



NUI Galway
OÉ Gaillimh

Introduction to NLP

Semantic Analysis

Dr. Paul Buitelaar

Data Science Institute, NUI Galway



Learning Outcomes of This Lecture

Overview of main tasks in semantic analysis of language

Understand how to develop a Word Sense Disambiguation approach

Insight into Semantic Role Labeling

Insight into Coreference Resolution



Overview

Semantic Analysis

Word Sense Disambiguation

Semantic Role Labelling

Coreference Resolution

Overview

Semantic Analysis

Word Sense Disambiguation

Semantic Role Labelling

Coreference Resolution



Semantic Analysis

Semantic analysis of language data is concerned with how words refer to concepts about objects in the world

Challenge is **ambiguity**: which word(s) refer to which concept(s)?

Levels of Semantic Analysis

Lexical Semantics

meaning of words

Compositional Semantics

combined meaning of words in a phrase or sentence

Discourse Semantics

combined meaning of words across several sentences

Semantic Analysis Tasks

Lexical Semantic Analysis: **word sense disambiguation**

identify correct word meaning or 'word sense'

Compositional Semantic Analysis: **semantic role labeling**

identify correct 'semantic role' for each word/phrase

Discourse Semantic Analysis: **coreference resolution**

identify which words/phrases refer to the same 'entity' across sentences

Overview

Semantic Analysis

Word Sense Disambiguation

Semantic Role Labelling

Coreference Resolution



Lexical Semantic Ambiguity

Words can have different meanings due to **homonymy**



table in the interpretation of ‘furniture’

table in the interpretation of ‘accounting’

3 YEARS ACCOUNTING

| | | MAR18 - MAY18 | | JUN18 - AUG18 | | SEP18 - NOV18 | | DEC18 - FEB19 | | MAR19 - MAY19 | | JUN19 - AUG19 | | SEP19 - N | |
|-----------------------------|---|---------------|-------------|---------------|-------------|---------------|-------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | | Q1 | | Q2 | | Q3 | | Q4 | | Q5 | | Q6 | | Q7 | |
| COST ITEM | DESCRIPTION | TOTAL BUDGET | HISTORIC | HISTORIC | HISTORIC | HISTORIC | HISTORIC | HISTORIC | HISTORIC | HISTORIC | CURRENT | FOREC | FOREC | FOREC | FOREC |
| A1 | Directly incurred: Staff | £1,000,000.00 | £20,000.00 | £20,000.00 | £50,000.00 | £50,000.00 | £100,000.00 | £100,000.00 | £100,000.00 | £100,000.00 | £100,000.00 | £1,300,000.00 | £1,300,000.00 | £1,300,000.00 | £1,300,000.00 |
| A2 | Directly incurred: Travel & subsistence | £150,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 | £6,000.00 |
| A3 | Directly incurred: Equipment | £1,500,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 | £50,000.00 |
| A4 | Directly incurred: Other cost | £400,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 | £20,000.00 |
| A4a | Directly incurred: Exceptions Other | £100,000.00 | | £5,000.00 | | | | | | £20,000.00 | | | | | |
| A5 | Directly allocated: Investigators | £300,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 | £30,000.00 |
| A6 | Directly allocated: Estates | £300,000.00 | £10,000.00 | | £10,000.00 | | | £10,000.00 | | £10,000.00 | | | | | £10,000.00 |
| A7 | Directly allocated: Other cost | £100,000.00 | | | | | | | | | | | | | |
| A8 | Indirect costs | £100,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 | £15,000.00 |
| A9 | Exceptions: Staff | £20,000.00 | | | | | | | | | | | | | |
| A10 | Exceptions: Travel & Subsistence | £10,000.00 | | £2,000.00 | | | | £2,000.00 | | | | | £2,000.00 | | |
| A11 | Exceptions: Student Internships | £10,000.00 | | | | | | | | | | | | | |
| A12 | Exceptions: Other cost | £20,000.00 | | | | | | | | | | | | | |
| Total Cost (for each claim) | | £4,010,000.00 | £151,000.00 | £148,000.00 | £181,000.00 | £173,000.00 | £251,000.00 | £223,000.00 | £1,431,000.00 | £1,431,000.00 | £1,431,000.00 | £1,431,000.00 | £1,431,000.00 | £1,431,000.00 | £1,431,000.00 |
| Total Cost (cumulative) | | | £151,000.00 | £299,000.00 | £480,000.00 | £653,000.00 | £904,000.00 | £1,127,000.00 | £2,558,000.00 | £2,558,000.00 | £2,558,000.00 | £2,558,000.00 | £2,558,000.00 | £2,558,000.00 | £2,558,000.00 |



Word Sense Disambiguation

Identify correct word meaning or ‘word sense’ for ambiguous words:

*To add additional columns, drag the field to the [**table ACCOUNTING**] until you see an insertion point.*

*Whether on the mantelpiece or on the dining [**table FURNITURE**], vases are a beautiful addition to any home.*



