Ollscoil na Gaillimhe
University of Galway

Introduction to NLP

Semantic Analysis

Dr. Paul Buitelaar
Data Science Institute, University of Galway

Includes updates by Omnia Zayed

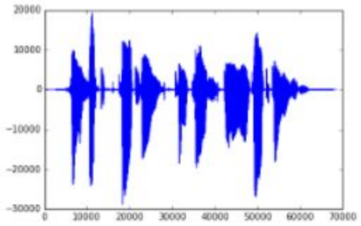
Learning Objectives

- Understand basic concepts and challenges in semantic analysis
- Understand and apply approaches to Word Sense Disambiguation
- Understand other tasks in semantic analysis: Semantic Role Labelling, Coreference Resolution

Remember

In Lecture 2 on Linguistic Concepts, we discussed the different **linguistic levels of analysis**

Phonology - acoustic signal

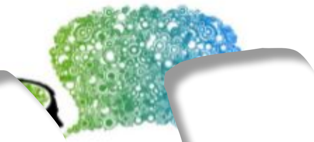


Morphology – word structure

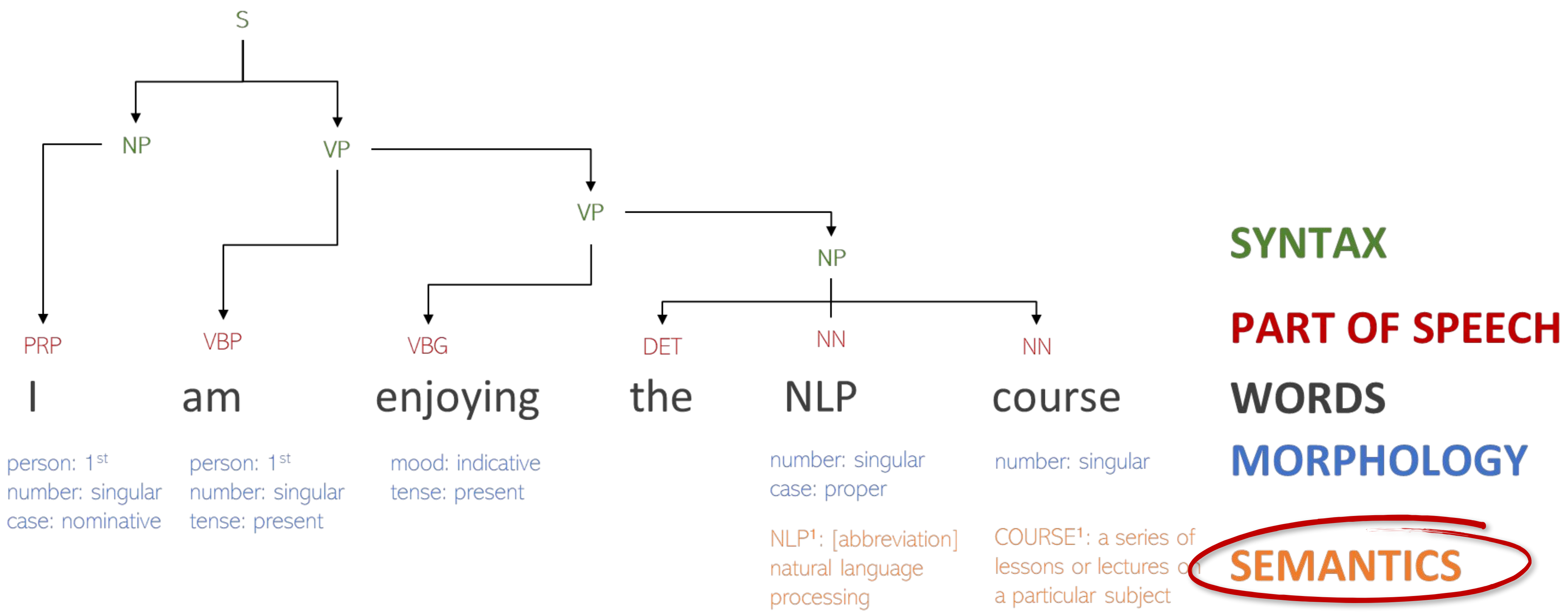
Syntax – sentence and phrase structure

Semantics – meaning and reference

Pragmatics - communication



Remember



S
NP
VP
VP
NP
PRP
VBP
VBG
DET
NN
NN
I
am
enjoying
the
NLP
course
person: 1st
number: singular
case: nominative
person: 1st
number: singular
tense: present
mood: indicative
tense: present
number: singular
case: proper
number: singular
NLP1: [abbreviation] natural language processing
COURSE1: a series of lessons or lectures on a particular subject
SYNTAX
PART OF SPEECH
WORDS
MORPHOLOGY
SEMANTICS

Remember

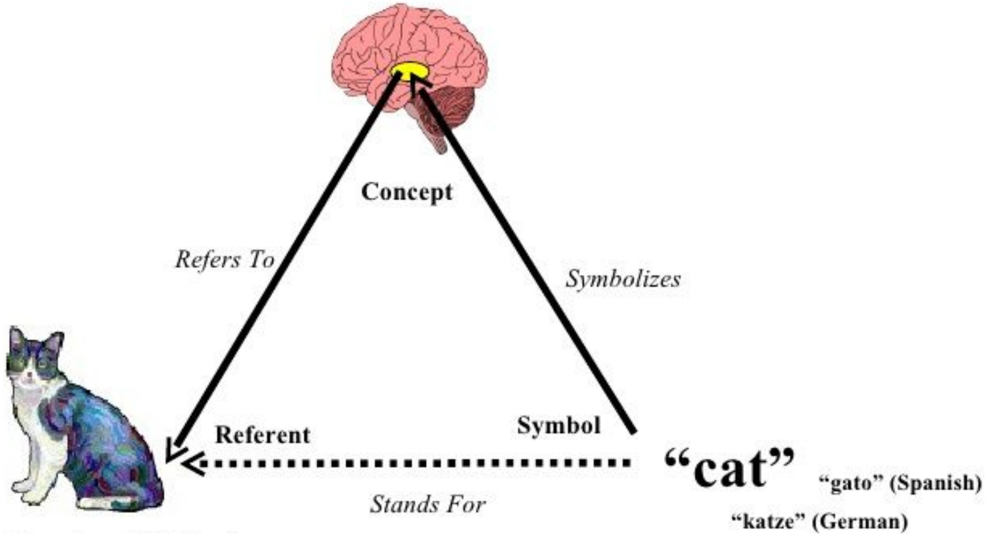
Semantic analysis of language data is concerned with

how linguistic expressions/symbols (words, phrases, sentences)

refer to concepts (entities, relations, events)

about referents (objects in the world)

Semantic relation *between linguistic symbols and concepts and how they refer to objects in the world is commonly referred to as* ***'meaning'***



Semantic Analysis

Semantic Analysis in Linguistics

Analyse how **linguistic expressions** (words, phrases, sentences) refer to **meaning representations** (concepts about objects in the world)

Semantic Analysis in NLP

Develop automatic methods for **constructing unambiguous meaning representations** for linguistic expressions

Levels of Semantic Analysis

Lexical Semantics

*XYZ corporation bought the **stock**.*

*She made gravy with a base of beef **stock**.*

*The store has a very low turnover of **stock**.*

Word Sense Disambiguation: identify correct word meaning or ‘sense’

Levels of Semantic Analysis

Compositional Semantics

*XYZ **corporation** <u>bought</u> the **stock**.*

*They <u>sold</u> the **stock** to XYZ **corporation**.*

*The **stock** was <u>bought</u> by XYZ **corporation**.*

*The <u>purchase</u> of the **stock** by XYZ **corporation**...*

*The **stock** <u>purchase</u> by XYZ **corporation**...*

Semantic Role Labelling: identify 'semantic role' for each word/phrase

Levels of Semantic Analysis

Discourse Semantics

__XYZ corporation__ bought the stock. __They__ had the backing from their bank.

XYZ corporation bought __the stock__. __It__ was valued rather high.

Coreference Resolution: identify which words/phrases refer to the same ‘entity’ across sentences

Buzz Groups

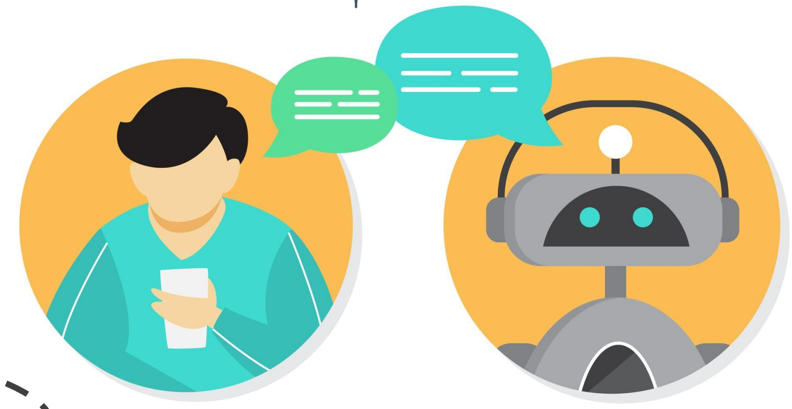
Why do we need semantic analysis?

Semantic Analysis in Applications - SRL, CR

"Google acquired Fitbit in January 2021"

"Fitbit bought by Google in 2021"

"Last year, Google completed its acquisition of Fitbit."



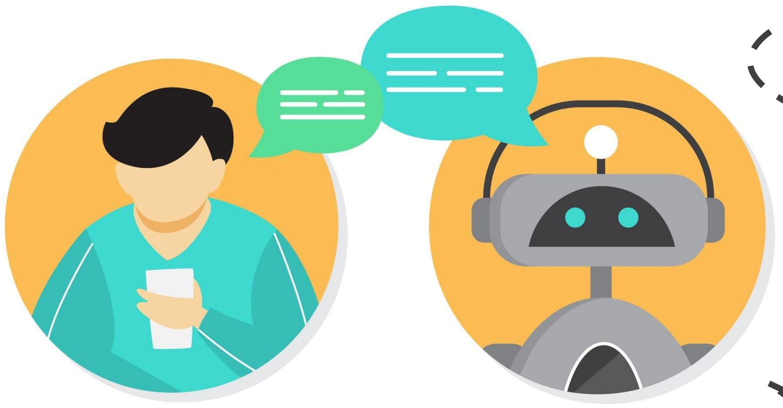
"Which company did Google buy last year?"

"Google bought Fitbit"

Question Answering and Chatbots

Images source https://towardsdatascience.com/automatic-question-answering-ac7593432842

Semantic Analysis in Applications - WSD



Question Answering and Chatbots

"Do you mean an overview of your interest in this company?"

"What is my interest here?"

"Do you mean the interest on your account"

Ollscoil na Gaillimhe
University of Galway

Word Sense Disambiguation

What is Word Sense Disambiguation?

