FNLP - Week 8: Semantic Role Labelling and Lexical Semantics

Antonio León Villares

March 2022

Contents

1 Semantic Role Labelling

1.1 Semantic Roles

- **What are semantic roles?**
  - also known as **thematic roles**
  - capture **semantic commonality** between words
    - for example:

      "Sasha broke the window"

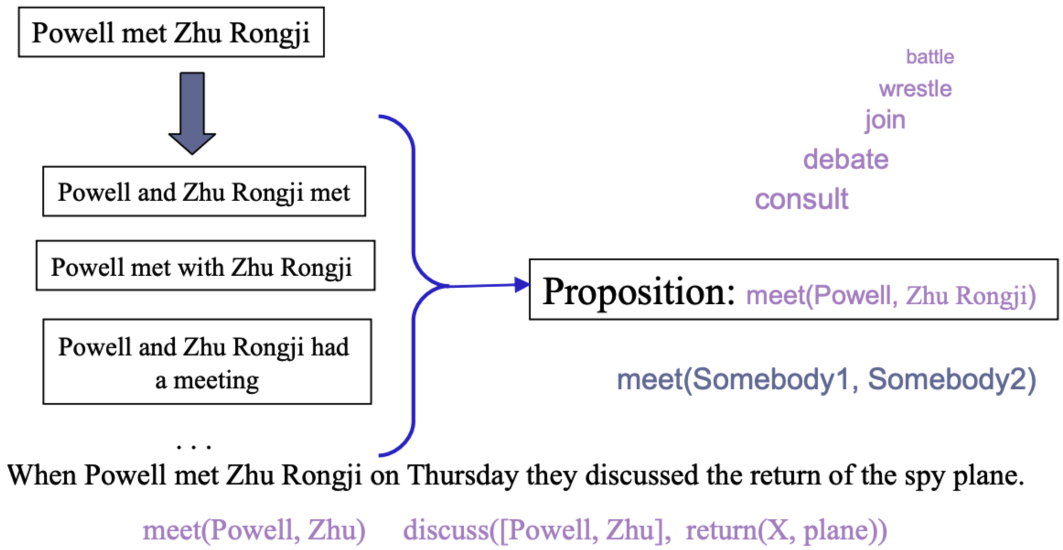
      "Pat opened the door"
    - Sasha is acting as a **breaker**; Pat acts as an **opener**
    - different actions, but similar **theme**: sentient being performing an action
- **What are some examples of semantic roles?**
  - no universal set of semantic roles; however, we can use:

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 1: In the example above, "Pat" and "Sasha" act as **agents**. The "door" and "window" (acted upon by the agents) are **themes**.

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*.

- **What is the purpose of semantic role labelling?**
  - assign **semantic roles** to sentence spans
  - informally: allows us to understand *who did what to whom when, where and how*



- **Why is semantic role labelling useful?**
  - **syntax** and **semantics** not powerful enough to understand the meaning of sentences
  - **semantic role labelling** provides a (shallow) **meaning representation**, which is **general**
    - for example:

      (1) John broke the window

      (2) The window broke

      (3) The window was broken by John

      (4) John smashed the window

      (5) The window was destroyed by John

      (6) John tore down the window

    - as humans, we understand all these sentences convey the same meaning: a window was broken by a person called John
    - however, simple **syntactic** and **lexical** representations could understand these as sentences representing different phenomena
    - with semantic role labelling, we can see that, for instance, verbs like "smashed" and "broke" fulfill similar roles (acting on a **theme** like "the window)
    - this process in which a **verb** realises an action on different subjects is known as **diathesis alternation**
  - useful:
    - provide **inferences** not possible even with parse trees (i.e understanding what a question is asking [*"Was Minecraft acquired by Microsoft?"*], and what type of information can be used to answer a question [*"In September 2014, Microsoft acquired Minecraft."*])
    - act as **intermediate language** in **machine translation**

1.2 Issues with Semantic Roles

1. No **universal** set of roles
2. Items with **same role** don't always behave in the same way, so might need to **fragment** them

$$\textit{“Sandy opened the door with a key”} \implies \textit{“The key opened the door”}$$

$$\textit{“Sandy ate the salad with a fork”} \not\Longrightarrow \textit{“The fork ate the salad”}$$

3. Hard to **formally** define **semantic roles**

    - i.e an **agent** could be defined as:

      animate, volitional, sentient, causal

    - however, certain noun phrases don't satisfy **all**
    - partial solution: define **generalised semantic roles** (i.e **proto-agent**, **proto-patient**)
    - use **heuristics**: exhibiting more agent-like properties (i.e sentiency), implies higher likelihood of being a **proto-agent**
    - alternatively, define very fine-grained roles: **PropBank** and **FrameNet**