Word Senses

Word senses are defined by a **semantic lexicon** such as WordNet

A semantic lexicon defines word senses **explicitly**, by use of lists

Word senses can also be defined **implicitly** by use of clusters

Here we focus on explicitly defined word senses

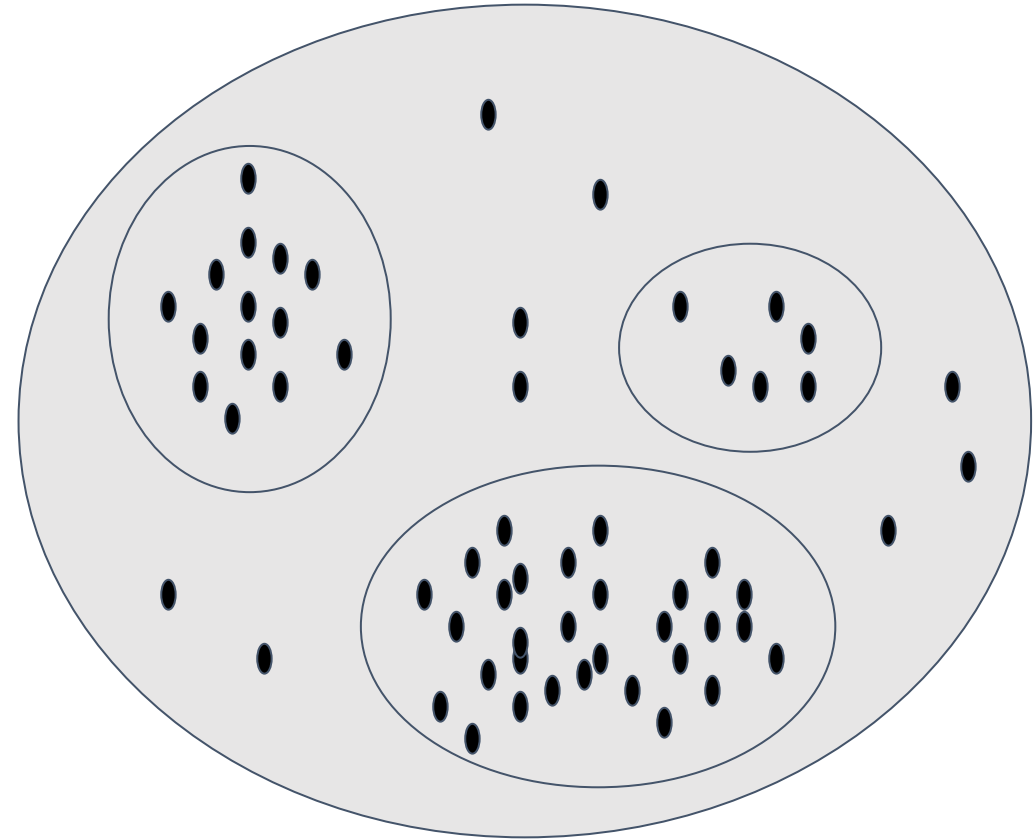
Explicit vs. Implicit Word Senses

table

Sense 1: FURNITURE

Sense 2: ACCOUNTING

Sense 3: GEOGRAPHICAL



WordNet Senses for *table*

Remember that WordNet uses 'synsets' to express word senses

- S: (n) **table#1**, tabular array#1 (a set of data arranged in rows and columns) "*see table 1*"
- S: (n) **table#2** (a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs) "*it was a sturdy table*"
- S: (n) **table#3** (a piece of furniture with tableware for a meal laid out on it) "*I reserved a table at my favorite restaurant*"
- S: (n) mesa#1, **table#4** (flat tableland with steep edges) "*the tribe was relatively safe on the mesa but they had to descend into the valley for water*"
- S: (n) **table#5** (a company of people assembled at a table for a meal or game) "*he entertained the whole table with his witty remarks*"
- S: (n) board#4, **table#6** (food or meals in general) "*she sets a fine table*"; "*room and board*"

Data Labeling - WordNet Glosses

| Word | Sense | SynSets | WordNet 'glosses' (examples) |
|--------------|-------|--------------------------|---|
| <i>table</i> | #1 | table#1, tabular array#1 | see <i>table</i> 1 |
| | #2 | table#2 | <i>it was a sturdy table</i> |
| | #3 | table#3 | <i>I reserved a table at my favourite restaurant</i> |
| | #4 | table#4, mesa#1 | <i>the tribe was relatively safe on the table</i> |
| | #5 | table#5 | <i>he entertained the whole table</i> |
| | #6 | table#6, board#4 | <i>she sets a fine table</i> |

Data Labeling - Corpus Data

| Word | Sense | Corpus Data (manual annotation) |
|-------|-------|---|
| table | #1 | <p><i>To add additional columns, drag the field to the table until you see an insertion point.</i></p> <p><i>In addition to rows, you can also divide a table into columns.</i></p> |
| | #2 | <p><i>Whether on the mantelpiece or centre stage on the dining table, vases are a beautiful addition to any home.</i></p> <p><i>Our dining tables and chairs are of the highest quality, crafted from solid premium wood.</i></p> |

Basic Algorithm for WSD - 'Simplified Lesk'

function SIMPLIFIED LESK(*word*, *sentence*) **returns** best sense of *word*

best-sense \leftarrow most frequent sense for *word*

max-overlap \leftarrow 0

context \leftarrow set of words in *sentence*

for each *sense* **in** senses of *word* **do**

signature \leftarrow set of words in the gloss and examples of *sense*

overlap \leftarrow COMPUTE OVERLAP(*signature*, *context*)

if *overlap* > *max-overlap* **then**

max-overlap \leftarrow *overlap*

best-sense \leftarrow *sense*

end

return(*best-sense*)

Jurafsky & Martin p365

Lesk Algorithm – Example WSD of *table*

*The toolbar includes a basic **table** drawing feature that lets you divide the canvas into rows and columns that you can use to organize your text.*

Lesk Algorithm – WordNet Senses

*The toolbar includes a basic **table** drawing feature that lets you divide the canvas into rows and columns that you can use to organize your text.*

| Word | Sense | SynSets |
|-------|-------|--------------------------|
| table | #1 | table#1, tabular array#1 |
| | #2 | table#2 |



Lesk Algorithm – Signature

*The toolbar includes a basic **table** drawing feature that lets you divide the canvas into rows and columns that you can use to organize your text.*

| Word | Sense | Signature (from corpus data) |
|--------------|-------|--|
| <i>table</i> | #1 | <i>add additional columns divide drag field insertion point rows see</i> |
| | #2 | <i>beautiful centre chairs crafted dining highest home mantelpiece premium quality solid stage sturdy vases wood</i> |