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

table in the interpretation of ***furniture***

3 YEARS ACCOUNTING

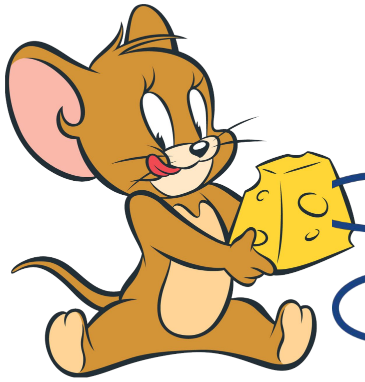
			MAR18 - MAY18	JUN18 - AUG18	SEP18 - NOV18	DEC18 - FEB19	MAR19 - MAY19	JUN19 - AUG19	SEP19 - N
			Q1	Q2	Q3	Q4	Q5	Q6	Q7
COST ITEM	DESCRIPTION	TOTAL BUDGET	HISTORIC	HISTORIC	HISTORIC	HISTORIC	HISTORIC	CURRENT	FORECA
A1	Directly incurred: Staff	£1,000,000.00	£20,000.00	£20,000.00	£50,000.00	£50,000.00	£100,000.00	£100,000.00	£1,300,0
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
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	£50,00
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	£20,00
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	£30,00
A6	Directly allocated: Estates	£300,000.00	£10,000.00		£10,000.00		£10,000.00		£10,00
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	£15,00
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**	**£1,431,0**
	Total Cost (cumulative)		**£151,000.00**	**£299,000.00**	**£480,000.00**	**£653,000.00**	**£904,000.00**	**£1,127,000.00**	**£2,558,0**

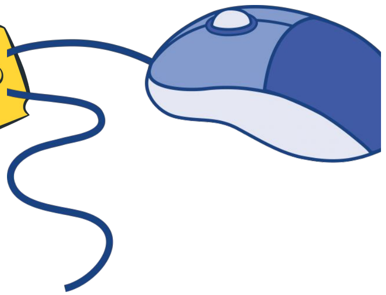
table in the interpretation of ***data***

Word Senses

mouse

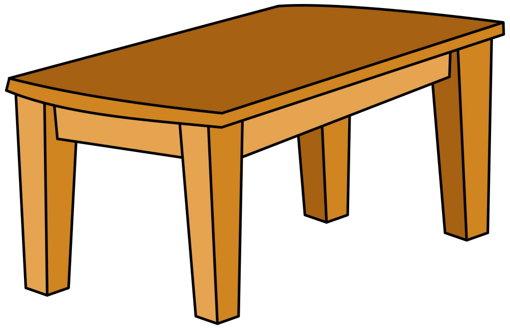
table

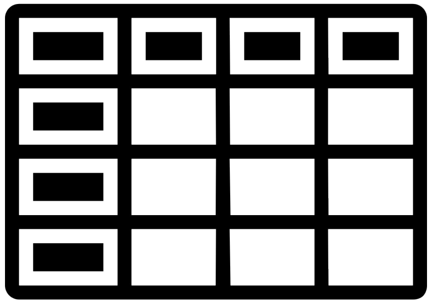




mouse#1
ANIMAL

mouse#2
OBJECT

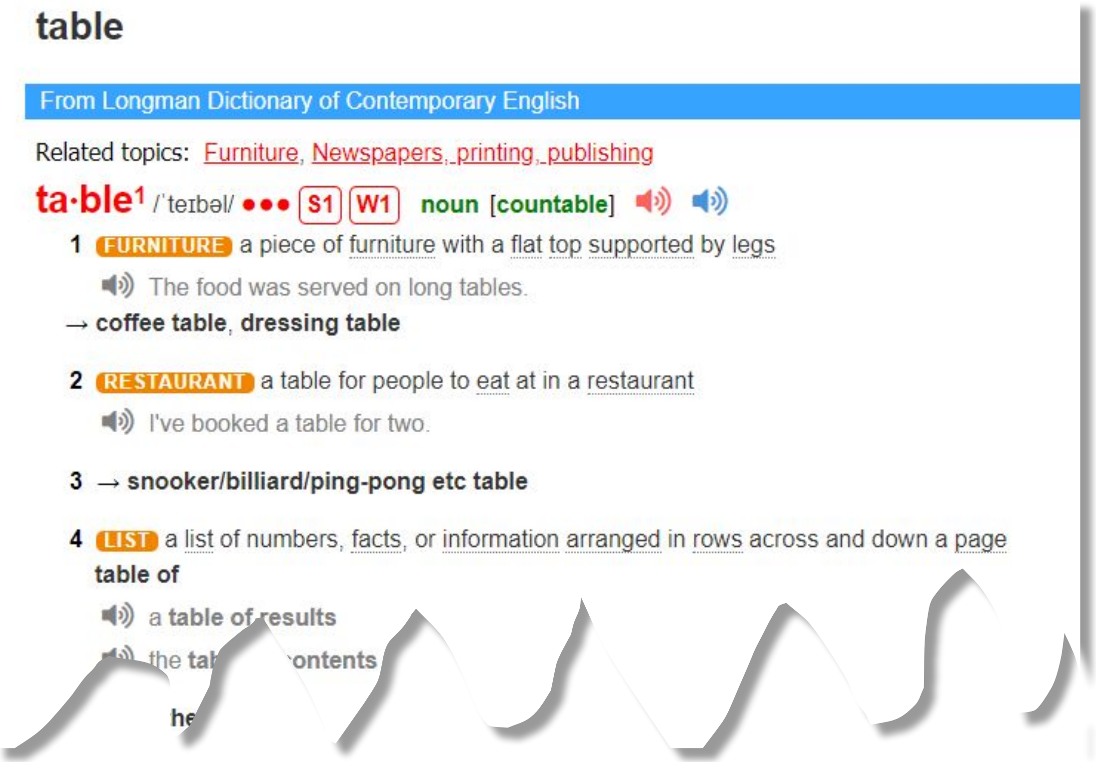




table#1
FURNITURE

table#2
DATA

Word Senses - Dictionary / Lexicon



table

From Longman Dictionary of Contemporary English

Related topics: Furniture, Newspapers, printing, publishing

ta·ble[1] /ˈteɪbəl/ ●●● S1 W1 noun [countable]

1 FURNITURE a piece of furniture with a flat top supported by legs

The food was served on long tables.

→ coffee table, dressing table

2 RESTAURANT a table for people to eat at in a restaurant

I've booked a table for two.

3 → snooker/billiard/ping-pong etc table

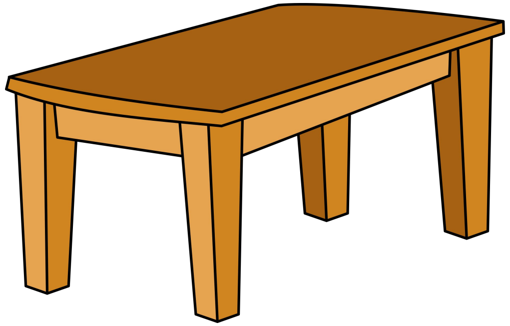
4 LIST a list of numbers, facts, or information arranged in rows across and down a page

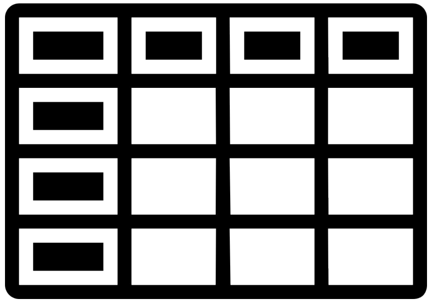
table of

a table of results

the tab ontents

he





table#1
FURNITURE

table#2
DATA

Word Senses - Summary

- A sense, or **word sense**, is a discrete representation of one meaning of a word
- Words can have different meanings due to **homonymy** (same sound or spelling but different meaning)
- Word senses can be defined explicitly in a **semantic lexicon** such as WordNet

Word Senses - Explicit vs. Implicit

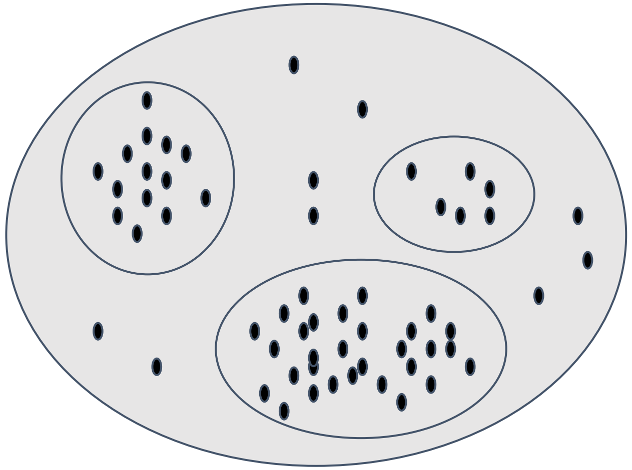
Use lists to define **explicit** word senses (symbolic)

Use clusters to define **implicit** word senses (sub-symbolic)

