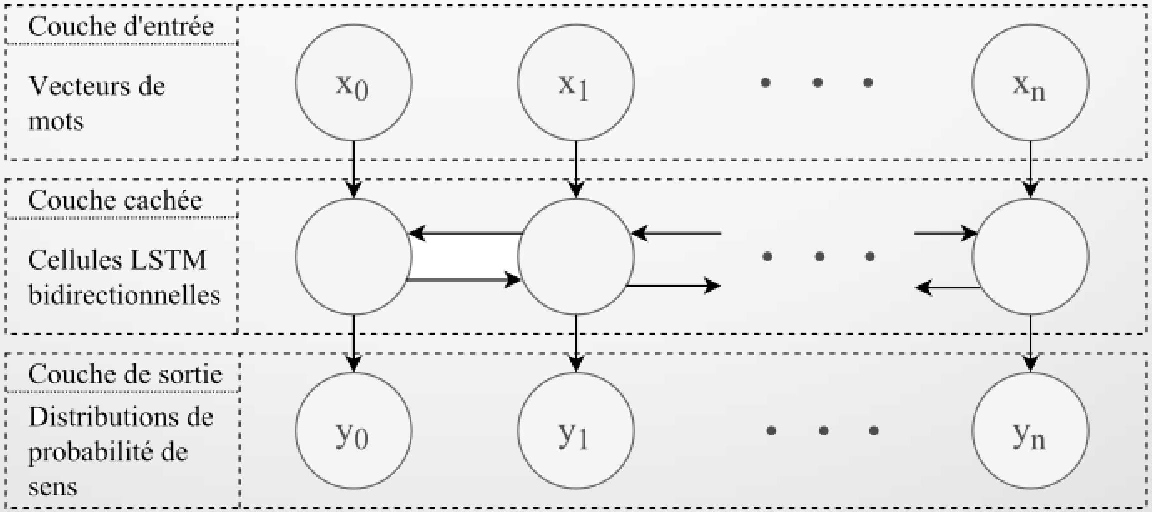
[Raganato et al., 2017]

- Directly predict sense for each word
- Predict word when no sense can be assigned
- Multi-task learning (POS + WSD)
- Reproductibility is possible
- Can't learn on partially annotated data



[Vial et al., 2018]

- Input layer : pre-trained vectors (Glove (Pennington et al., 2014))
- Hidden layer : Bidirectional LSTM (size : 1000)
- Output layer : size number of senses (~ 100 000)
- Dropout : 50%



Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural Word Sense Disambiguation, Global WordNet Conference 2019

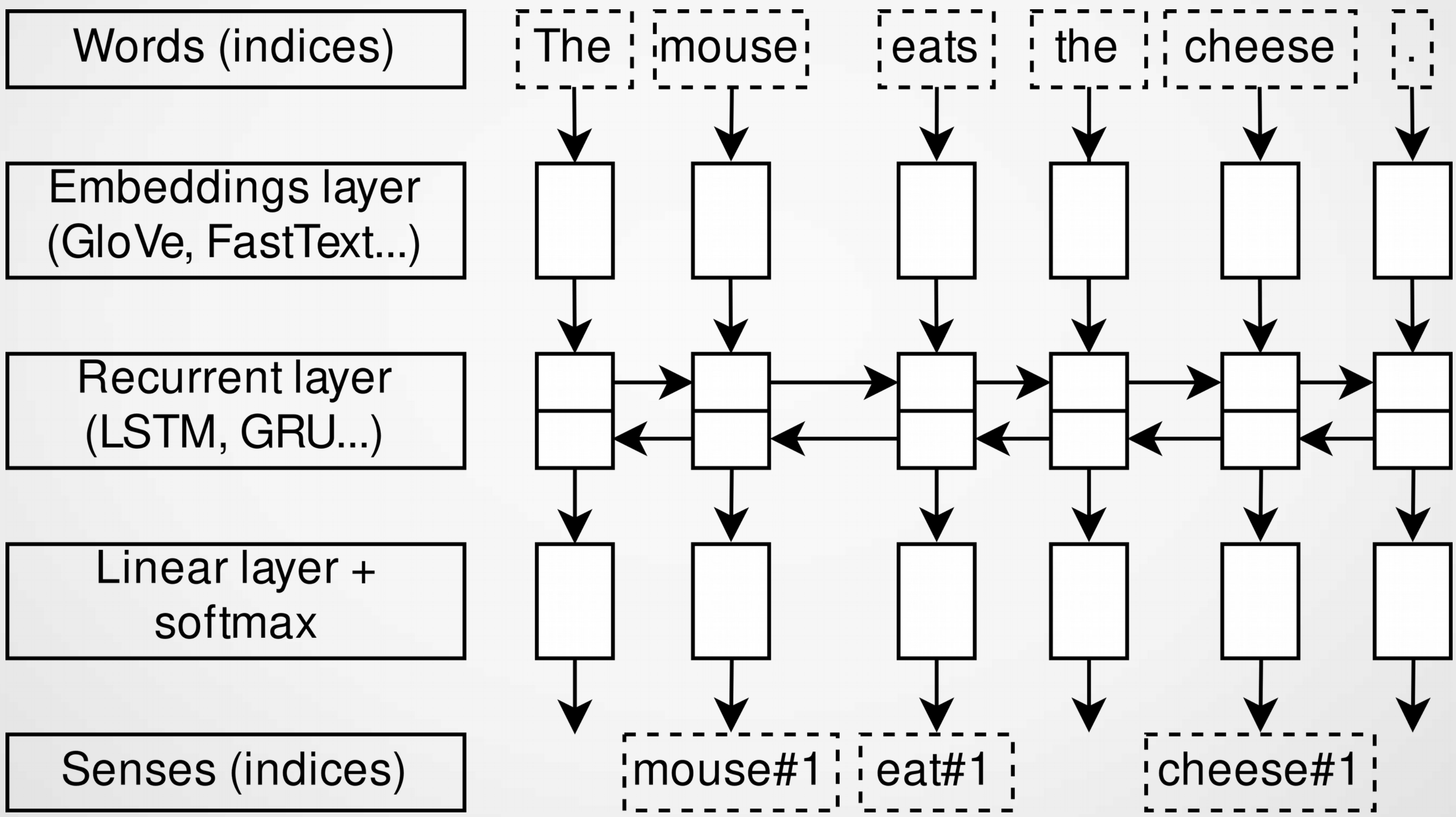
Loïc Vial, Benjamin Lecouteux, Didier Schwab