1.3 The Proposition Bank: PropBank

- **What is PropBank?**
  - sentences annotated with **semantic roles**, with respect to **verb senses**
  - instead of a general role for each verb, assign each verb with its specific roles
  - consistent with Penn TreeBank
  - **NomBank** does same thing, but annotating noun predicates
- **How is PropBank stuctured?**
  - each verb annotated with a **set of roles**

    Frameset **break.01** "break, cause to not be whole":
    Arg0: breaker
    Arg1: thing broken
    Arg2: instrument
    Arg3: pieces

    Figure 2: Such entries in PropBank are called **frame file**

- **What are the Args in the frame files?**
  - *Argn* refers to a specific role assigned to the verb
  - no convention with roles, so used numbered arguments
  - not much structure with regards to roles (either **number of roles**, or specific **meaning** of *Argn*)
    - except for *Arg*0 (**proto-agent**) and *Arg*1 (**proto-patient**)

agree.01
Arg0: Agreer
Arg1: Proposition
Arg2: Other entity agreeing

Ex1: [$_{\text{Arg0}}$ The group] *agreed* [$_{\text{Arg1}}$ it wouldn't make an offer].
Ex2: [$_{\text{ArgM-TMP}}$ Usually] [$_{\text{Arg0}}$ John] *agrees* [$_{\text{Arg2}}$ with Mary] [$_{\text{Arg1}}$ on everything].

fall.01
Arg1: Logical subject, patient, thing falling
Arg2: Extent, amount fallen
Arg3: start point
Arg4: end point, end state of arg1
Ex1: [$_{\text{Arg1}}$ Sales] *fell* [$_{\text{Arg4}}$ to \$25 million] [$_{\text{Arg3}}$ from \$27 million].
Ex2: [$_{\text{Arg1}}$ The average junk bond] *fell* [$_{\text{Arg2}}$ by 4.2%].

Figure 3: We can see that "agree" has a proto-agent (an **agreer**) and a proto-patient (the thing agreed on, a **proposition**.
However, "fall" doesn't have a proto-agent ($Arg0$) (since "fall" alone affects the proto-patient).

- **What other arguments does PropBank consider?**
  - has **non-numbered** arguments $ArgMs$
  - represent **modifications/adjunct meanings**
  - used, for example, to have better understanding of **temporal location**

TMP	when?	yesterday evening, now
LOC	where?	at the museum, in San Francisco
DIR	where to/from?	down, to Bangkok
MNR	how?	clearly, with much enthusiasm
PRP/CAU	why?	because ... , in response to the ruling
REC		themselves, each other
ADV	miscellaneous	
PRD	secondary predication	...ate the meat raw

- **How can PropBank be useful?**

[$_{\text{Arg0}}$ Big Fruit Co.] increased [$_{\text{Arg1}}$ the price of bananas].
[$_{\text{Arg1}}$ The price of bananas] was increased again [$_{\text{Arg0}}$ by Big Fruit Co.]
[$_{\text{Arg1}}$ The price of bananas] increased [$_{\text{Arg2}}$ 5%].

Figure 4: Despite having 3 different sentences, PropBank gauges the commonality between them by using the verb "increased".