table#1: FURNITURE

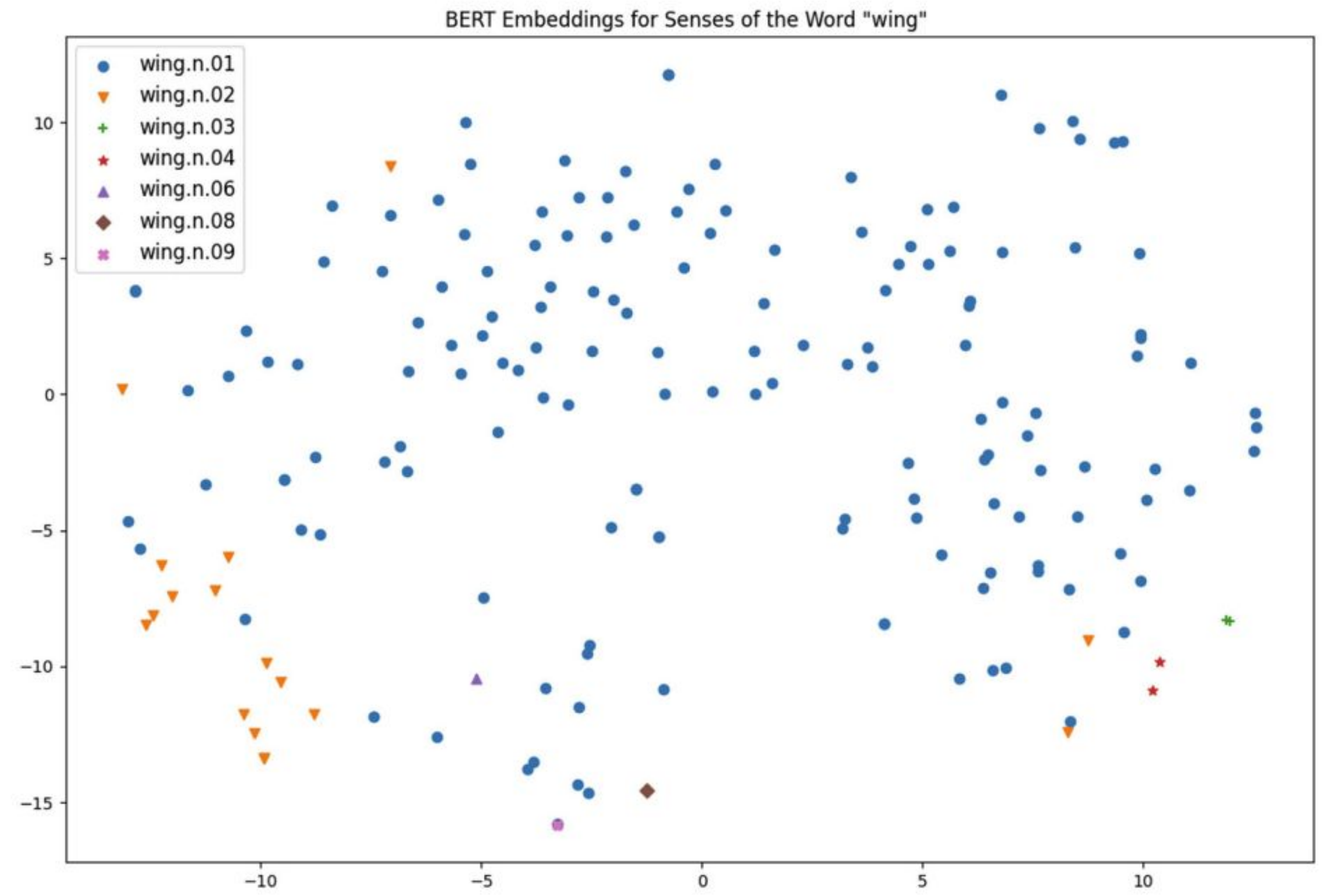
table#2: ACCOUNTING

table#3: GEOGRAPHICAL



Here we focus on explicitly defined (symbolic) word senses as in WordNet

Word Senses - Clustering



McCrae, J. P., Fransen, T., Ahmadi, S., Buitelaar, P., & Goswami, K. (2022). Towards an Integrative Approach for Making Sense Distinctions. *Frontiers in Artificial Intelligence*, 3.

Word Senses - not only Nouns

*They rarely **serve** red meat, preferring to prepare seafood.*

*He **served** as U.S. ambassador to Norway in 1976 and 1977.*

*He might have **served** his time, come out and led an upstanding life.*

Word Senses in WordNet

Remember that WordNet uses ‘synsets’ to define word senses

Word Senses in WordNet

*… drag the field to the **table** until you see an insertion point …*
*… on the dining **table**, vases are a beautiful addition to any home …*

- 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"*

Word Sense Disambiguation

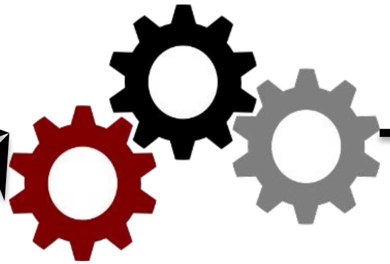
… drag the field to the ***[table#1]*** *until you see an insertion point …*
… on the dining ***[table#3]****, vases are a beautiful addition to any home …*

- 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"*

WSD Task

a **word in context**

*An electric guitar and **bass** player stand off to one side*



WSD Algorithm

disambiguated (correct) word sense

*An electric guitar and **bass**[7 INSTRUMENT] player stand off to one side*

fixed **inventory of word senses**

- S: (n) **bass#1** (the lowest part of the musical range)
- S: (n) **bass#2**, bass part#1 (the lowest part in polyphonic music)
- S: (n) **bass#3**, basso#1 (an adult male singer with the lowest voice)
- S: (n) sea bass#1, **bass#4** (the lean flesh of a saltwater fish of the family Serranidae)
- S: (n) freshwater bass#1, **bass#5** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- S: (n) **bass#6**, bass voice#1, basso#2 (the lowest adult male singing voice)
- S: (n) **bass#7** (the member with the lowest range of a family of musical instruments)
- S: (n) **bass#8** (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Ollscoil na Gaillimhe
University of Galway

Buzz Groups

Can you think of different approaches to WSD?

Approaches to WSD

Unsupervised

- Knowledge-based
- Graph-based

Supervised

- Word Sense Annotated Corpora

Semi-supervised

- Distant supervision

Image source: https://www.coursera.org/learn/machine-learning

Approaches to WSD

Unsupervised

- Knowledge-based
- Graph-based

Supervised

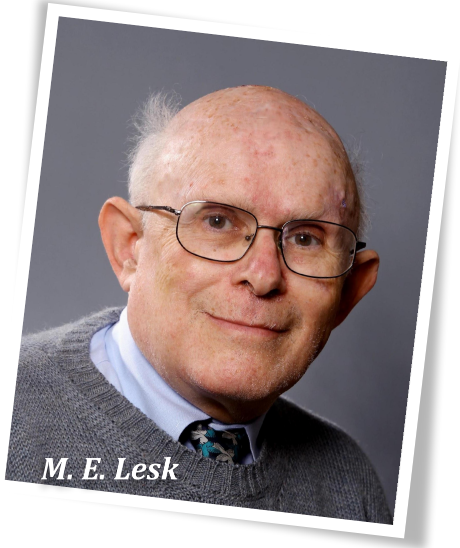
- Word Sense Annotated Corpora

Semi-supervised

- Distant supervision

Image source: https://www.coursera.org/learn/machine-learning

Unsupervised WSD - Knowledge-based



Lesk

Oldest, most effective baseline **WSD algorithm**

Uses a dictionary (e.g. WordNet) as **knowledge base**

Choose the sense that shares the most words in its signature (dictionary definition) with the context of the target word

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

(Simplified) Lesk Algorithm

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

best-sense ← most frequent sense for *word*
max-overlap ← 0
context ← set of words in *sentence*
for each *sense* **in** senses of *word* **do**
  signature ← set of words in the gloss and examples of *sense*
  overlap ← COMPUTEOVERLAP(*signature*, *context*)
  if *overlap* > *max-overlap* **then**
    max-overlap ← *overlap*
    best-sense ← *sense*
end
return(*best-sense*)

Jurafsky & Martin, 3rd ed., p.367

Lesk Algorithm – Example

*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 – Dictionary Definition

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**	**dictionary definition - here WordNet gloss**
table	#1	*‘a set of data arranged in rows and columns’*
	#2	*‘a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs’*

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 (e.g. nouns in dictionary definition - here WordNet gloss)**
table	#1	*'a set of data arranged in rows and columns'*
	#2	*'a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs'*

Lesk Algorithm – Context/Signature Overlap

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

Word	Sense	Signature	
table	#1	*data <u>rows</u> <u>columns</u>*	**2**
	#2	*furniture top legs*	**0**

Lesk Algorithm – Best Sense

*The toolbar includes a basic [**table#1**] 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