State of the art in WSD (on the 19th Nov 2019)

-

Best Paper Award TALN 2019 (French version)

State of the art neural approach for supervised Word Sense Disambiguation



Drawbacks of current supervised systems

- Output vocabulary (number of sense tags) is large

  WordNet 3.0 = 206 941 senses

  → Output layer of a typical neural model

  = ~200M parameters

- Sense annotated corpora = costly resource

  SemCor: largest manually annotated corpus

  → Only 16% of all WordNet senses are represented

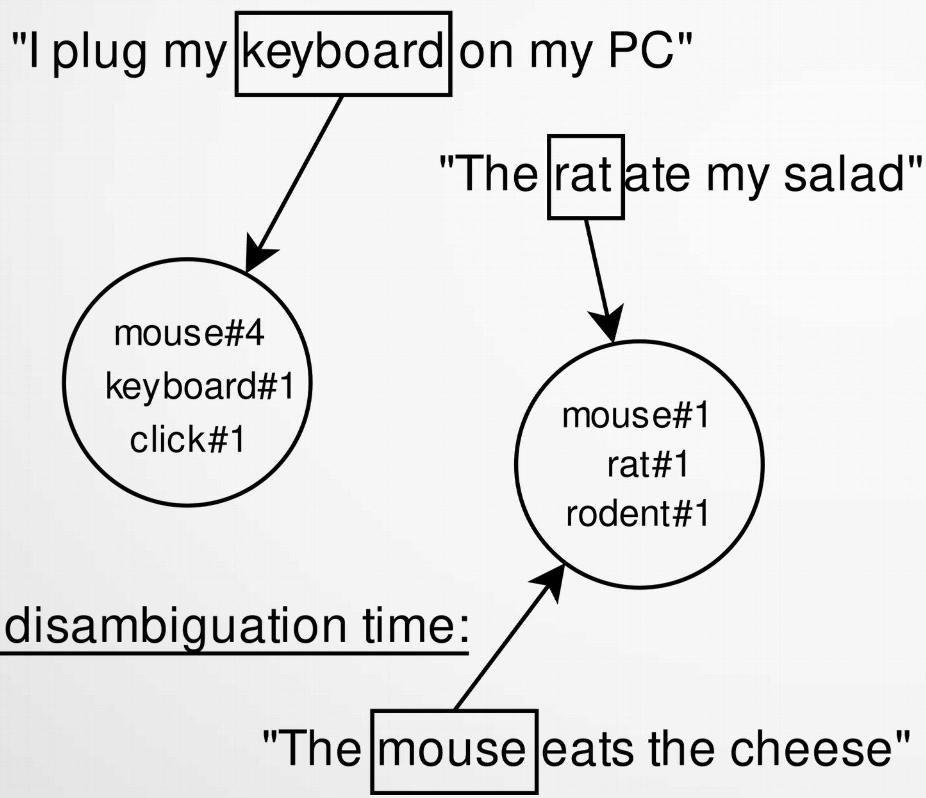
Sense Vocabulary Compression

- Principle:
  - Form groups of similar senses, for instance:
    - group n°1 : {mouse#1, rat#1, rodent#1…}
    - group n°2 : {mouse#4, keyboard#1, click#4...}
- Learn to predict group tags instead of sense tags during training
- Find back the “true” sense at disambiguation time, from the lemma of the target word

Sense Vocabulary Compression

Example

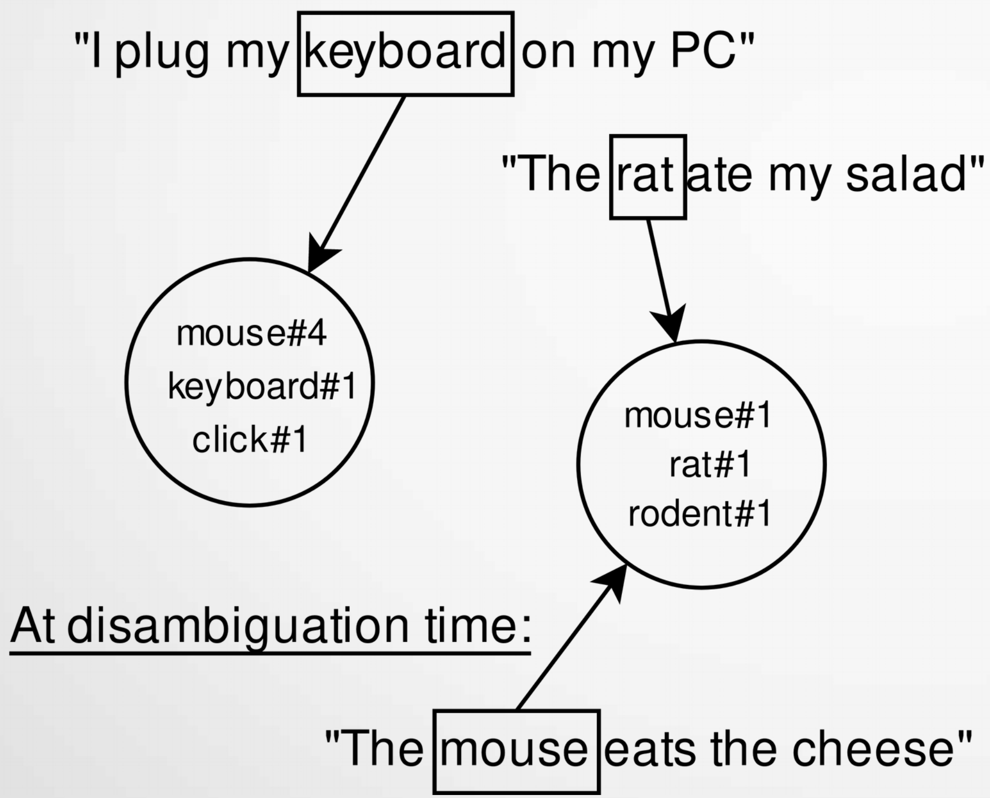
At training time:



Sense Vocabulary Compression

Example

At training time:



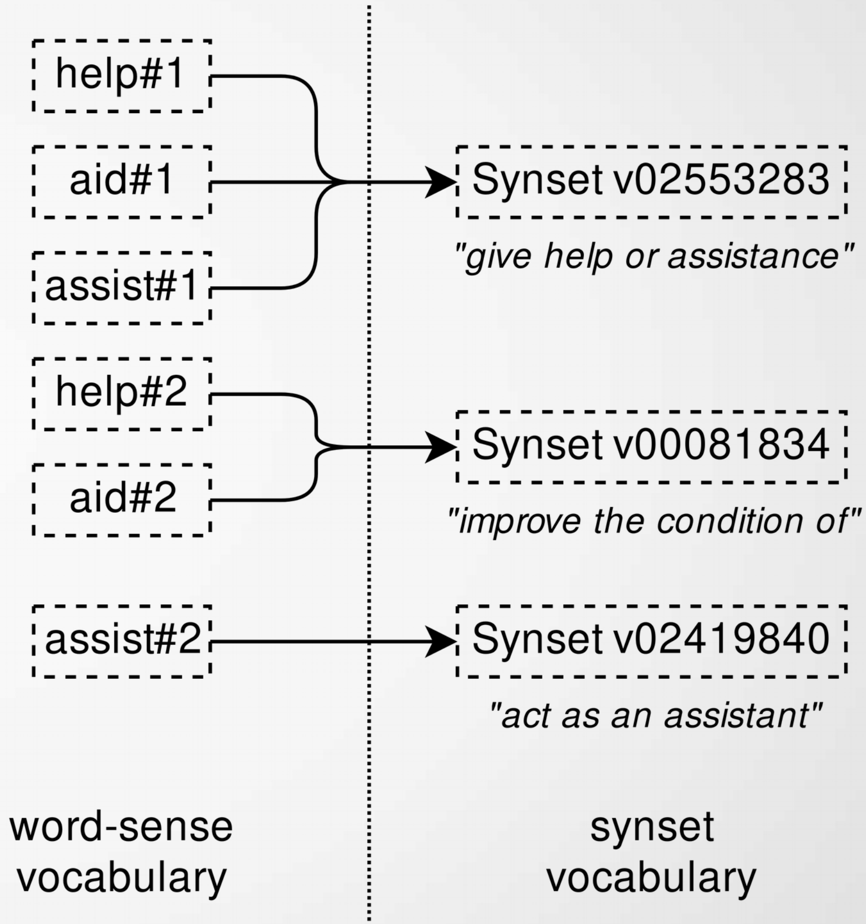
Advantages

- Smaller number of senses
  - Smaller size of neural models
  - Shorter training time
- Increased coverage
- Better generalization ?

Baseline method: from senses to synsets

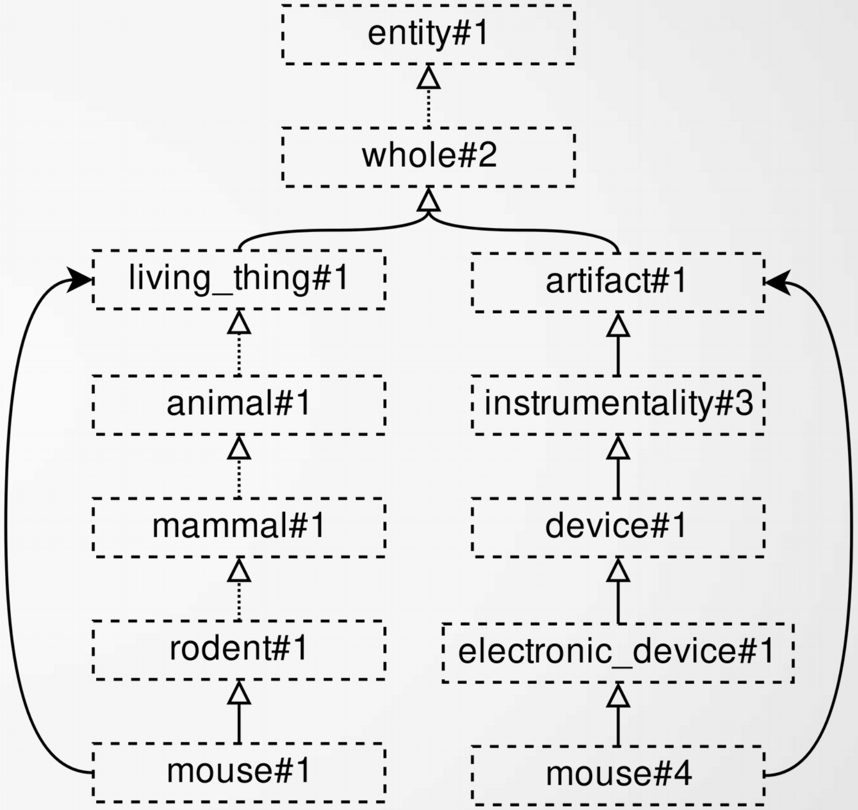
- In WordNet, senses are grouped into sets of synonyms called “synsets”
- State of the art systems rarely indicate whether they predict synset tags or sense tags
- It has a significant effect though:

Method	Vocabulary size	Compression rate	SemCor coverage
Senses	206 941	0 %	16 %
Synsets	117 659	43 %	22 %



Proposed method n°1: compression through the hypernymy and hyponymy relationships