- **How is PropBank applied on parses?**

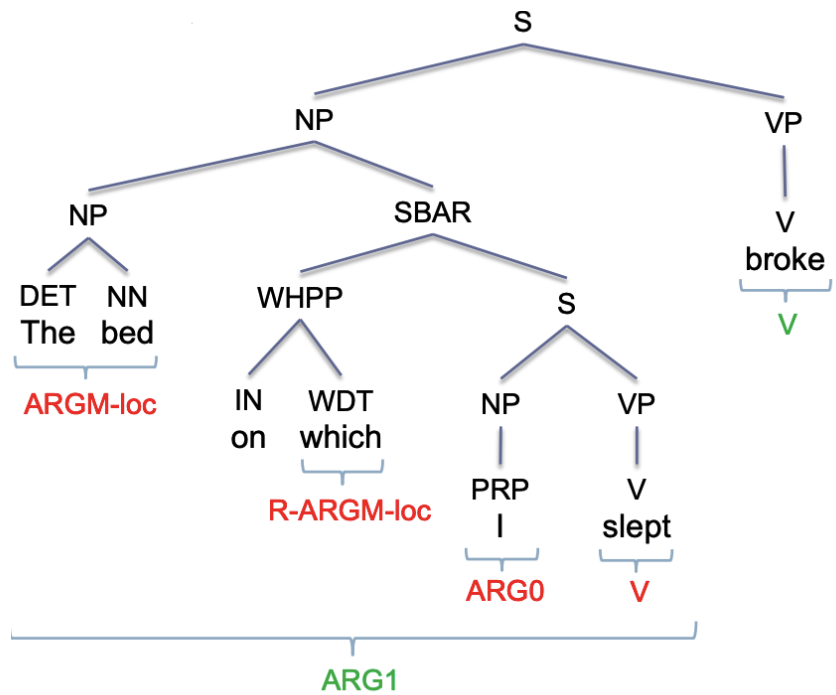
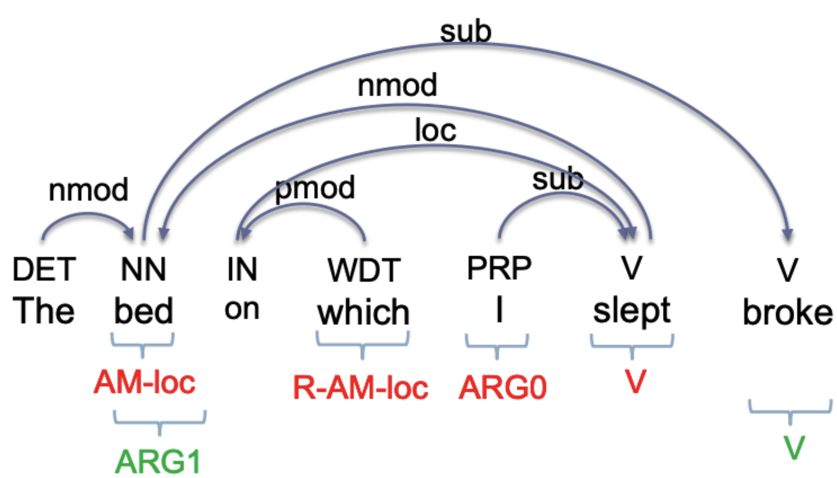


Figure 5: We annotate on top of the constituency syntax labels-



- **What are the issues with PropBank?**
  1. **Incomplete Role Consistency** (for example, synonyms don't have the same roles assigned)
  2. Only focuses on **verbs**
     - issue: things expressible via verbs can be expressed via nouns
     - for example, consider applying SRL to answer a question on meetings
     - can be expressed via verbs ("James and Arvid met yesterday"), but also nouns ("James and Arvid had a meeting yesterday")

- PropBank sees these as fundamentally different, since one involves **meeting**, but the other involves **having**

3. Overly tied to **syntax**
   - expanding on the above:

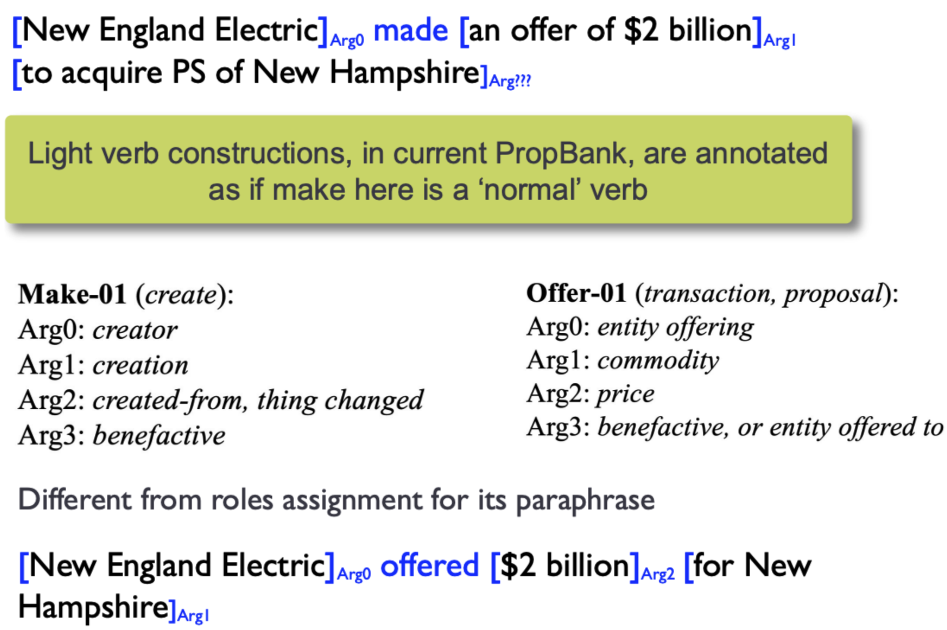


Figure 6: PropBank can't see that the fact that there was an offer is what matters, but it ignores this in the first sentence, since the verb "made" is used.

1.4 FrameNet

- **What are semantic frames?**
  - notion that our knowledge of a **concept** is based of our understanding of a set of **background concepts**
  - for example in:

    reservation, flight, travel, buy, price, cost, fare, rates, meal, plane

    we understand that the words are related, but we don't understand **how** - in this case, information concerning air-travel
  - a **frame** is the **background knowledge** which **unites** our understanding of a group of words:

> *"A* ***semantic frame*** *is a conceptual structure describing a situation, object, or event along with associated properties and participants"*

  * for example, consider the sentence:

    Lansky left Australia to study the piano at the Royal College of Music.