Approaches to WSD

Unsupervised

- Knowledge-based
- Graph-based

Supervised

- Word Sense Annotated Corpora

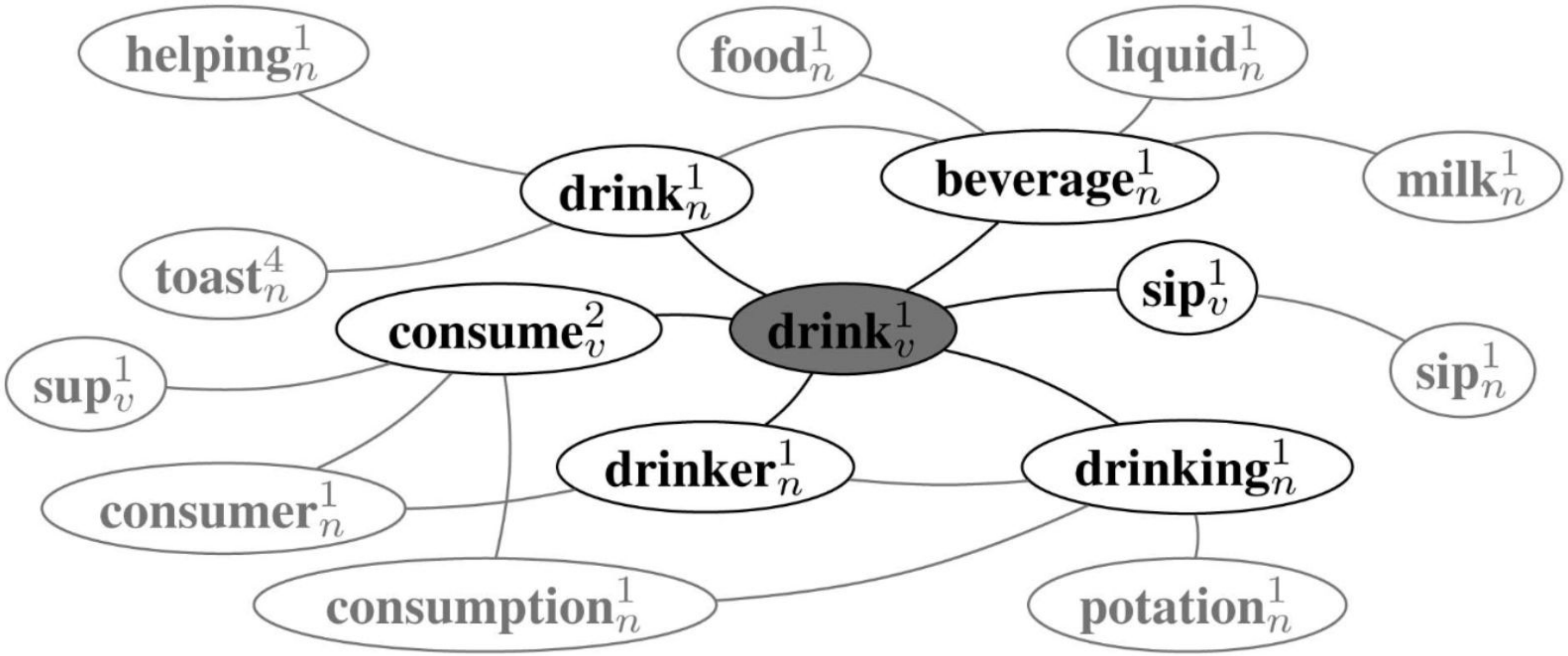
Semi-supervised

- Distant supervision

Image source: https://www.coursera.org/learn/machine-learning

Unsupervised WSD using Graph Distance

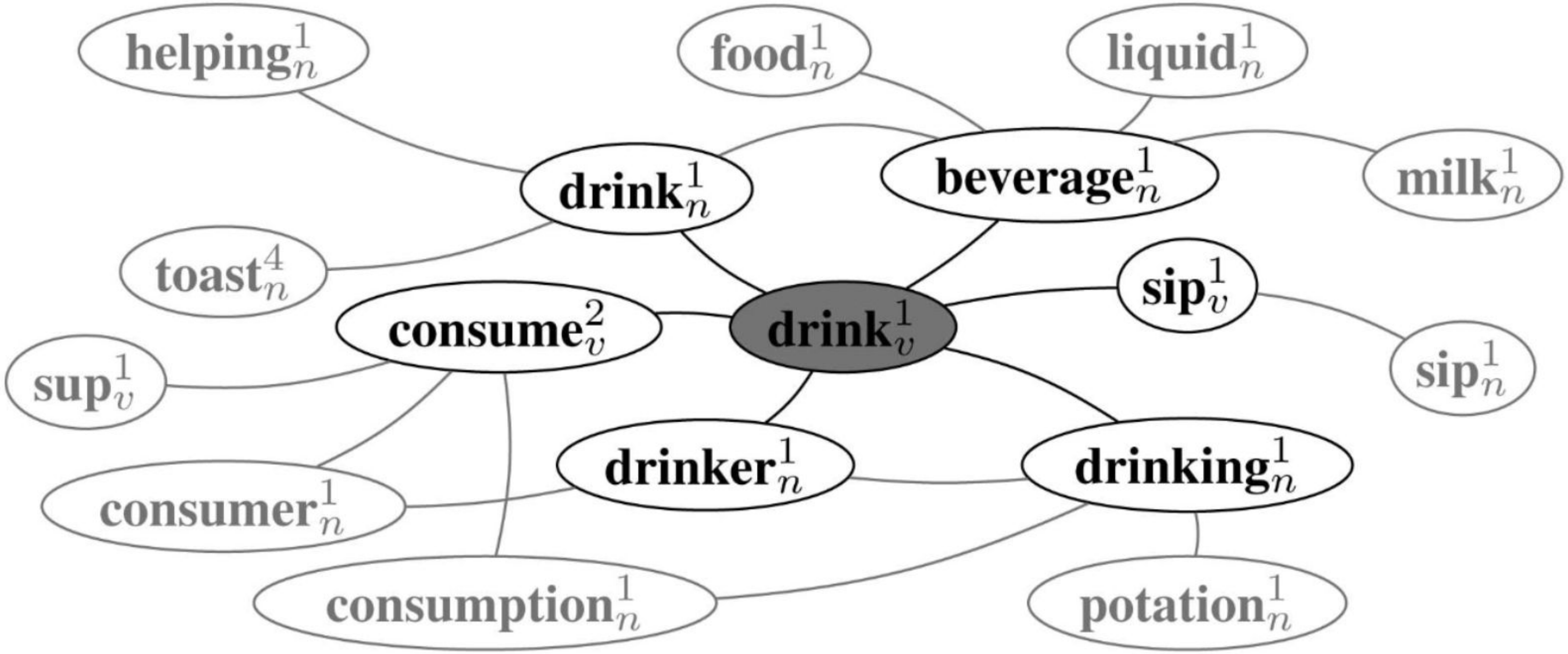
WordNet graph for word sense ***drink (verb) #1: 'take in liquids'***



Navigli, R. and Lapata, M. (2010). An experimental study of graph connectivity for unsupervised word sense disambiguation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(4), 678–692