- Hypernymy and hyponymy relationships connect together all nouns (and many verbs !) in WordNet
- Idea: associate the most specific concepts to more general concepts
- Constraint: always being able to discriminate the different senses of every word
- Example:
  - mouse#1 → animal → living thing
  - mouse#4 → device → artifact



Proposed method n°1: compression through the hypernymy and hyponymy relationships

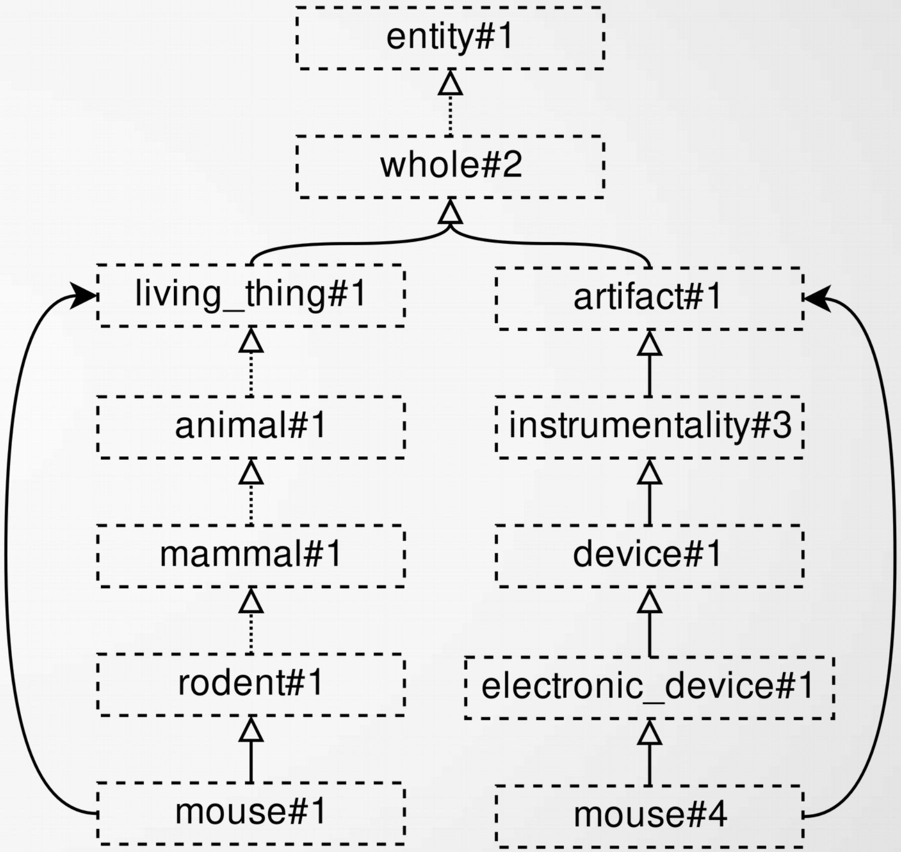
- **Method:**

  1) For every lemma of WordNet, for every pair of its senses, find their common ancestor, and mark the children of its ancestor as "necessary"

  2) Map every sense of WordNet to its first ancestor in the hypernymy hierarchy that has been previously marked as "necessary"

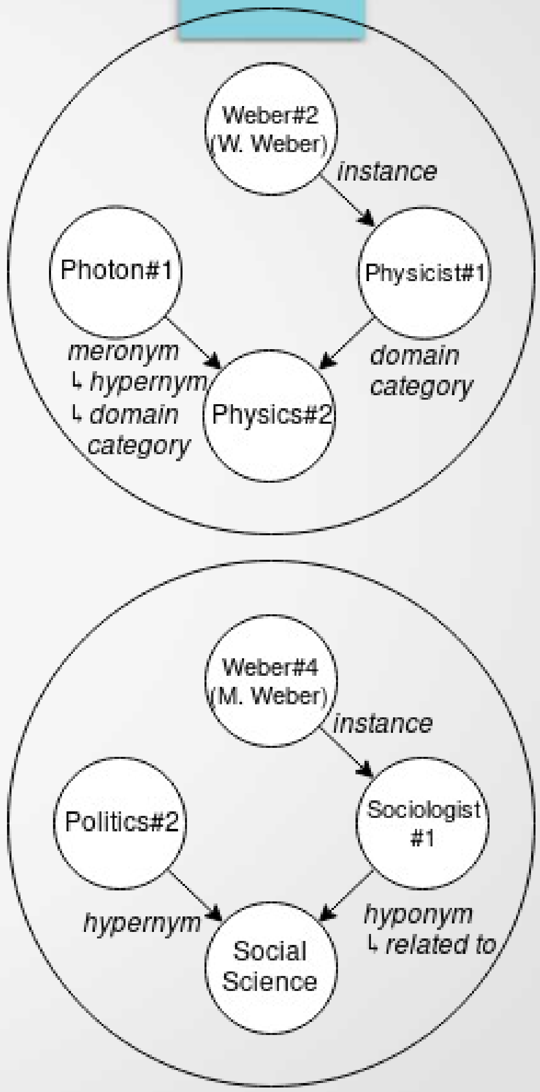
- **Results**:

Method	Vocabulary size	Compression rate	SemCor coverage
Sens	206 941	0 %	16 %
Synsets	117 659	43 %	22 %
Hypernyms	39 147	81 %	32 %