Lesk Algorithm – Overlap

*The toolbar includes a basic **table** drawing feature that lets you divide the canvas into rows and columns that you can use to organize your text.*

| Word | Sense | Signature |
|------|-------|-----------|
|------|-------|-----------|

| | | |
|--------------|----|--|
| <i>table</i> | #1 | <i>add additional <u>columns</u> <u>divide</u> drag field insertion point rows see</i> |
| | #2 | <i>beautiful centre chairs crafted dining highest home mantelpiece premium quality solid stage sturdy vases wood</i> |

Lesk Algorithm – Best Sense

*The toolbar includes a basic **table** drawing feature that lets you divide the canvas into rows and columns that you can use to organize your text.*

| Word | Sense | Signature |
|------|-------|-----------|
|------|-------|-----------|

| | | |
|--------------|----|--|
| table | #1 | <i>add additional <u>columns</u> <u>divide</u> drag field insertion point rows see</i> |
| | #2 | <i>beautiful centre chairs crafted dining highest home mantelpiece premium quality solid stage sturdy vases wood</i> |

WSD - Vector Space Model

WSD target:

*To add columns and rows, drag the field to the **table** until you see an insertion point.*

Corpus data for senses:

*In addition to rows, you can also divide a **table#1** into columns.*

*This **table#2** is of the highest quality, crafted from solid wood.*

Vocabulary (nouns in WSD target and corpus data):

[columns, rows, field, point, quality, wood]

Vectors (binary):

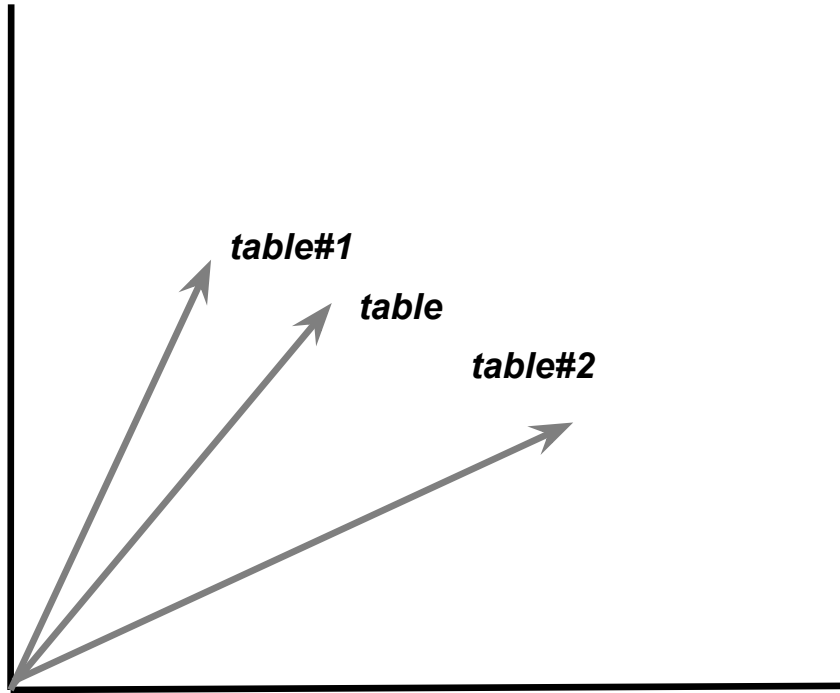
table ***[1,1,1,1,0,0]***

table#1 ***[1,1,0,0,0,0]***

table#2 ***[0,0,0,0,1,1]***



WSD - Vector Space Visualization



WSD Training

Supervised

- WordNet glosses

- Manual annotation of corpus data with word senses (see link below)

Semi-supervised

- 'Distant supervision' on Wikipedia categories

Unsupervised

- Graph-based WSD

Wikipedia Categories - *table*



WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Current events](#)
[Random article](#)
[About Wikipedia](#)
[Contact us](#)
[Donate](#)

[Contribute](#)

[Help](#)
[Learn to edit](#)
[Community portal](#)
[Recent changes](#)
[Upload file](#)

[Tools](#)

[What links here](#)
[Related changes](#)
[Special pages](#)

 Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#)

[Read](#)

[Edit](#)

[View history](#)



Table


From Wikipedia, the free encyclopedia

For use of tables to display information in Wikipedia, see [Wikipedia:Tables](#).

"Tabular" redirects here. For the typewriter key, see [tab key](#). For the bone at the back of the skull, see [tabular bone](#).

Table may refer to:

- [Table \(furniture\)](#), a piece of furniture with a flat top and one or more legs
- [Table \(information\)](#), a data arrangement with rows and columns
- [Table \(database\)](#)
- [Calligra Tables](#), a spreadsheet application
- [Mathematical table](#)
- [Table \(landform\)](#)
- [Table \(parliamentary procedure\)](#)
- [Tables \(board game\)](#)
- [The Table](#), a volcanic tuya in British Columbia, Canada
- Table, surface of the [sound board \(music\)](#) of a string instrument
- *[Al-Ma'ida](#)*, the fifth *sura* of the Qur'an, usually translated as "The Table"
- [Water table](#)

 Look up [table](#) or [tabular](#) in Wiktionary, the free dictionary.



NUI Galway
OÉ Gaillimh

<https://en.wikipedia.org/wiki/Table>

table as “furniture, information, landform”

[https://en.wikipedia.org/wiki/Table_\(furniture\)](https://en.wikipedia.org/wiki/Table_(furniture))

A **table** is an item of furniture with a flat top and one or more legs, used as a surface for working at, eating from or on which to place things. ... There are also a range of specialized types of **tables**, such as drafting **tables**, used for doing architectural drawings, and sewing **tables**.

[https://en.wikipedia.org/wiki/Table_\(information\)](https://en.wikipedia.org/wiki/Table_(information))

A **table** is an arrangement of data in rows and columns, or possibly in a more complex structure. ... **Tables** appear in print media, handwritten notes, computer software, architectural ornamentation, traffic signs, and many other places. ... In books and technical articles, **tables** are typically presented apart from the main text in numbered and captioned floating blocks.