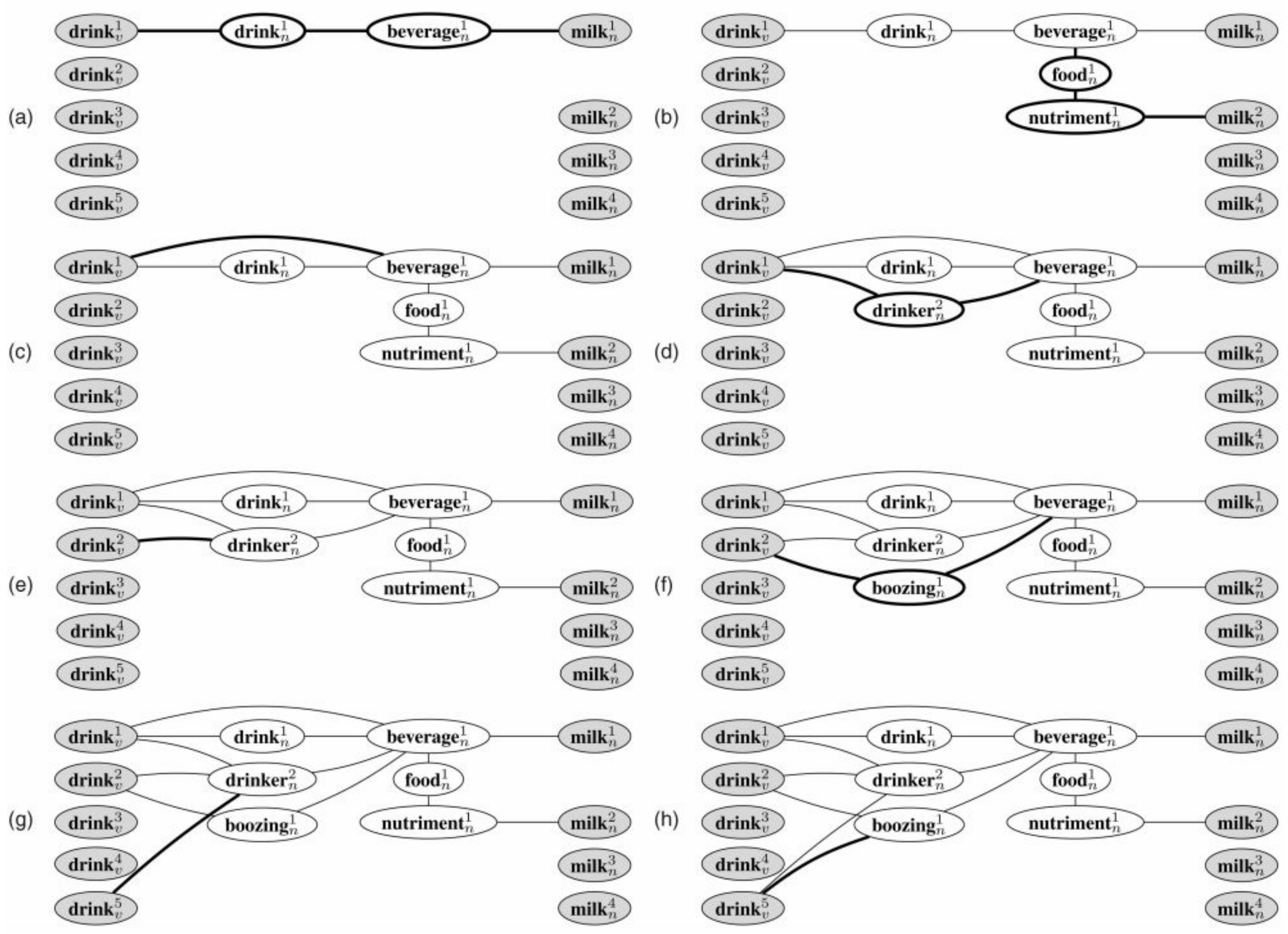
Graph-based Methods for WSD

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



Graph Construction

*“She **drank** some **milk**.”*



Paths in the Graph with Distance

*"She **drank** some **milk**."*

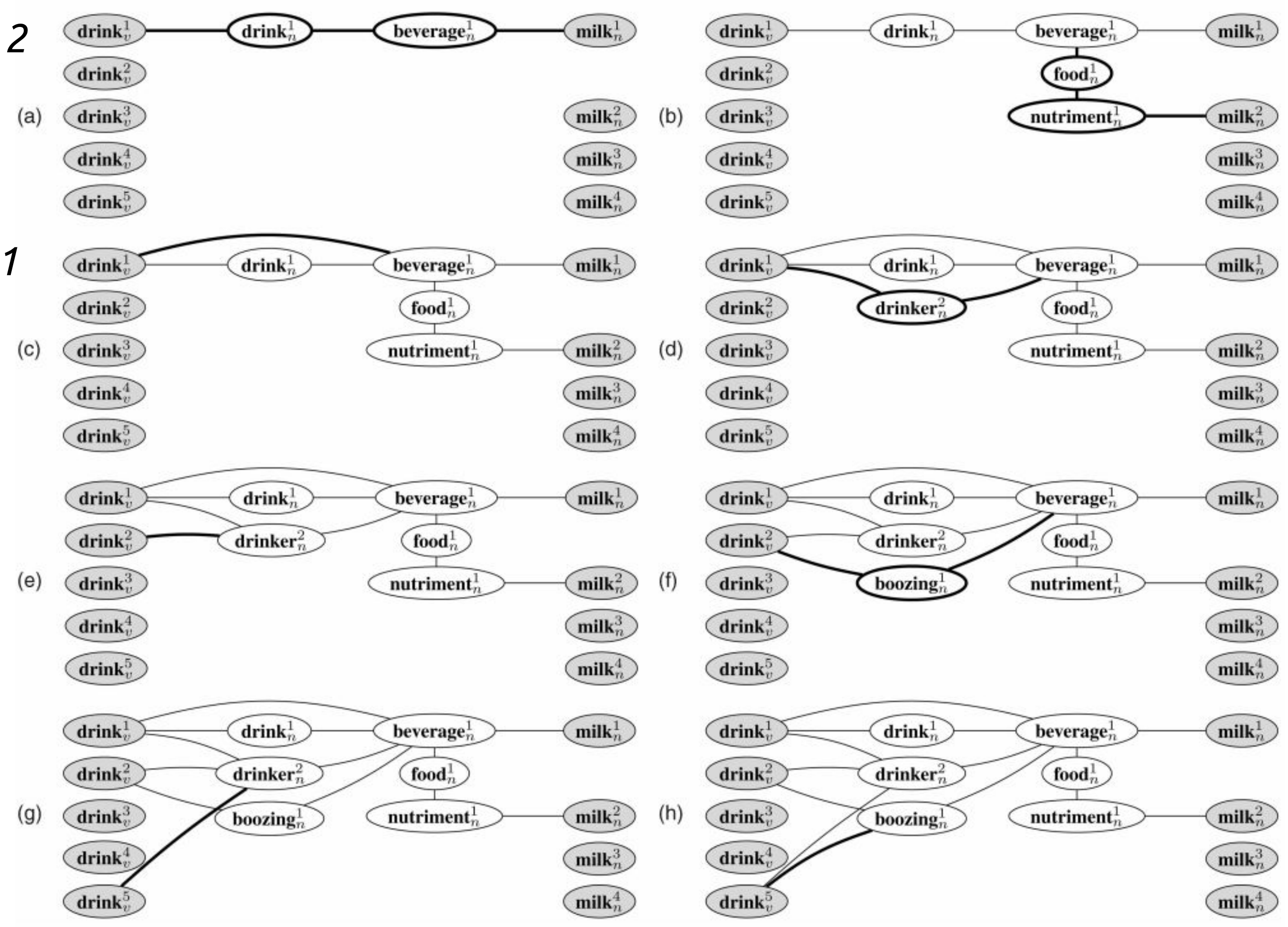
d(u,v) = 2

d(u,v) = 4

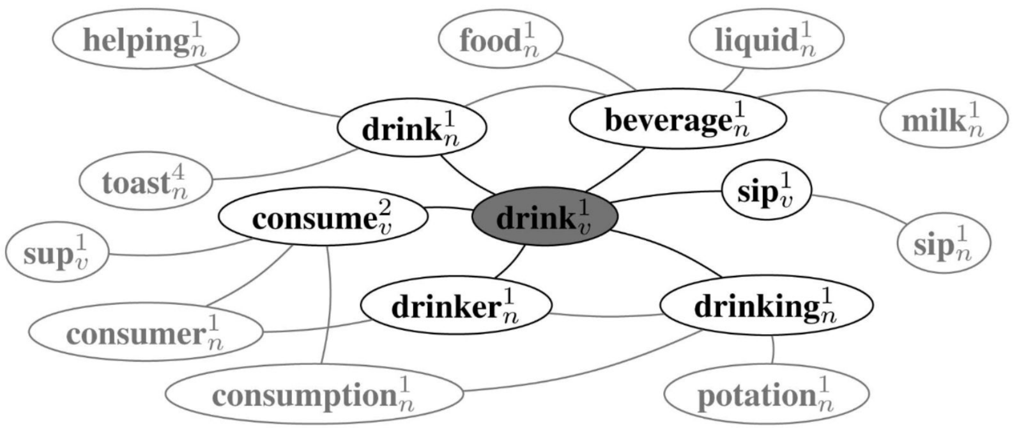
d(u,v) = 1

d(u,v) = 2

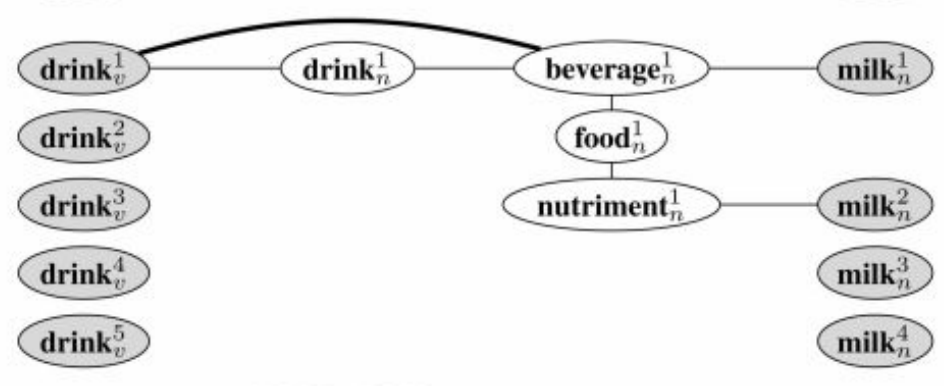
d(u,v) = 4



Shortest Path



$d(u,v) = 1$



*"She **[drank#1]** some **[milk#1]**."*

Approaches to WSD

Unsupervised

- Knowledge-based
- Graph-based

Supervised

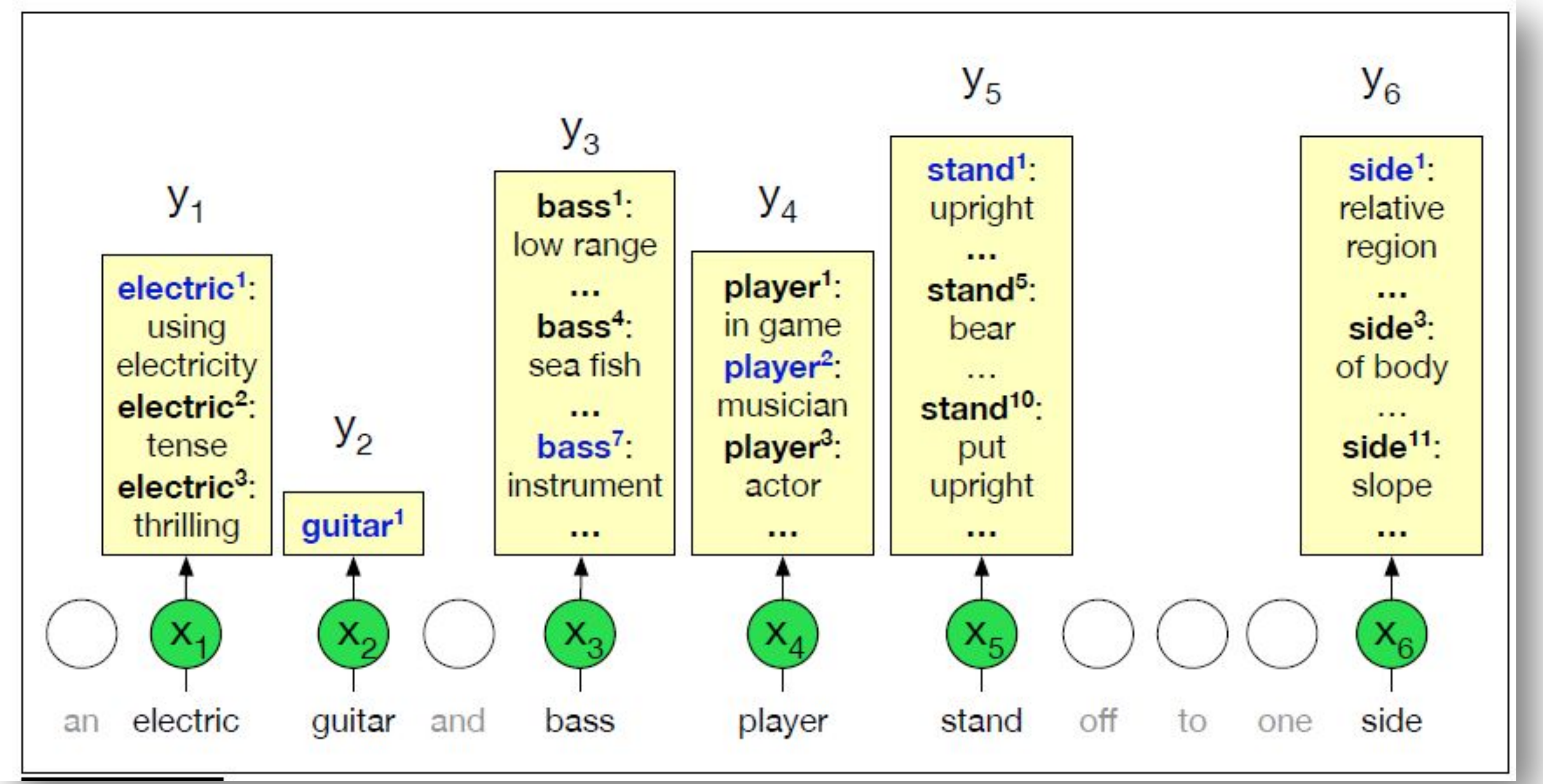
- Word Sense Annotated Corpora

Semi-supervised

- Distant supervision

Image source: https://www.coursera.org/learn/machine-learning

Supervised WSD - Sense Annotated Corpora



all-words WSD task, mapping from input words (x) to WordNet senses (y) (Jurafsky & Martin, 3rd ed., p.364)

WSD - Vector Space Model

WSD target:

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

Labeled (word sense annotated) corpus data:

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 labeled data):

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

WSD - Vector Space Model

WSD target:

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

Labeled (word sense annotated) corpus data:

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

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

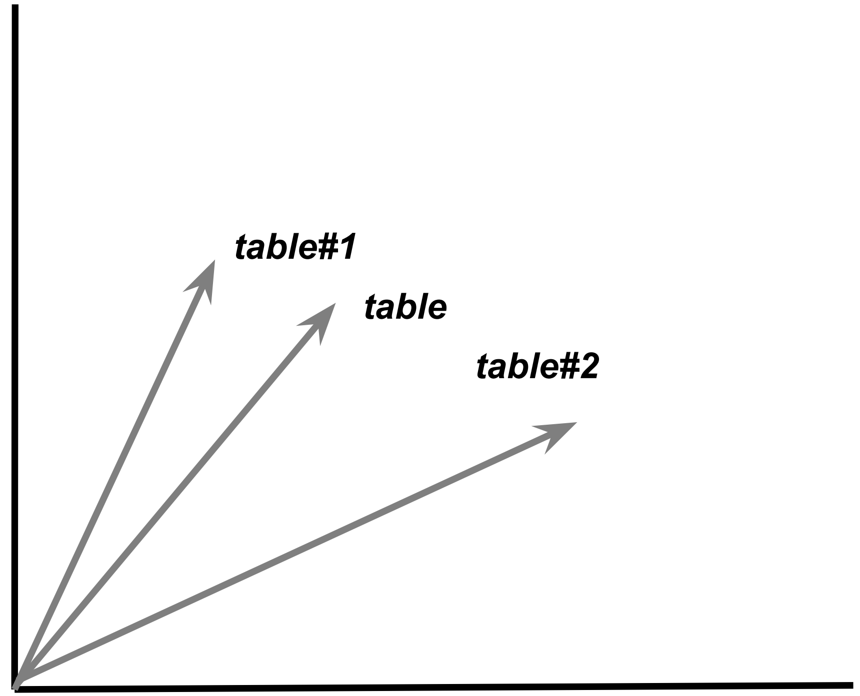
Vocabulary (nouns in WSD target and labeled data):

[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 Model



Supervised WSD - Features

SVM classifier using features of surrounding words (e.g. window of 2 on each side):

*An electric [guitar and **bass** player stand] off to one side.*

part-of-speech tags

guitar, NN, and, CC, player, NN, stand, VB

collocation features

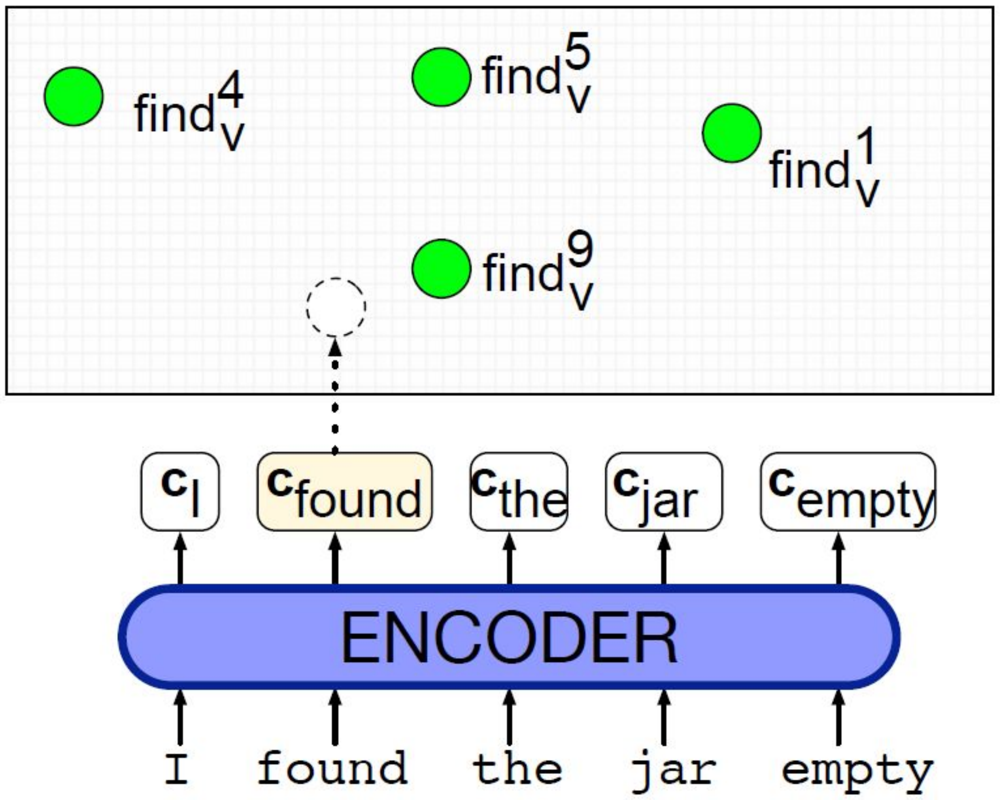
and guitar, player stand

weighted sum *g* of embeddings

g(E(guitar), E(and), E(player), E(stand))