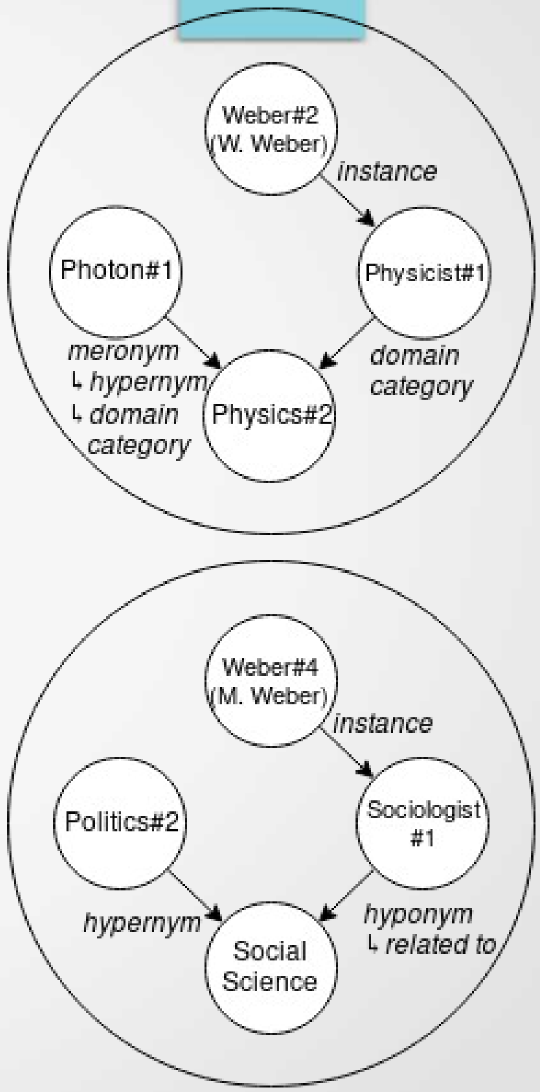
Proposed method n°2: compression through all semantic relationships

- Numerous other semantic relationships are present in WordNet (meronymy, antonymy, domain…)
- Can we go even further by using all the relationships offered ?
- Idea: build iteratively groups of senses linked by any semantic relationship
- Example:
  - {Weber#4, sociologist, social science...}
  - {Weber#2, physicist, physics, photon...}



Proposed method n°2: compression through all semantic relationships

- **Method**:
  - Initialization: every group contains a different sense
  - Iteratively:
    - Select the smallest group g_1
    - Select the smallest group g_2 linked to g_1 by any semantic link
    - Merge g_1 and g_2 together iff the operation still allows to discriminate every sense of every word of WordNet



- **Results**:

Method	Vocabulary size	Compression rate	SemCor coverage
Senses	206 941	0 %	16 %
Synsets	117 659	43 %	22 %
Hypernyms	39 147	81 %	32 %
All relations	11 885	94 %	39 %

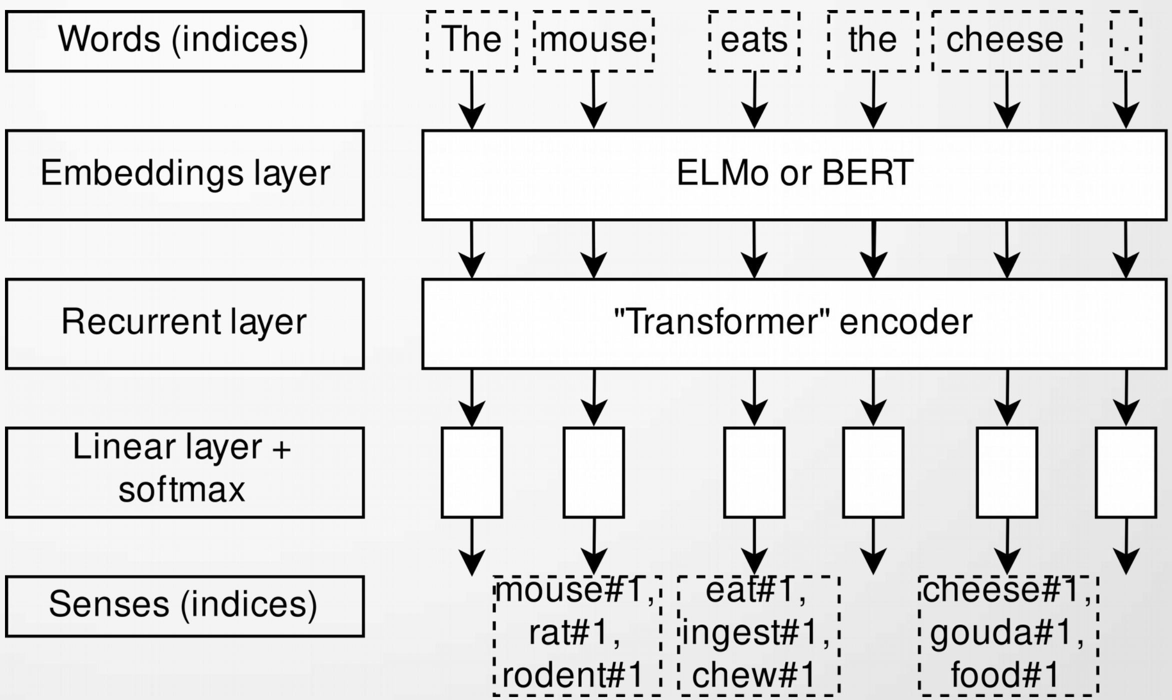
Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural WSD: by L. Vial, B. Lecouteux and D. Schwab

Evaluation of our compression methods

- Implementation of a state of the art neural system:
  - Input embeddings: GloVe → ELMo / BERT
  - Recurrent layer: LSTM → Transformer
  - Output: depends on the vocabulary compression method
- Training corpus:

  1) SemCor

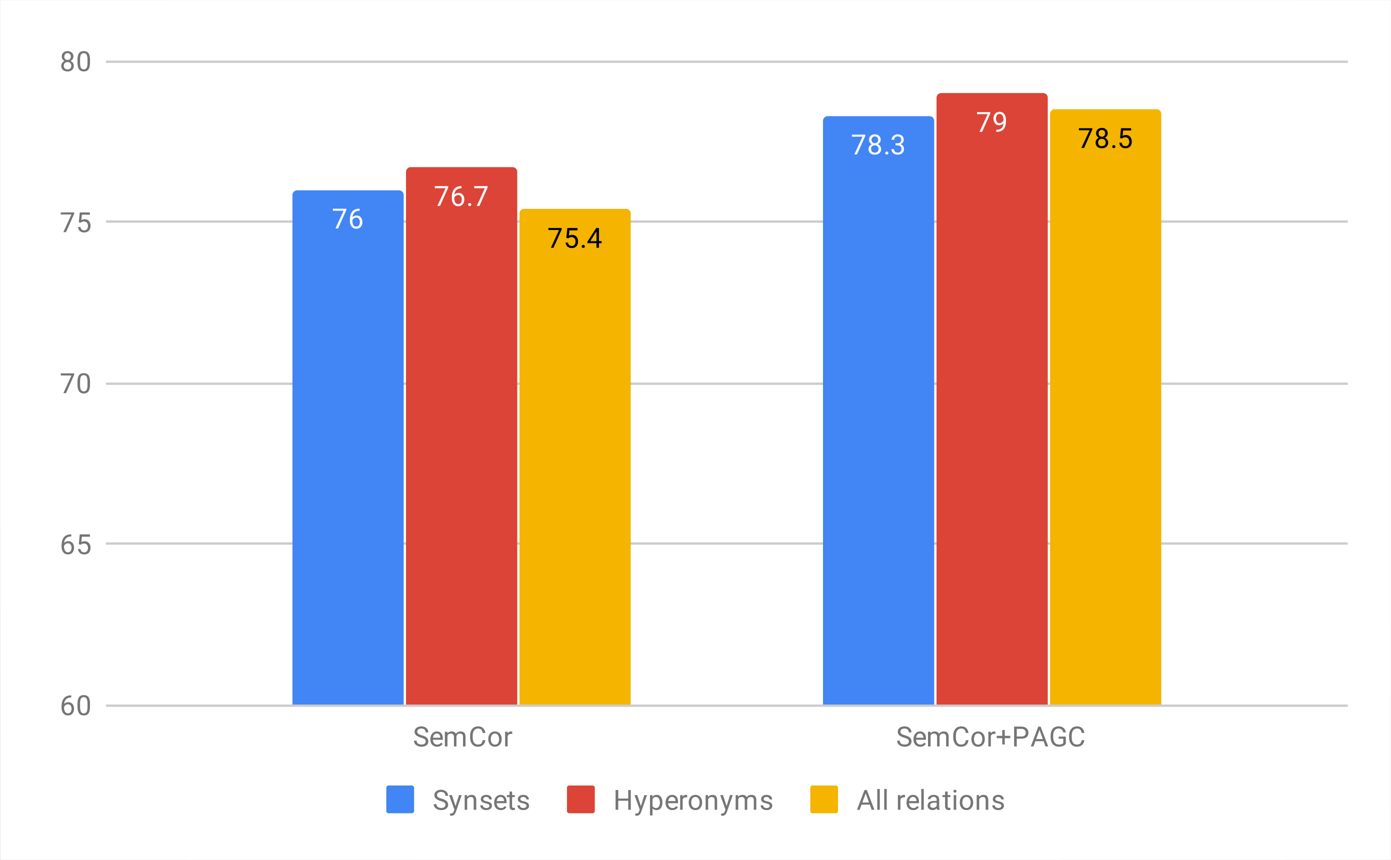
  2) SemCor+PAGC



Results of our compression methods

- **Evaluation corpus:**
  - Concatenation of fine-grained all-words WSD tasks from SensEval 2/3 and SemEval 2007/2013/2015