[https://en.wikipedia.org/wiki/Table_\(landform\)](https://en.wikipedia.org/wiki/Table_(landform))

The term "flat" is relative when speaking of **tables**, and often the naming or identification of a **table** (or **table** hill or mountain) is based on the appearance of the terrain feature from a distance or from below it.



Labelled Data by Wikipedia Category

[https://en.wikipedia.org/wiki/Table_\(furniture\)](https://en.wikipedia.org/wiki/Table_(furniture))

A **table:furniture** is an item of furniture with a flat top and one or more legs, used as a surface for working at, eating from or on which to place things. ... There are also a range of specialized types of **tables:furniture**, such as drafting **tables:furniture**, used for doing architectural drawings, and sewing **tables:furniture**.

[https://en.wikipedia.org/wiki/Table_\(information\)](https://en.wikipedia.org/wiki/Table_(information))

A **table:information** is an arrangement of data in rows and columns, or possibly in a more complex structure. ... **Tables:information** appear in print media, handwritten notes, computer software, architectural ornamentation, traffic signs, and many other places. ... In books and technical articles, **tables:information** are typically presented apart from the main text in numbered and captioned floating blocks.

[https://en.wikipedia.org/wiki/Table_\(landform\)](https://en.wikipedia.org/wiki/Table_(landform))

The term "flat" is relative when speaking of **tables:landform**, and often the naming or identification of a **table:landform** (or **table:landform** hill or mountain) is based on the appearance of the terrain feature from a distance or from below it.



Aligning Wikipedia & WordNet for WSD

Wikipedia: Table_(furniture) - WordNet table#2

A **table#2** is an item of furniture with a flat top and one or more legs, used as a surface for working at, eating from or on which to place things.

Wikipedia: Table_(information) - WordNet table#1

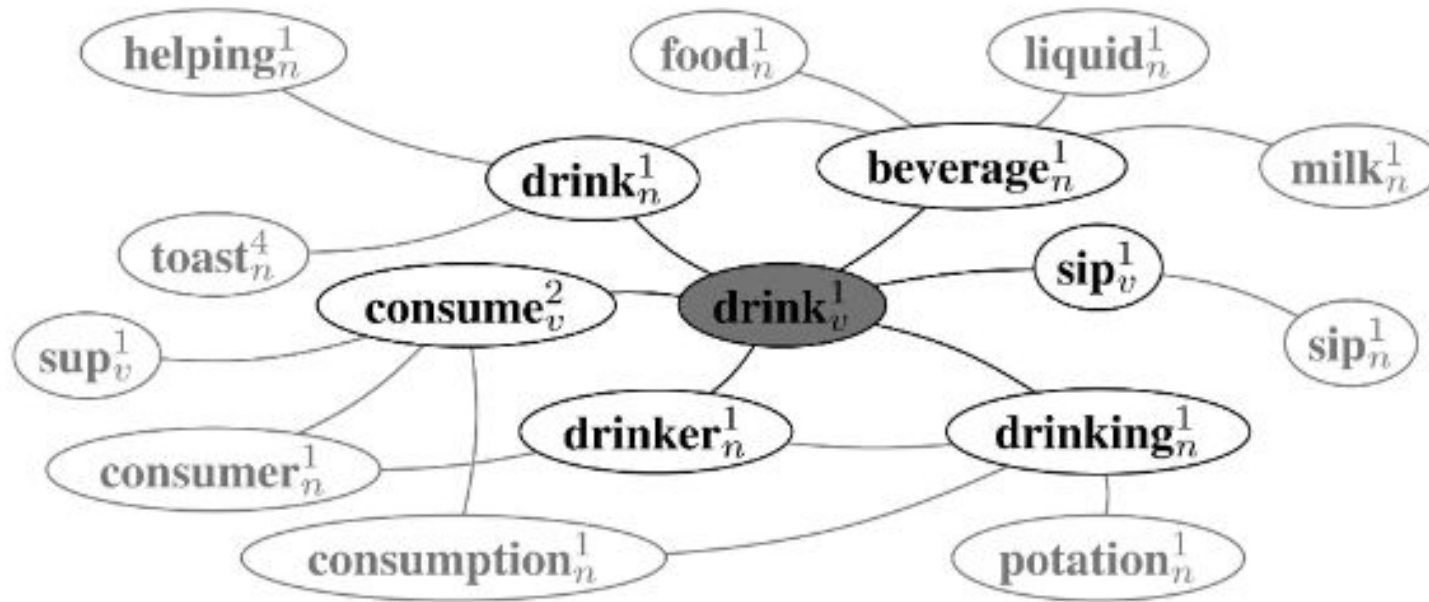
A **table#1** is an arrangement of data in rows and columns, or possibly in a more complex structure.

Wikipedia: Table_(landform) - WordNet table#4

The term "flat" is relative when speaking of **tables#4**, and often the naming or identification of a **table#4** (or **table#4** hill or mountain) is based on the appearance of the terrain feature from a distance or from below it.