[guitar, NN, and, CC, player, NN, stand, VB, and guitar, player stand, g(E(guitar), E(and), E(player), E(stand))]

Supervised WSD - Nearest-Neighbor Algorithm



WSD Training

- Pass each sentence in a word sense labeled dataset (e.g. SemCore: Miller et al 1993) through any contextual embedding (e.g., BERT)
- For each token of each sense of each word, average contextual representations to produce contextual sense embeddings

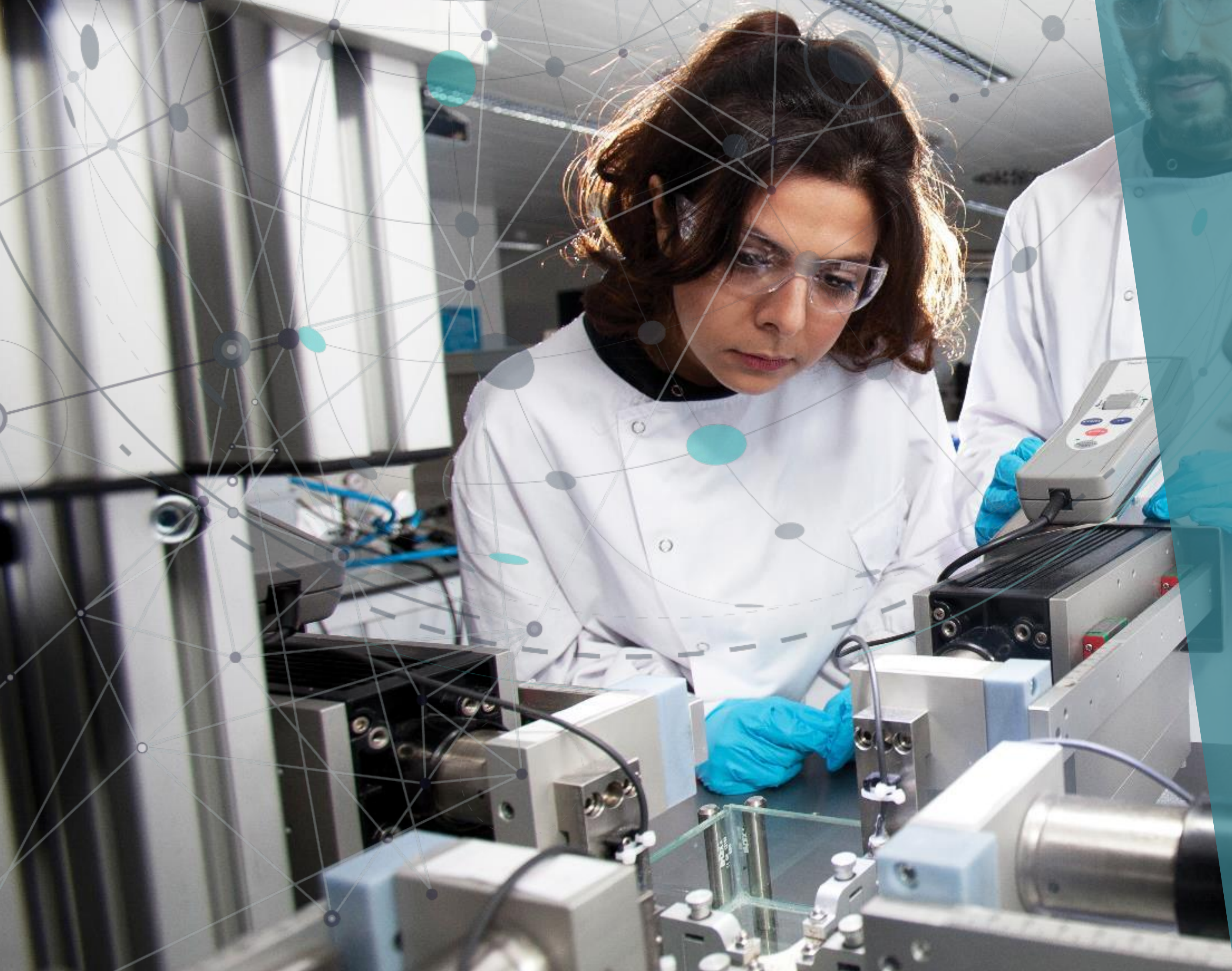
WSD Application

- Compute contextual embedding for target (ambiguous) word and select nearest neighbor contextual sense embedding from training set

Jurafsky & Martin, 3rd ed., p.365

Miller, G. A., Leacock, C., Tengi, R., & Bunker, R. T. (1993). A semantic concordance. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, 21-24, 1993*.

Ollscoil na Gaillimhe
University of Galway



Buzz Groups

Can you think of a challenge in supervised WSD?

Supervised WSD - Lack of Training Data

Only a **few sense annotated corpora exist**, even for English

No sense annotated corpora for most **other languages**

No sense annotated corpora for **domain-specific areas** (e.g. finance)

Language use changes rapidly, sense annotation may be outdated soon

Approaches to WSD

Unsupervised

- Knowledge-based
- Graph-based

Supervised

- Word Sense Annotated Corpora

Semi-supervised

- Distant supervision

Semi-Supervised WSD

Distant Supervision by use of Wikipedia as a source of sense-labeled data

We do not consider labeled data in a corpus but use **related data** at a ‘distance’ as ‘weakly labeled data’

WIKIPEDIA
The Free Encyclopedia

Using Wikipedia Categories as WordNet Senses

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

Article | Talk | Read | Edit | View hi

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

- 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"*

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

Wikipedia Text as ‘weakly labeled sense data’

*“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_(information)

⇒ Align **Wikipedia Category information** with **WordNet sense table#1**

- ★ *A **[table#1:information]** is an arrangement of data in rows and columns, or possibly in a more complex structure.*
- ★ ***[table#1:information]** appear in print media, handwritten notes, computer software, architectural ornamentation, traffic signs, and many other places.*
- ★ *In books and technical articles, **[table#1:information]** are typically presented apart from the main text in numbered and captioned floating blocks.*

Wikipedia Text as ‘weakly labeled sense data’

*“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_(furniture)

⇒ Align **Wikipedia Category furniture** with **WordNet sense table#2**

★ *A **[table#2: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 **[table#2:furniture]**, such as drafting **[table#2:furniture]**, used for doing architectural drawings, and sewing **[table#2:furniture]**.*

WSD - SemEval Tasks (Ideas for MSc Projects)

SemEval-2021 Task 2: Multilingual and Cross-lingual Word-in-Context Disambiguation: https://competitions.codalab.org/competitions/27054

SemEval 2020 Task 1: Unsupervised Lexical Semantic Change Detection: https://competitions.codalab.org/competitions/20948

SemEval-2015 Task 13: Multilingual All-Words Sense Disambiguation and Entity Linking: https://alt.qcri.org/semeval2015/task13/

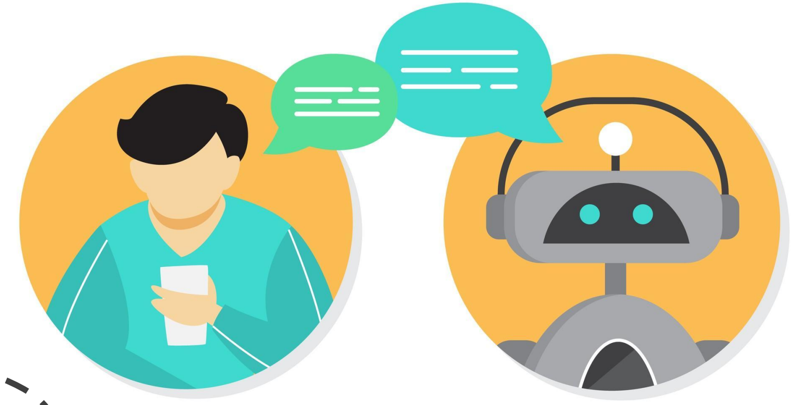
Ollscoil na Gaillimhe
University of Galway

Semantic Role Labelling

Remember: SRL in Applications

"Google acquired Fitbit in January 2021"

"Which company did Google buy last year?"



"Google bought Fitbit"

Question Answering and Chatbots

Images source https://towardsdatascience.com/automatic-question-answering-ac7593432842

What is Semantic Role Labelling (SRL)?

Process of understanding **who** did **what** to **whom**, **with what**, **where** ...

NLP task of **assigning labels to entities** (expressed by words or phrases) indicating their **semantic role in an event**

For example: several **entities** with different **semantic roles** in this *hitting* **event:**

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

Where: AGENT (AGT), EXPERIENCER (EXP), INSTRUMENT (INS)

Note that semantic roles remain the same across different sentences:

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

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

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

Semantic Roles - Overview

Thematic Role	Definition
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

Figure 19.1 Some commonly used thematic roles with their definitions.

Thematic Role	Example
AGENT	*The waiter* spilled the soup.
EXPERIENCER	*John* has a headache.
FORCE	*The wind* blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke *the ice*...
RESULT	The city built a *regulation-size baseball diamond*...
CONTENT	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	He poached catfish, stunning them *with a shocking device*...
BENEFICIARY	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	I flew in *from Boston*.
GOAL	I drove *to Portland*.

Figure 19.2 Some prototypical examples of various thematic roles.

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

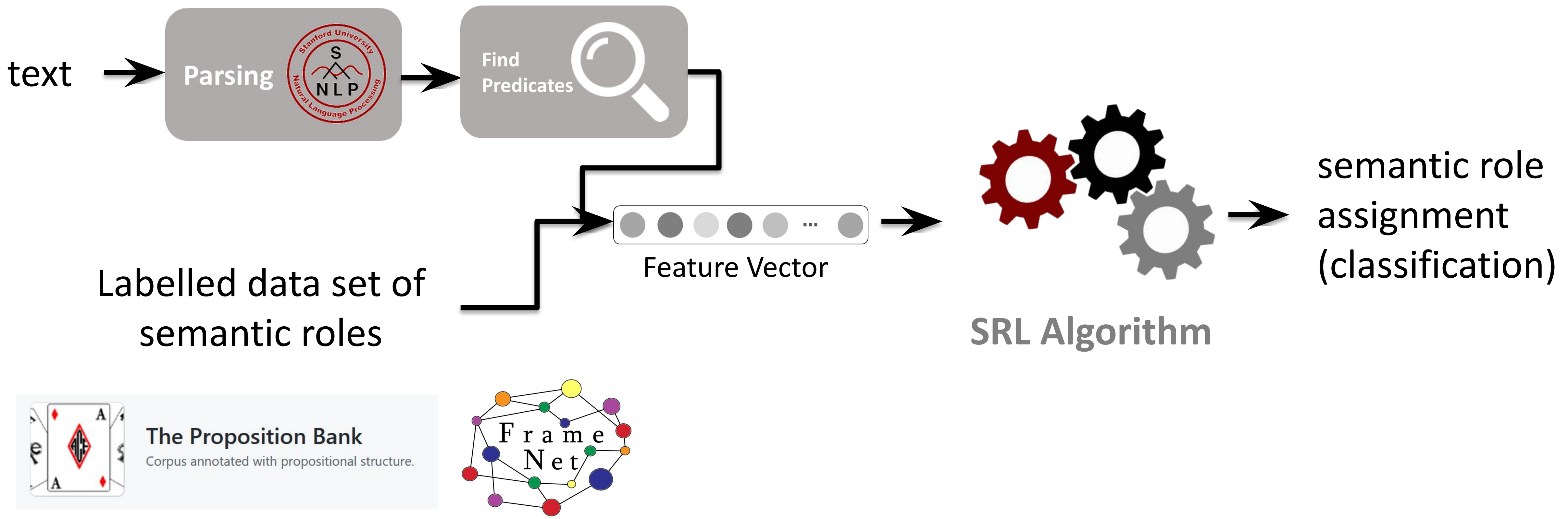
Jurafsky & Martin p374/375

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 available**

SRL Approach



Remember: FrameNet

Based on **Frame Semantics** (Fillmore 1968)

"frame [is] a description of a type of event … and the participants in it"

Frames express Semantic Roles - for example:

- ***cooking*** typically relates to an **Apply_Heat event**
- involving a person doing the cooking (**Cook**)
- the food that is to be cooked (**Food**)
- something to hold the food while cooking (**Container**)
- and a source of heat (**Heating_instrument**)

Labeled Data:

*"[John (**Cook**)] [made **Apply_Heat event]** [a great cake (**Food**)] [in the oven (**Heating_instrument**)]."*

Fillmore, C. J. (1968). The case for case. In Bach, E.W. and Harms, R. T. (Eds.), Universals in Linguistic Theory, p1–88

PropBank

The Proposition Bank
Corpus annotated with propositional structure.