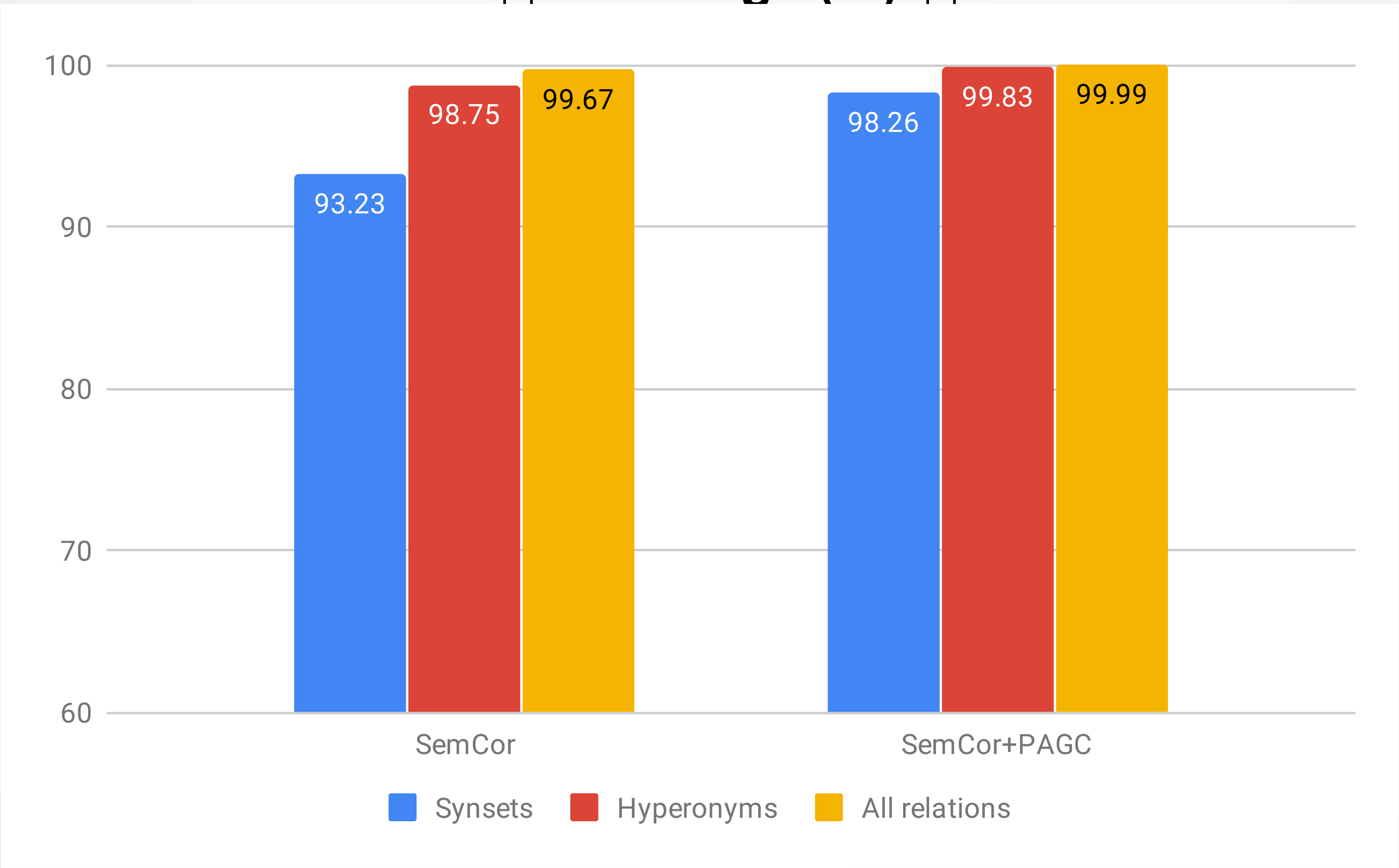
↑↑ F1 Score (%) ↑↑



Results of our compression methods

- **Evaluation corpus:**
  - Concatenation of fine-grained all-words WSD tasks from SensEval 2/3 and SemEval 2007/2013/2015