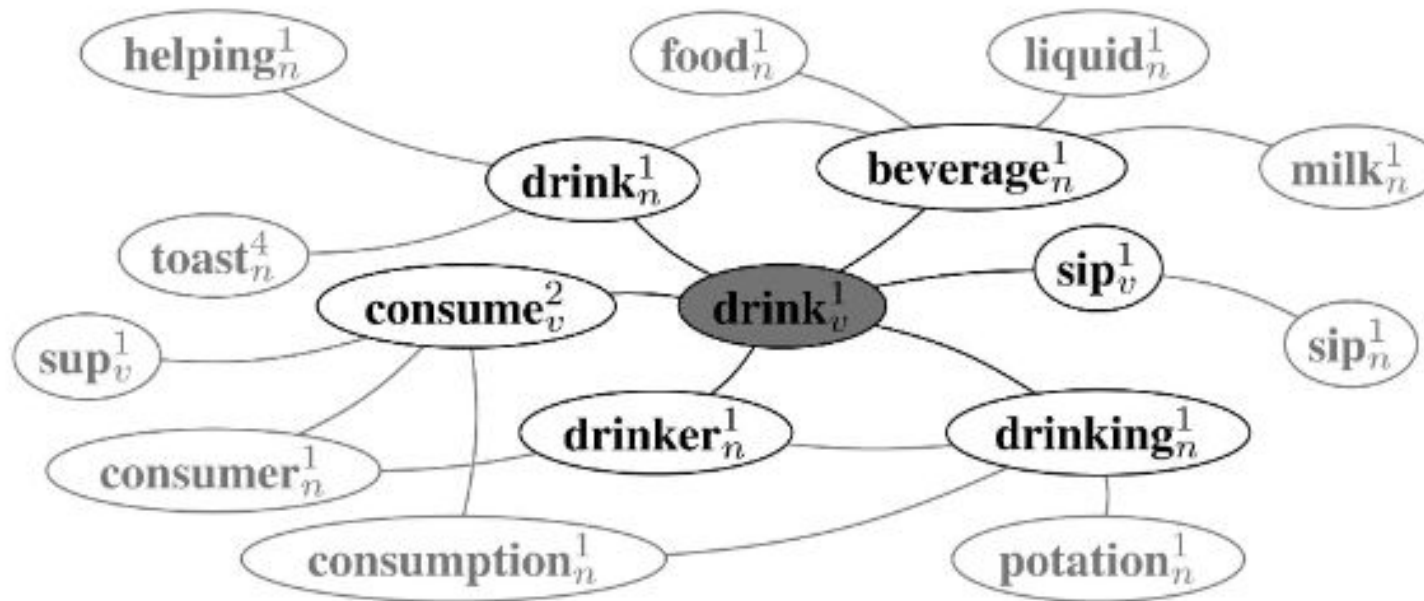
Graph-based Methods for WSD

WordNet graph for the verb **drink** with word sense: **drink#1, imbibe#3 - take in liquids**



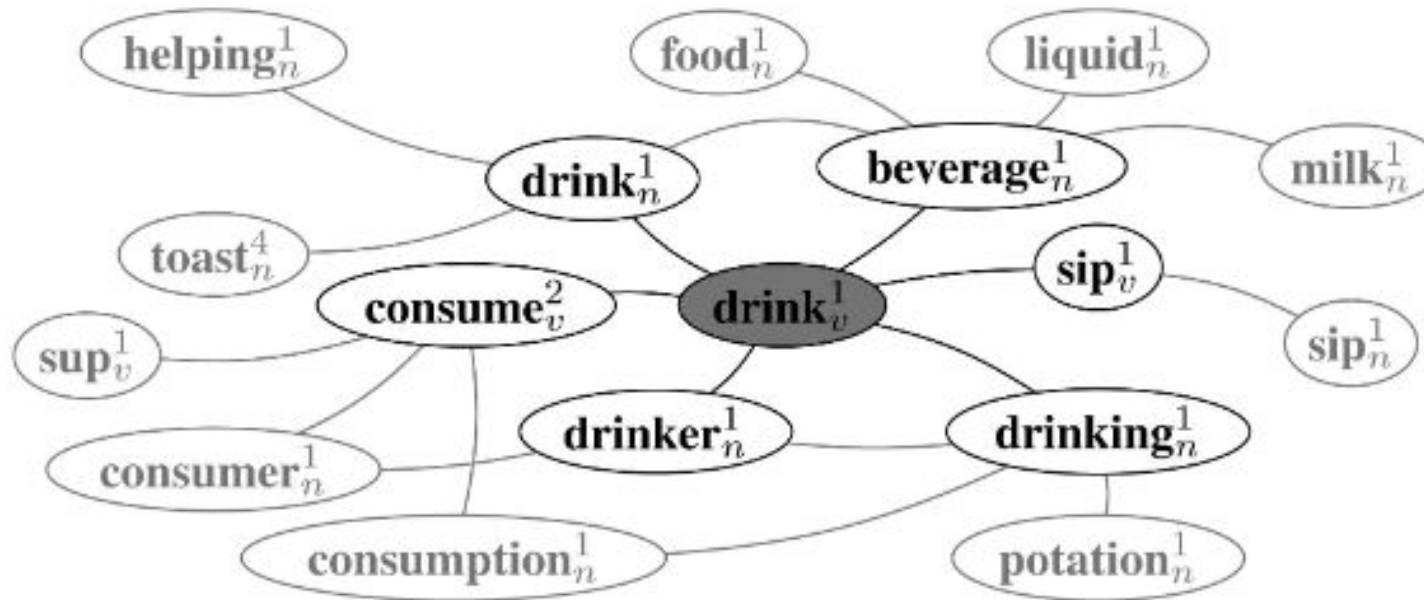
Graph-based Methods for WSD

Can we use this graph to disambiguate words in the sentence “*She drank some milk.*”