∗ **frames** constitute the underlying “concepts” which dominate the meaning of this sentence: in this case, “education” and “departing”

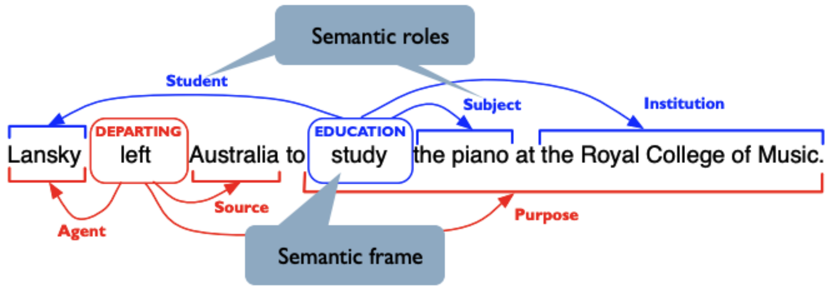


Figure 7: We can annotate it using semantic frames, and the semantic roles.

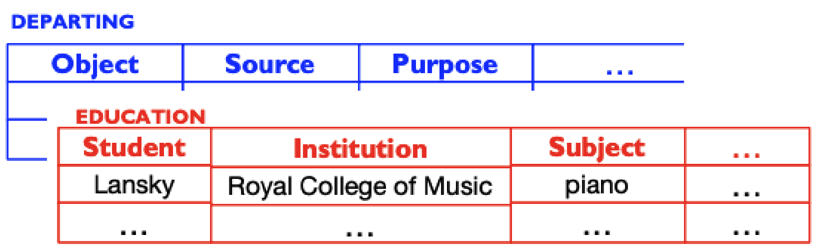


Figure 8: Semantic frames are typically organised as a **relational database**.

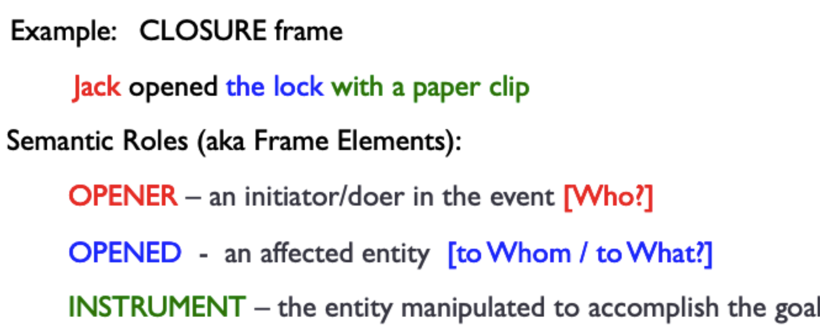


Figure 9: Semantic roles are the ones which define the frames.

- **What is FrameNet?**
  - a **lexicographic database**, containing:
    - ∗ **frames**
      - · also **frame elements** (containing semantic role of frame)
      - · each word has a frame associated, alongside some representative examples of the frame in use

[ITEM Oil] *rose* [ATTRIBUTE in price] [DIFFERENCE by 2%].

[ITEM It] has *increased* [FINAL_STATE to having them 1 day a month].

[ITEM Microsoft shares] *fell* [FINAL_VALUE to 7 5/8].

[ITEM Colon cancer incidence] *fell* [DIFFERENCE by 50%] [GROUP among men].

a steady *increase* [INITIAL_VALUE from 9.5] [FINAL_VALUE to 14.3] [ITEM in dividends]

a [DIFFERENCE 5%] [ITEM dividend] *increase*...

Figure 10: Examples used in the frame `change_position_on_scale`.

- **frame elements** can be **core** (frame-specific) and **non-core** (similar to $ArgM$ in PropBank)

Core Roles	
ATTRIBUTE	The ATTRIBUTE is a scalar property that the ITEM possesses.
DIFFERENCE	The distance by which an ITEM changes its position on the scale.
FINAL_STATE	A description that presents the ITEM's state after the change in the ATTRIBUTE's value as an independent predication.
FINAL_VALUE	The position on the scale where the ITEM ends up.
INITIAL_STATE	A description that presents the ITEM's state before the change in the ATTRIBUTE's value as an independent predication.
INITIAL_VALUE	The initial position on the scale from which the ITEM moves away.
ITEM	The entity that has a position on the scale.
VALUE_RANGE	A portion of the scale, typically identified by its end points, along which the values of the ATTRIBUTE fluctuate.
Some Non-Core Roles	
DURATION	The length of time over which the change takes place.
SPEED	The rate of change of the VALUE.
GROUP	The GROUP in which an ITEM changes the value of an ATTRIBUTE in a specified way.