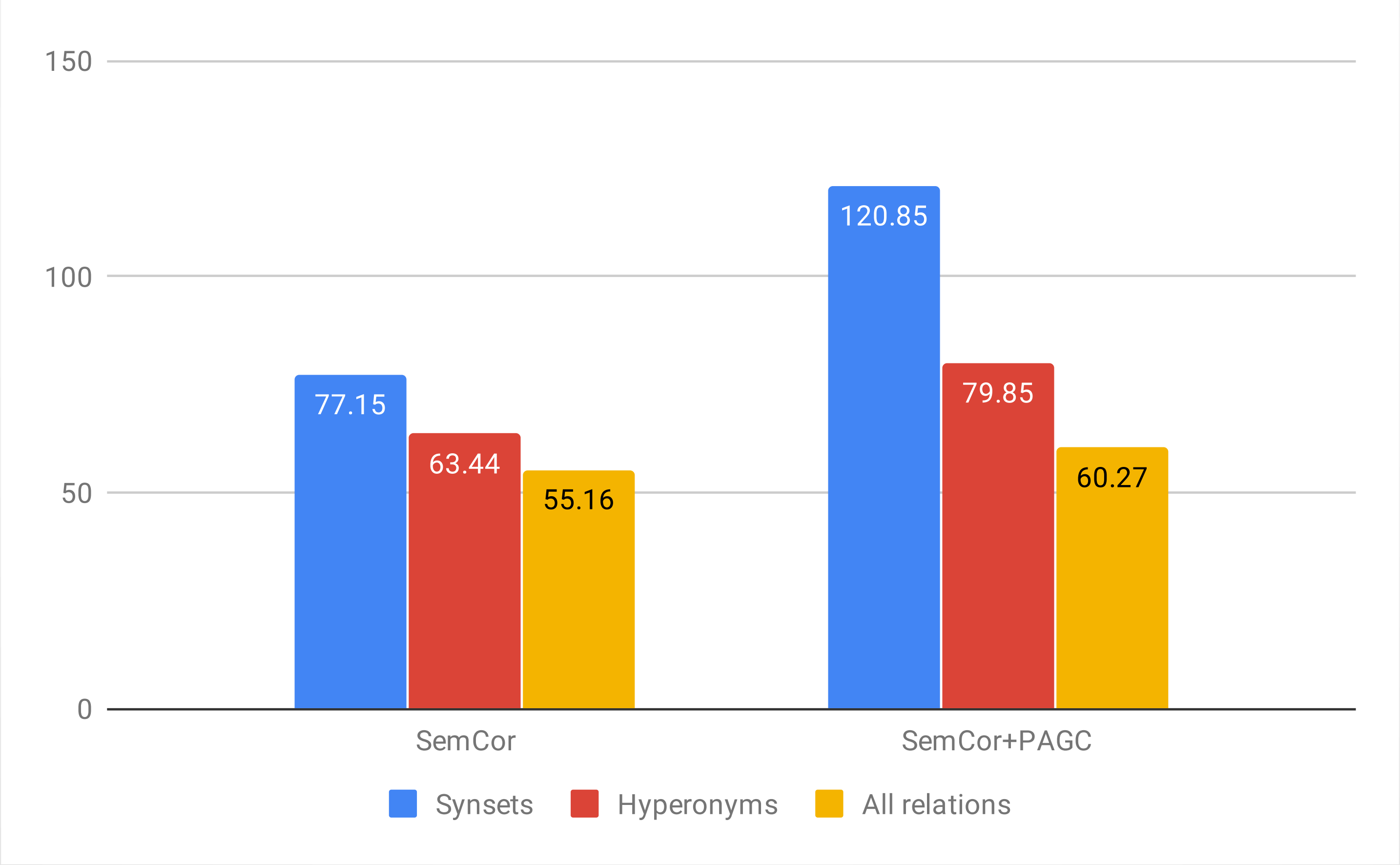
↑↑ Coverage (%) ↑↑



Results of our compression methods

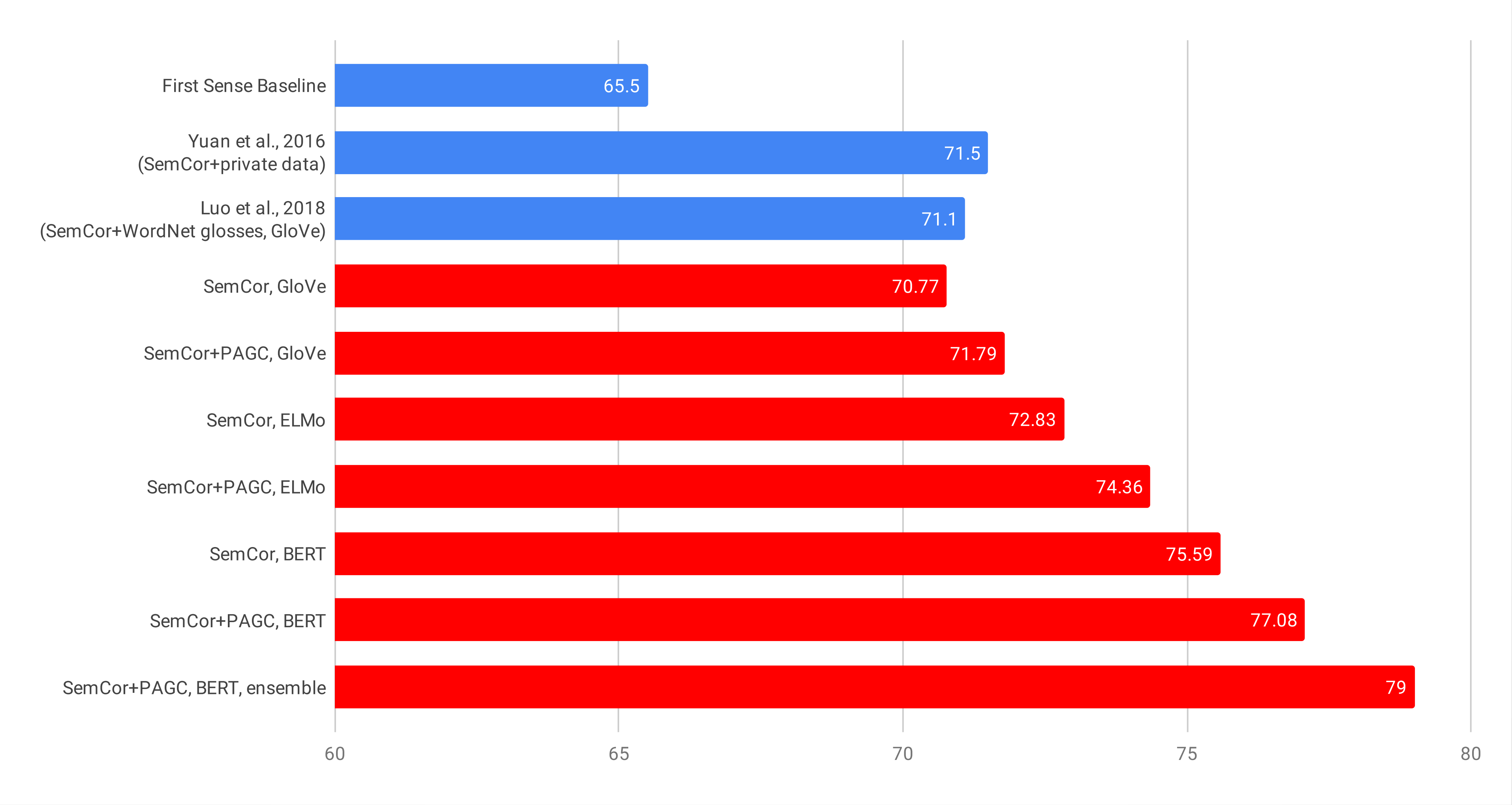
- **Evaluation corpus:**
  - Concatenation of fine-grained all-words WSD tasks from SensEval 2/3 and SemEval 2007/2013/2015

↓↓ Size of neural models (millions of parameters) ↓↓



Hyperparameter study and comparison with other works

↑↑ F1 Score (%) with compression through hypernyms ↑↑



Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural WSD: by L. Vial, B. Lecouteux and D. Schwab

Conclusion

- Sense Vocabulary Compression :

  → Easy to implement method

  → Improves the coverage and generalization ability of neural WSD systems

  → Reduces the number of parameters of neural models

- New “contextualized” word embeddings (ELMo, BERT) :

  → Greatly improve the performance of neural WSD systems

  → Improve the state of the art by almost 10 points

- Our code and our pre-trained models are available: https://github.com/getalp/disambiguate

Bibliography

[Gale et al., 1991] One sense per discourse, Gale, William A. and Church, Kenneth W. and Yarowsky, David, HLT '91 Proceedings of the workshop on Speech and Natural Language, Pages 233-237, Harriman, New York, 1991

[Lesk, 1986] Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In SIGDOC '86: Proceedings of the 5th annual international conference on Systems documentation, pages 24-26, New York, NY, USA. ACM

[Navigli & Lapata, 2010]R. Navigli, M. Lapata, An experimental study on graph connectivity for unsupervised Word Sense Disambiguation, IEEE Transactions on Pattern, 2010

Analysis and Machine Intelligence 32 (2010) 678–692.

[Navigli & Ponzetto, 2012] R. Navigli and S. Ponzetto. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. Artificial Intelligence, 193, Elsevier, 2012, pp. 217-250

[Rivest, 1987] Learning Decision Lists, Ronald Rivest, Machine Learning, 2, pp 229-246, 1987

[Sánchez-de-Madariaga & Fernández-del-Castillo, 2009] The bootstrapping of the Yarowsky algorithm in real corpora Ricardo Sánchez-de-Madariaga & José R. Fernández-del-Castillo, Journal Information Processing and Management: an International Journal, Volume 45 Issue 1, Pages 55-69, January, 2009

[Sarkar, 2009] Bootstrapping a Classifier Using the Yarowsky Algorithm, Anoop Sarkar

[Schütze, 1992] Dimensions of Meaning,H. Schütze Proceedings of the 1992 ACM/IEEE Conference on Supercomputing, Supercomputing '92, Minneapolis, Minnesota, USA, 1992

[Yarowsky,1993] One Sense Per Collocation, Yarowsky, David, Proceedings of the Workshop on Human Language Technology, HLT '93, pages 266-271, 1993