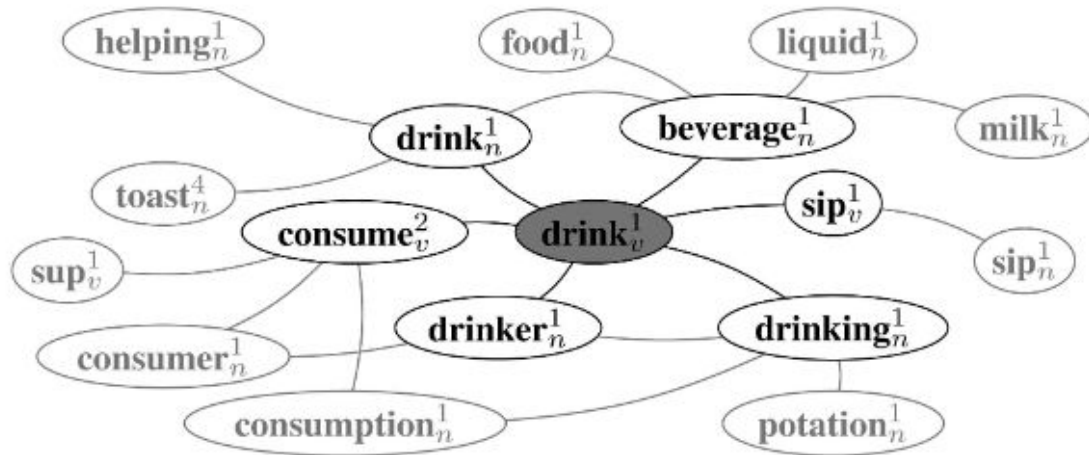


Graph-based Methods for WSD

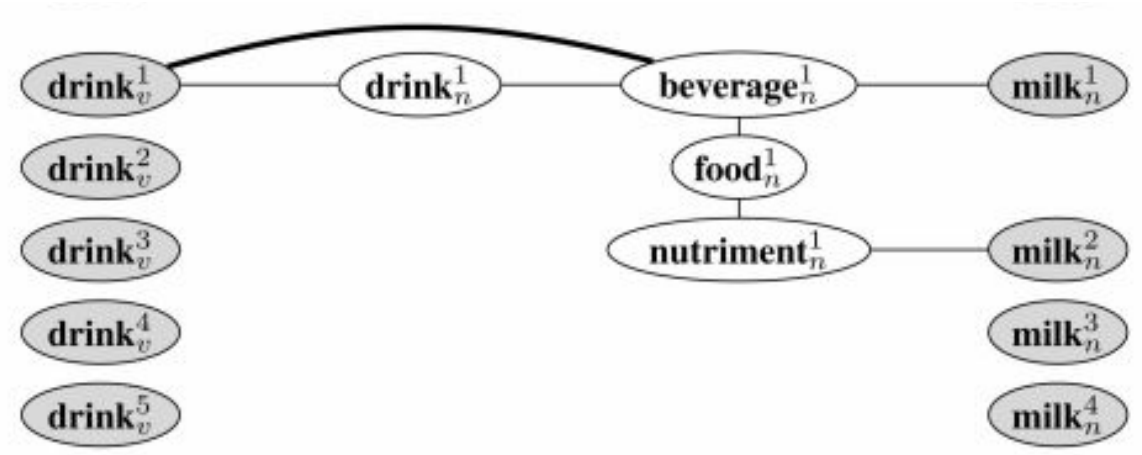
Identify shortest path between “*drank*” (verb “*to drink*”) and “*milk*” (noun)