Figure 11: Core and non-core frame elements for the frame `change_position_on_scale`.

- **frame** definitions

> *The frame **`change_position_on_scale`** is defined by:*
> *"This frame consists of words that indicate the change of an **Item**'s position on a scale (the **Attribute**) from a starting point (**Initial** value) to an end point (**Final** value)"*

- **frame** relations
  - frames can **inherit** properties, or represent **causation**
  - for example, `change_position_on_scale` can be associated with `Cause_change_position_on_scale`

- **Why is FrameNet useful?**
  - doesn't only focus on **verbs**, so better for more complex inferences

[$_{\text{Arg1}}$ The price of bananas] increased [$_{\text{Arg2}}$ 5%].
[$_{\text{Arg1}}$ The price of bananas] rose [$_{\text{Arg2}}$ 5%].
There has been a [$_{\text{Arg2}}$ 5%] rise [$_{\text{Arg1}}$ in the price of bananas].

Figure 12: Unlike with PropBank, FrameNet can be used to understand that these 3 sentences are similar: the "price of bananas" has gone up; and "5%" is the amount by which it has gone up.

- **What are some issues of FrameNet?**
  1. Very **Small** (only 5k annotated sentences; PropBank has 40k)
  2. Frames have **Different Granularity** (for example, 10 frames related to hair, 2 frames related to education)
  3. Unrelated to **Syntax** (arguments in FrameNet not necessarily syntactic constituents, so harder to build ML models)
- **What is SemLink?**
  - provides **alignment** between resources:
    * link *Argn* across different verbs in PropBank
    * link frames in FrameNet
    * link frames in FrameNet and roles in PropBank

1.5 Supervised Semantic Role Labelling

- **How is SRL carried out nowadays?**
  - using **supervised learning**
  - develop a **parser**, trained on annotated text
  - requires a lot of data
  - nowadays, very strong, feature-based parsers can still make "simple" mistakes:

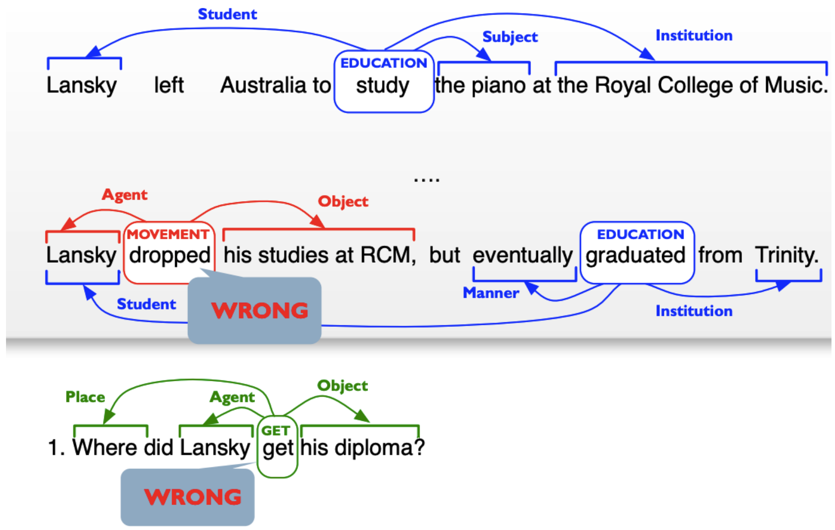


Figure 13: It misses the fact that "dropped" is meant in an academic context. It also thinks of "get" as "grabbing an object", as opposed to "obtaining an academic accreditation". Hence, it doesn't "understand" the question, nor can it associate it to the second example sentence.

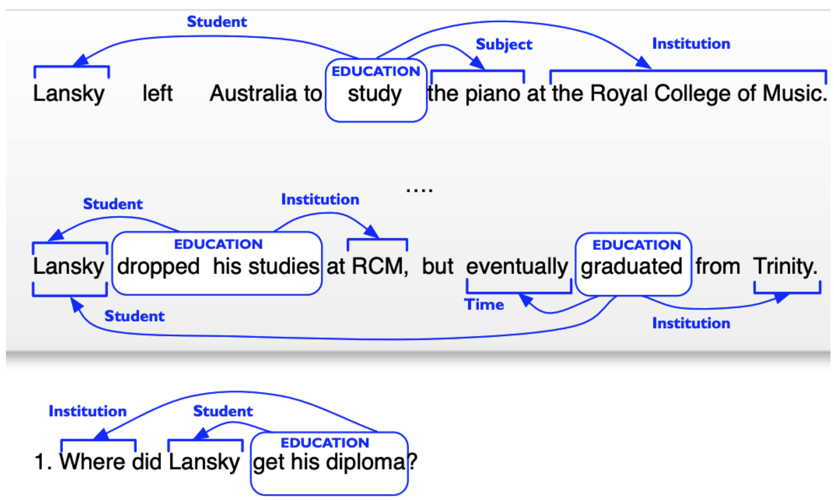


Figure 14: The correct annotations, using FrameNet notation.