The Proposition Bank (or ‘PropBank’), is a corpus annotated with semantic roles (Palmer et al 2005)

Each sense of each verb has a specific set of roles, which are given only numbers rather than names: Arg0, Arg1, Arg2, and so on.

Labeled Data:

*“[AIB **Arg0**] [raised **increase.01**] [interest **Arg1]** [incrementally by 2% **Arg2**].”*

- **Arg0**: causer of increase
- **Arg1**: thing increasing
- **Arg2**: amount increased by

Jurafsky & Martin p377/378

Palmer M, Kingsbury P, Gildea D (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles". *Computational Linguistics*. **31** (1): 71–106.

SRL Features - Syntax

Use constituency parsing features

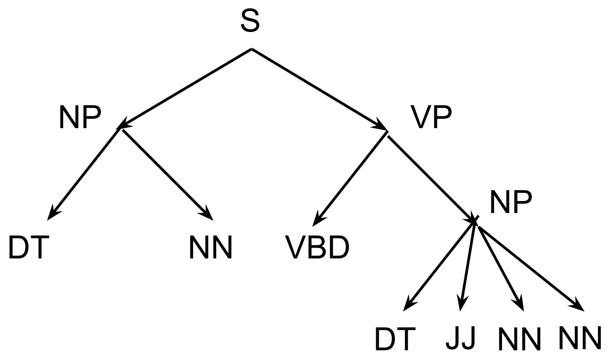
The chancellor announced a new policy measure.

(S

(NP (DT *The*) (NN *chancellor*))

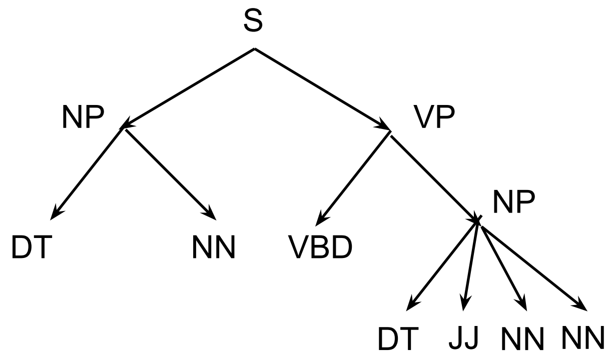
(VP (VBD *announced*)

(NP (DT *a*) (JJ *new*) (NN *policy*) (NN *measure*))))



Gildea, D. and Jurafsky, D. (2002). Automatic labeling of semantic roles. Computational Linguistics, 28(3), 245–288.

SRL Features



Labeled Data

[AGENT The chancellor] announced new policy measures.

Feature Extraction

Predicate:	***announce***
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:	***chancellor***
Sub-categorization for parent of predicate:	**VP→VBD,NP**

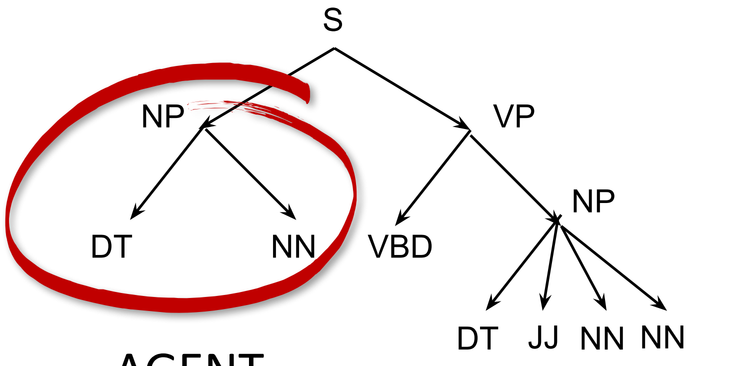
Feature Vector

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

SRL Algorithm

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

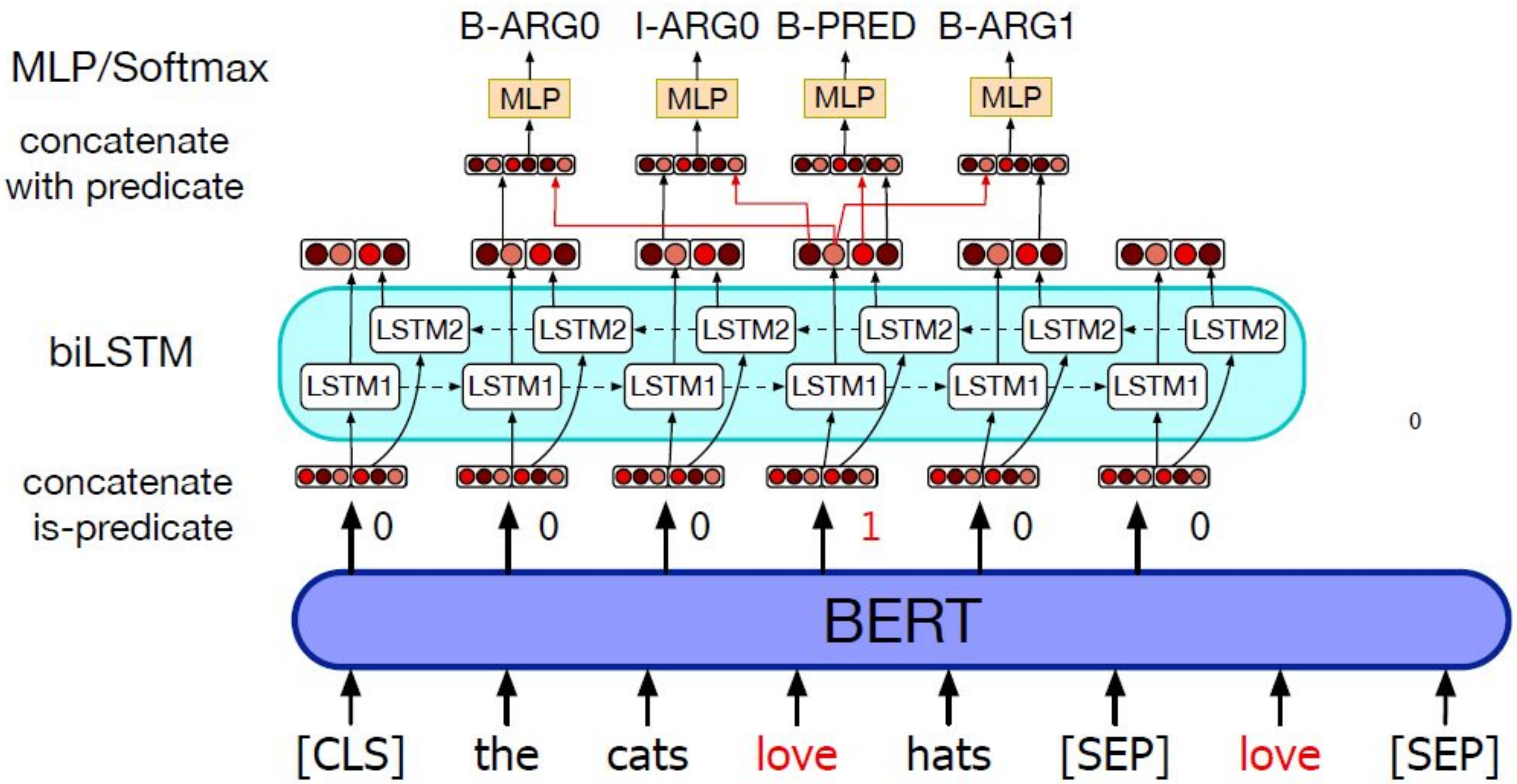
  parse ← PARSE(*words*)
  for each *predicate* **in** *parse* **do**
    for each *node* **in** *parse* **do**
      featurevector ← EXTRACTFEATURES(*node*, *predicate*, *parse*)
      CLASSIFYNODE(*node*, *featurevector*, *parse*)



AGENT

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

Neural SRL Algorithm



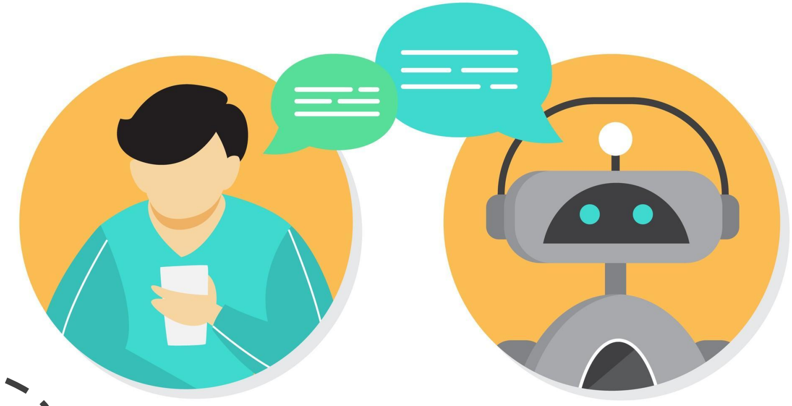
0

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Coreference Resolution

Remember: CR in Applications

"Last year, Google completed its acquisition of Fitbit."



"Which company did Google buy last year?"

"Google bought Fitbit"

Question Answering and Chatbots

Image source https://towardsdatascience.com/automatic-question-answering-ac7593432842

CR in Applications - Other Example

"How late is the 404 bus to Dangan?"