Graph-based Methods for WSD

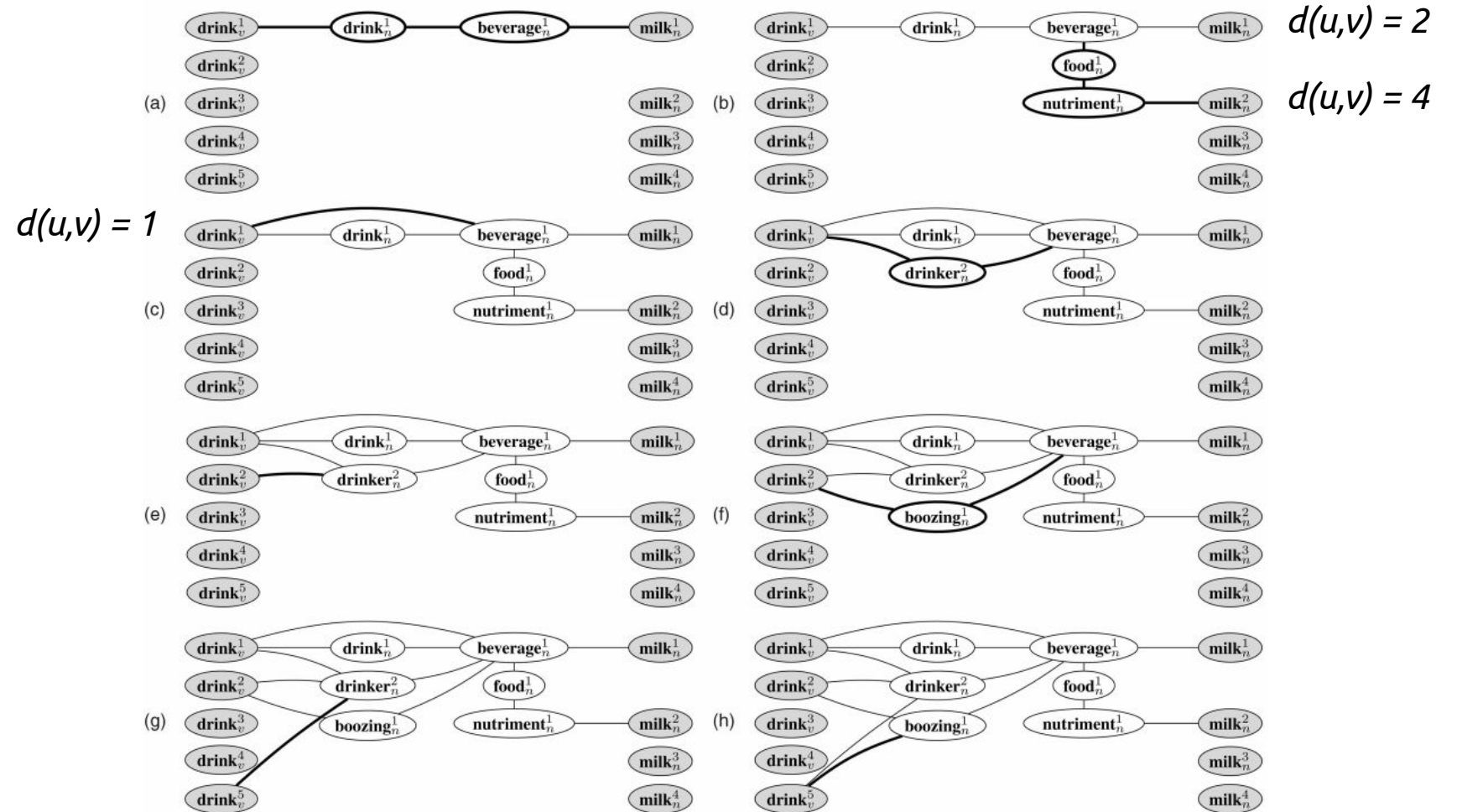


$$d(u, v) = 1$$



Graph-based Methods for WSD

"She drank some milk."



Overview

Semantic Analysis

Word Sense Disambiguation

Semantic Role Labelling

Coreference Resolution

Entities, Roles & Events

Consider the following sentence

“Kristina hit Scott with a baseball.”

There are several actors/items (‘entities’) with different ‘roles’ in the *hitting* ‘event’

Kristina

Scott

baseball



Entities, Roles & Events

Different sentences with 'entities' in the same 'roles' can express the same 'event'

Kristina hit *Scott* with a baseball.

Scott was hit by *Kristina* with a baseball.

Scott was hit with a baseball by *Kristina*.

Semantic Roles

We can define 'semantic roles' for each entity of an event expressed by a sentence, e.g.

AGENT (**AGT**), EXPERIENCER (**EXP**), INSTRUMENT (**INS**)

[Kristina AGT] hit [Scott EXP] [with a baseball INS] .

[Scott EXP] was hit by [Kristina AGT] [with a baseball INS] .

[Scott EXP] was hit [with a baseball INS] by [Kristina AGT].



Semantic Roles - Overview

| | |
|-------------|---|
| AGENT | The volitional causer of an event |
| EXPERIENCER | The experiencer of an event |
| FORCE | The non-volitional causer of the event |
| THEME | The participant most directly affected by an event |
| RESULT | The end product of an event |
| CONTENT | The proposition or content of a propositional event |
| INSTRUMENT | An instrument used in an event |
| BENEFICIARY | The beneficiary of an event |
| SOURCE | The origin of the object of a transfer event |
| GOAL | The destination of an object of a transfer event |

The waiter spilled the soup.

John has a headache.

The wind blows debris from the mall into our yards.

Only after Benjamin Franklin broke the ice...

The city built a regulation-size baseball diamond...

Mona asked "You met Mary Ann at a supermarket?"

He poached catfish, stunning them with a shocking device...

Whenever Ann Callahan makes hotel reservations for her boss...

I flew in from Boston.

I drove to Portland.

Note: semantic roles are also referred to as 'thematic roles'

Jurafsky & Martin p374



Semantic Roles - Use

Semantic roles enable ‘event identification’ across different sentences, which can be very useful in applications such as Question Answering, e.g.

Question: *“Which company bought Google last year?”*

Answer Candidates: *“Fitbit bought by Google in 2019”*

“Last year it was Fitbit that Google bought”

Semantic Roles - Challenges

Semantic roles are not very well defined with different names and interpretations of roles across the literature

Automatic 'semantic role labeling' depends largely on supervised training with limited labeled data existing

Semantic Role Labeling

Identification of semantic roles of each ‘argument’ of each ‘predicate’ in a sentence

Basic algorithm for Semantic Role Labeling (SRL):

```
function SEMANTICROLELABEL(words) returns labeled tree
```

```
  parse ← PARSE(words)
```

```
  for each predicate in parse do
```

```
    for each node in parse do
```

```
      featurevector ← EXTRACTFEATURES(node, predicate, parse)
```

```
      CLASSIFYNODE(node, featurevector, parse)
```

Jurafsky & Martin Ch20



Semantic Role Labeling - Feature Extraction

[AGENT The chancellor] announced [THEME new policy measures].

| | |
|--|--------------------------|
| Predicate: | <i>announce</i> |
| Voice (whether predicate is active or passive): | active |
| Position (before/after) of classified constituent (phrase) with regard to predicate: | before |
| Head Word of classified constituent (phrase): | <i>chancellor</i> |

Semantic Role Labeling - Feature Extraction

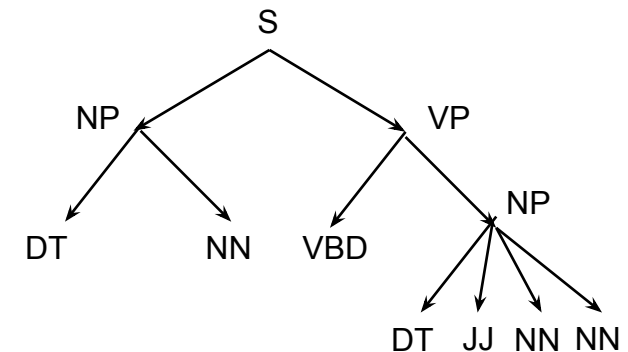
[AGENT The chancellor] announced [THEME new policy measures].