- **How are SRL models trained?**
  - can be done **end-to-end**, but typically follow 3 steps:

1. **Multiclass Classifcation**
   - identifying **predicates** and their **word sense** or **frame**
2. **Sequence Labelling**
   - define the argument spans taken by the predicates
3. **Multiclass Classification**
   - classify the argument spans according tot he roles

- **How can supervised SRL be enhanced?**
  - use **features** dependent on **syntactic structure**
  - apply **constraints** on valid labelling (can be applied with DP):
    * a role appears only **once** in sentence
    * role consistent with predicate (i.e using the definition of a frame)
    * constraints on syntax
- **How can we deal with little/no annotations?**
  - problem, for example, when using SRL with different languages
  - solution:
    * transfer resources/models from other labelled languages
    * semi-supervised/unsupervised learning (i.e HMMs)

2 Lexical Semantics: Defining the Meaning of Words

> *"How many legs does a dog have if you call its tail a leg? Four. Calling a tail a leg doesn't make it one."*
> *Attributed to Abraham Lincoln*

2.1 Motivating Lexical Semantics

- **What is a lexeme?**
  - a pairing between a word form (**orthographic** or **phonological**) and its **meaning**
- **What is a lexicon?**
  - a finite set of **lexemes**
- **What is a lemma?**
  - **grammatical** form which represents a **lexeme**
  - usually a **base form** (i.e "carpet" is the lemma of "carpets", and for verbs we typically use the infinitive))
- **Why are lemmas part of speech specific?**
  - the same word can have different parts of speech (i.e "table" can be a noun or a verb)
  - hence, we need a lemma for each possible form
- **What are lexical semantics?**
  - the study of the meaning of **individual** words
  - key to deriving **sentential semantics**
  - currently, lexical semantics built by using symbols like $dog(x)$, but this doesn't convey the meaning of "dog" (i.e a four legged mammal, domesticated, etc...)
  - **words** act as an **interface** between our world and **meaning**

2.2 The Difficulty in Endowing Meaning To Words: Word Relations

To motivate this, we consider the process of building a question answering machine. Such a machine has access to a knowledge base, and access to English text. This allows us to see how different word sense can be related.

2.2.1 Word Senses

- a specific meaning of a word
- for example, "bank" has 2 word sense:
  1. financial institution
  2. terrain next to a river
- a QA machine needs to be able to disambiguate:

> ***Q**: What plants are native to Scotland?*
> ***Corpus**: A new chemical plant was opened in Scotland*

- we understand from context that "plant" in Q refers to vegetation, whilst "plant" in the corpus refers to "a factory or workshop for the manufacture of a particular product".
- **word sense** can be defined by using the **number of possible translations** of words
  - for example, "interest" can be translated in 3 different ways

2.2.2 Synonyms

- 2 different words with identical/nearly identical word senses
- for example:
  1. "couch" and "sofa"
  2. "car" and "automobile"
- a QA machine needs to be able to match them:

> ***Q**: Where did David Cameron go on vacation?*
> ***Corpus**: David Cameron spent his holiday in Cornwall.*

- a QA system should understand that "vacation" and "holiday" are synonyms, so a suitable answer would be "Cornwall"

2.2.3 Antonyms

- 2 words with opposite word senses/meanings
- for example:
  - "long" and "short"
  - "big" and "little"
- antonyms typically describe senses in binary opposition/opposite ends of a scale (like "fast" vs "slow"), or which are **reversive** (like "rise" vs "fall" or "up" vs "down")

2.2.4 Hyponyms

- words whose sense is a **subset** of the sense of another word
- for example:
  1. "car" is a hyponym of "vehicle"
  2. "dog" is a hyponym of "mammal", which is a hyponym of "animal"
- a QA machine should have an **ontology** - a way of understanding $A \in B$ relationships:

> ***Q***: *Which animals loves to swim?*
> ***Corpus***: *Polar bears love to swim in the freezing waters of the Arctic.*

- a QA system should see that "Polar bears" are animals, and they like to swim, so they are a suitable response

2.2.5 Hypernyms

- words whose sense is a **superset** of the sense of another word
- for example:
  1. "vehicle" is a hypernym of "car"
  2. "animal" is a hypernym of "mammal", which is a hypernym of "dog"
- also called **superordinates**

2.2.6 Homonyms

- different words written in the same way, with completely different senses
- for example:
  - "bank"
  - "pen" (as a writing device, and as a animal enclosure)
  - "arm" (as a body part, and as a division of a company)