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

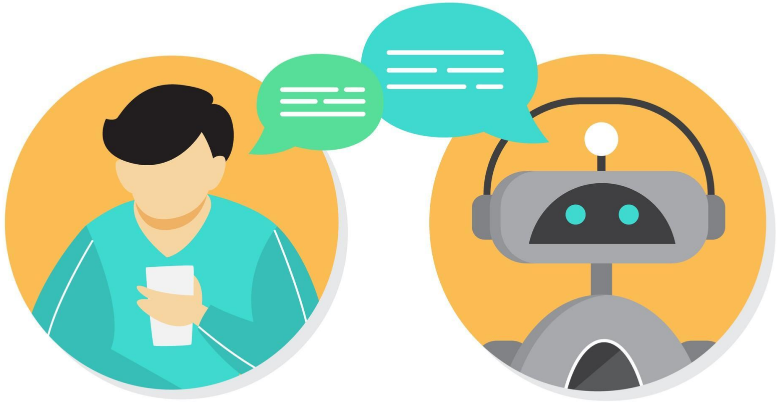
"Does it stop at the university?"

it : bus

"The bus to Dangan stops at the university"

"When will it stop there?"

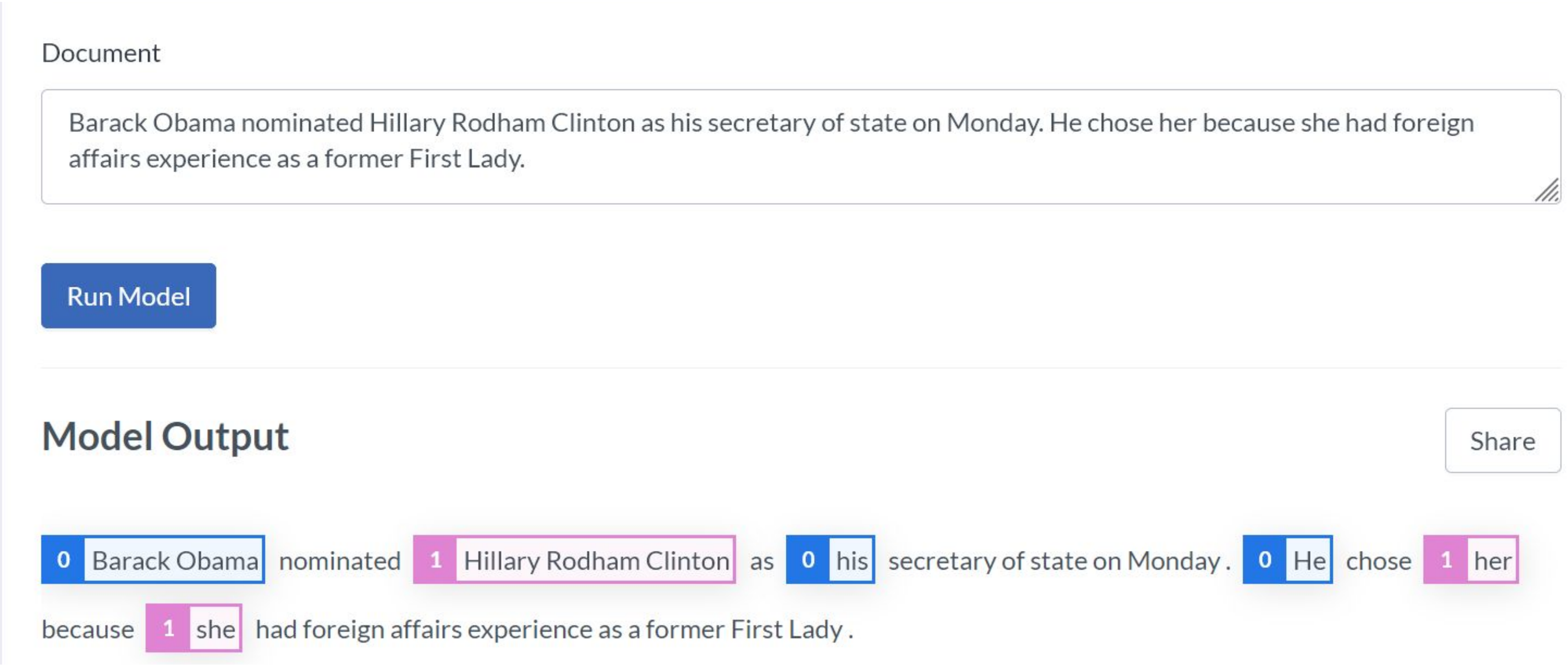
it : bus *there : the university*



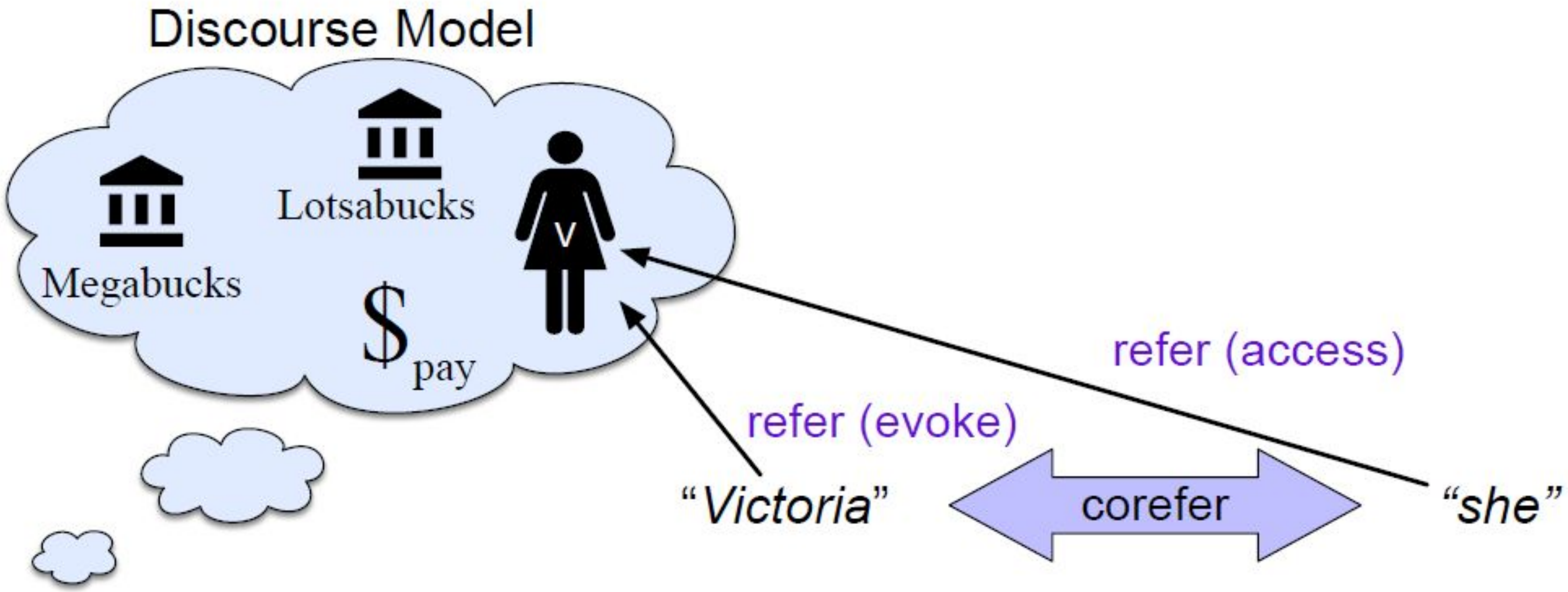
Question Answering and Chatbots

What is Coreference Resolution?

Task of identifying words and/or phrases **referring to the same entity**



Remember: Discourse Semantics



Subtasks: Mention Detection & 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.”

Step 1: 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**.”*

Mention Detection

Step 2: Identify all **mentions** that refer to the **same** real world entity

*“**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.”*

*“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

Example and slide adapted from: https://courses.cs.washington.edu/courses/csep517/20wi/slides/csep517wi20-Coreference.pdf

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"*

NLP Tools to Use

Part-of-speech tagger

Named Entity Recognizer (NER)

Constituency (Phrase) parser

Slide adapted from: https://www.cs.princeton.edu/courses/archive/fall19/cos484/lectures/lec21.pdf

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 ≠ John)*

Syntactic agreement

***Jack** gave **Mary** a gift. **She** was excited.* *She = Mary (She ≠ Jack)*

Other syntactic constraints

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

Semantic compatibility

*… the mining **conglomerate** … the **company** …* *company = conglomerate*

Grammatical Role - prefer entities in the subject position

***John** went to a movie with **Jack** as **he** was not busy.* *He = John (He ≠ Jack)*

Mention Pair Architecture

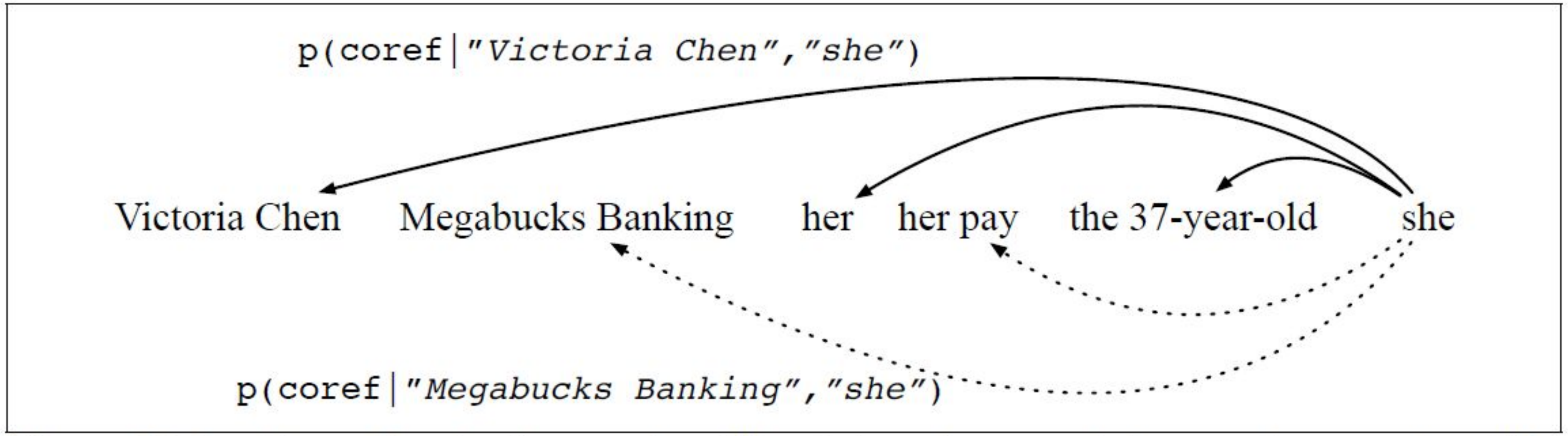


Figure 21.2 For each pair of a mention (like *she*), and a potential antecedent mention (like *Victoria Chen* or *her*), the mention-pair classifier assigns a probability of a coreference link.

Learning Outcomes

After completing this topic, you should be able to:

- Explain semantic analysis in NLP
- Discuss various approaches to Word Sense Disambiguation
- Understand the core idea of and features used for Semantic Role Labeling & Coreference Resolution

External Sources:

Chapters 18, 19 and 21 in Jurafsky and Martin, SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 3rd edition: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Lab of this week

Practical application on WSD