Phrase Type (NP, PP, ...) of classified constituent (phrase):

Minimal Path from classified constituent (phrase) to predicate:

Sub-categorization (phrase structure rule) for parent of predicate:

NP
NP↑S↓VP↓VBD
VP→VBD,NP



Semantic Role Labeling - Feature Extraction

[AGENT The chancellor] announced [THEME new policy measures].

| | |
|---|--------------------------|
| Predicate: | <i>announce</i> |
| Voice (whether predicate is active or passive): | active |
| Phrase Type (NP, PP, ...) of classified constituent: | NP |
| Minimal Path from classified constituent to predicate: | NP↑S↓VP↓VBD |
| Position (before/after) of classified constituent with regard to predicate: | before |
| Head Word of classified constituent: | <i>chancellor</i> |
| Sub-categorization for parent of predicate: | VP→VBD,NP |

AGENT [announce, active, NP, NP ↑S↓VP↓VBD, before, chancellor, VP→VBD,NP]

Overview

Semantic Analysis

Word Sense Disambiguation

Semantic Role Labelling

Coreference Resolution

Coreference Resolution

Identify words and/or phrases referring to the same entity

Application of coreference resolution for example in dialog systems:

USER: *"How late is the 404 bus to Dangan?"*

SYSTEM: *"The 404 bus to Dangan leaves at 12:35"*

USER: *"Does it stop at the university?"* ***it : bus***

SYSTEM: *"The bus to Dangan stops at the university"*

USER: *"When will it stop there?"* ***it : bus***

there : the university

Coreference Resolution - Example

“Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.”

Coreference Resolution - Mentions

Identify all mentions - coreference candidates

“Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.”



Coreference Resolution - Mentions

Mentions referring to **Barack Obama**

*“**Barack Obama** nominated Hillary Rodham Clinton as his secretary of state on Monday. **He** chose her because she had foreign affairs experience as a former First Lady.”*



Coreference Resolution - Mentions

Mentions referring to ***Hillary Rodham Clinton***

*“Barack Obama nominated **Hillary Rodham Clinton** as **his secretary of state** on Monday. He chose her because **she** had foreign affairs experience as **a former First Lady**.”*



Mention Detection and Mention Clustering

Mention Detection

“Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.”

Mention Clustering

“Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.”



Mention Detection - Kinds of Mentions

Mention is a span of text referring to an entity

There are three kinds of mentions:

- **Pronouns:** *I, your, it, she, him, ...*
- **Named Entities:** *person, location, organization, ...*
- **Noun Phrases:** *“a dog,” “the big fluffy cat stuck in the tree”*

Mention Detection - NLP Tools to Use

For detection of each kind of mention use:

- **Part-of-speech tagger** to identify pronouns
- **Named Entity Recognizer (NER) system** to identify named entities
- **Constituency (Phrase) parser** to identify noun phrases

Mention Clustering - Features

Recency - more recently mentioned entities are preferred

***John** went to a movie. **Jack** went as well. **He** was not busy.*

He = Jack (He \neq John)

Syntactic agreement

***Jack** gave **Mary** a gift. **She** was excited.*

She = Mary (She \neq Jack)

Other syntactic constraints

***John** bought **him** a new car.*

him \neq John

Semantic compatibility

*... the mining **conglomerate** ... the **company** ...
conglomerate*

company =

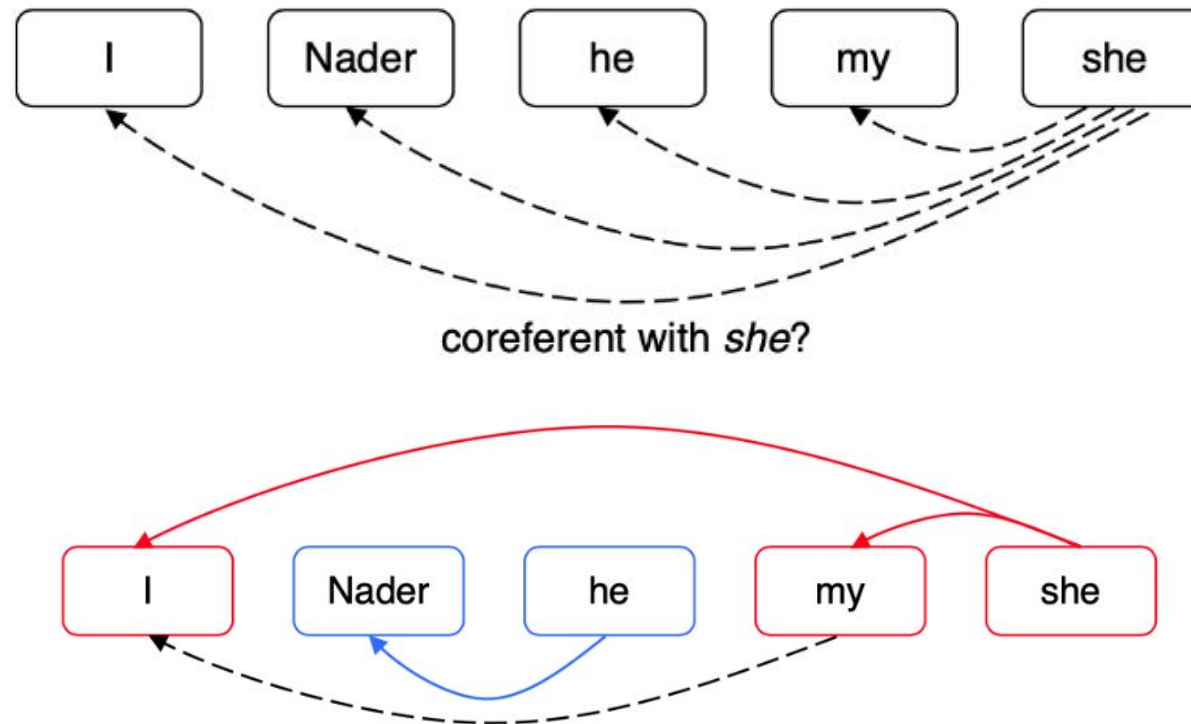
Grammatical Role - prefer entities in the subject position

***John** went to a movie with **Jack**. **He** was not busy.*

He = Jack (He \neq John)

Mention Clustering - Example

*"I voted for **Nader** because **he** was most aligned with **my** values," **she** said.*



Lab of this Week

Word Sense Disambiguation exercises with WordNet



NUI Galway
OÉ Gaillimh

QA