2.2.7 Polysemy

- the same word, but with the coexistence of different word senses
- for example:
  1. "bank" can refer to a financial institution, or as a place where biological samples (i.e sperm, blood) can be stored; this word is a **homonym** of "bank" (side of the river), but not **polysemous**
  2. "fixed" is polysemous, since we can say:

     "He fixed the door"

     "He fixed his hair"

2.2.8 Metonymy

- a subset of **polysemy** (also known as **regular polysemy**), by which the semantic relation between the senses can be systematic:

  1. in "The bank is on the corner of Nassau and Witherspoon.", "bank" refers to the fact that the building holding a financial institution is found at the given location
  2. similarly with "newspaper":

     "She read the newspaper calmly"

     "She sued the newspaper for defamations"

     In the first case, we refer to a written piece of text used to inform oneself of world affairs, whilst in the second case we refer to a company who writtes and prints newspapers.
  3. the "White House" refers both to the head of government of the US, and where the president of the US lives
  4. "chicken" refers both to a domesticated bird, and the meat derived from said bird for the purpose of eating

Pattern	Participating Senses	Example Sentences
Animal for fur	Mink, chinchilla, rabbit, beaver, raccoon*, alpaca*, crocodile*	The *mink* drank some water / She likes to wear *mink*
Animal/Object for personality	Chicken, sheep, pig, snake, star*, rat*, doll*	The *chicken* drank some water / He is a *chicken*
Animal for meat	Chicken, lamb, fish, shrimp, salmon*, rabbit*, lobster*	The chicken drank some water / The *chicken* is tasty
Artifact for activity	Shower, bath, sauna, baseball,	The *shower* was leaking / The *shower* was relaxing
Body part for object part	Arm, leg, hand, face, back*, head*, foot*, shoulder*, lip*,	John's *arm* was tired / The *arm* was reupholstered
Building for people	Church, factory, school, airplane,	The *church* was built 20 years ago / The *church* sang a song
Complement Coercion	Begin, start, finish, try	John *began* reading the book / John *began* the book
Container for contents	Bottle, can, pot, pan, bowl*, plate*, box*, bucket*	The *bottle* is made of steel / He drank half of the *bottle*
Word for question	Price, weight, speed	The *price* of the coffee was low / John asked the *price* of the coffee

Pattern	Participating Senses	Example Sentences
Figure for Ground	Window, door, gate, goal	The window is broken / The cat walked through the window
Grinding	Apple, chair, fly	The apple was tasty / There is apple all over the table
Instrument for action	Hammer, brush, shovel, tape, lock*, bicycle*, comb*, saw*	The hammer is heavy / She hammered the nail into the wall
Instance of an entity for kind	Tennis, soccer, cat, dog, class*, dinner*, chair*, table*	Tennis was invented in England / Tennis was fun today
Location / Place at location	Bench, land, floor, ground, box*, bottle*, jail*	The bench was made of pine / The coach benched the player
Object for placing at goal	Water, paint, salt, butter, frame*, dress*, oil*	The water is cold / He watered the plant.
Object for taking from source	Milk, dust, weed, peel, pit*, skin*, juice*	The milk tastes good / He milked the cow
Material for artifact	Tin, iron, china, glass, linen*, rubber*, nickel*, fur*	Watch out for the broken glass / He filled the glass with water
Occupation for role in action	Boss, nurse, guard, tutor	My boss is nice / He bossed me around

Pattern	Participating Senses	Example Sentences
Place for an event	Vietnam, Korea, Waterloo, Iraq	It is raining in *Vietnam* / John was shot during *Vietnam*
Place for an institution	White House, Washington, Hollywood, Pentagon, Wall Street*, Supreme Court	The *White House* is being repainted / The *White House* made an announcement
Plant for food or material	Corn, broccoli, coffee, cotton, lettuce*, eggs*, oak*, pine*	The large field of *corn* / The *corn* is delicious
Portioning	Water, beer, jam	She drank some *water* / She bought three *waters*
Publisher for product	Newspaper, magazine, encyclopedia, Wall Street Journal*, New York Times*,	The *newspaper* is badly printed / The *newspaper* fired three employees
Artist for product	Writer, artist, composer, Shakespeare, Dickens*, Mozart*, Picasso*	The *writer* drank a lot of wine / The *writer* is hard to understand
Object for contents	Book, CD, DVD, TV*, magazine*, newspaper*	The heavy, leather- bound *book* / The *book* is funny.
Visual Metaphor	Beam, belt, column, stick, bug*, leaf*	Most of the weight rests on the *beam* / There was a *beam* of light

2.2.9 Homophones

- words with the same pronunciation, but different spelling and sense
- for example:
  1. "would" and "wood"
  2. "two" and "too" and "to"
  3. "where" and "were"

2.2.10 Homographs

- a pseudo-subset of homophones: words with the same spelling, but different sense (and potentially different pronunciation)
- for example:
  1. "bass" (as a fish **BA-S** or as an instrument **BEH-I-S**)
  2. "content" ("That was good content" vs "Jane was content with her performance")

2.2.11 Holonyms

- a word that names the whole of which a given word is a part.
- for example:
    1. "car" is a holonym of "wheel" or "chassis"
    2. "face" is a holonym of "mouth" or "eye"

2.2.12 Meronyms

- a word that names a part of a whole.
- for example:
    1. "leg" is a meronym of "chair"
    2. "pag" is a meronym of a "book"

2.2.13 Similarity and Gradation

- words which are not quite **synonyms**, but have a similar or graded sense
- for example:
    1. "soggy", "damp" and "humid" are all different gradations of "wet"
- a QA machine needs to recognise these gradations:

> ***Q***: *What is a good way to remove wine stains?*
> ***Corpus***: *Salt is a great way to eliminate wine stains.*

- whilst "eliminate" is a much stronger word, it will lead to "removing" wine stains, so "salt" will be an appropriate answer

2.2.14 Additional Requirement: Inference

Being able to **infer** conclusions is key in QA systems:

> ***Q***: *Did Poland reduce its carbon emissions since 1989?*
> ***Corpus***: *Poland is a country in Central Europe. Due to the collapse of the industrial sector after the end of communism in 1989, all countries in Central Europe saw a fall in carbon emissions.*

Overall, we can draw the following lessons:

Some lessons to draw

- Words are typically semantically ambiguous
- There's a lot of regularity (and hence predictability) in the range of senses a word can take
- Those senses also influence the word's syntactic behaviour
- But all regularities admit (arbitrary) exceptions
- Word senses can be **productive**, making a dictionary model (like WordNet) inadequate
- But it's a dominant model in CL these days, and works quite well in lots of cases.

2.3 WordNet

- **What is WordNet?**
  - a **lexical database**
  - each word has an associated **synset** (set of synonymous words), denoting the possible **senses** of a word

S1: a sense of concern with and curiosity about someone or something, Synonym: involvement

S2: the power of attracting or holding one's interest (because it is unusual or exciting etc.), Synonym: interestingness

S3: a reason for wanting something done, Synonym: sake

S4: a fixed charge for borrowing money; usually a percentage of the amount borrowed

S5: a diversion that occupies one's time and thoughts (usually pleasantly), Synonyms: pastime, pursuit

S6: a right or legal share of something; a financial involvement with something, Synonym: stake

S7: (usually plural) a social group whose members control some field of activity and who have common aims, Synonym: interest group

Figure 15: "Interest" has 7 different senses in WordNet.

She pays 3% **interest** on the loan.
He showed a lot of **interest** in the painting.
Microsoft purchased a controlling **interest** in Google.
It is in the national **interest** to invade the Bahamas.
I only have your best **interest** in mind.
Playing chess is one of my **interests**.
Business **interests** lobbied for the legislation.

- **How are synsets organised?**
  - sysnsets specific to:
    - nouns (`.n`)
    - verbs (`.v`)
    - adjectives (`.a, .s`)
    - adverbs (`.r`)

    (WordNet is split into 3 databases: nouns, verbs & adjectives/adverbs)
  - if we look at `http://wordnetweb.princeton.edu/perl/webwn?s=CAR` we can see that **synonyms** are part of the same **synset**, whilst **polysemous words** are part of **multiple synsets**

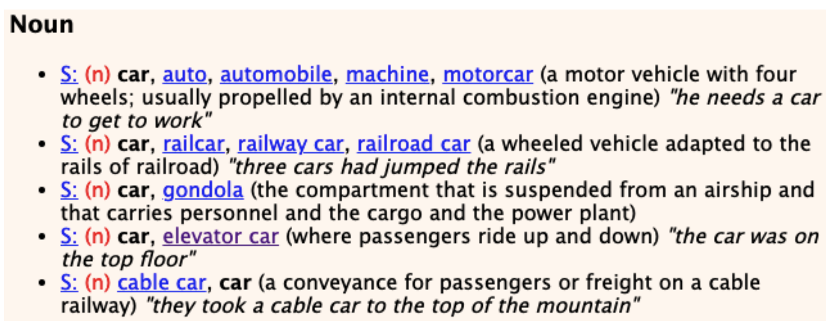


Figure 16: Sense `car.n.01` is associated with synonyms like "auto", "automobile". `car.n.01` and `car.n.04` are polysemous, so they are part of different synsets.

  - further, each **synset** can be associated with other **synsets**:
    1. *Hyponym/hypernm*: **IS-A** (i.e chair-furniture
    2. *Meronym*: **PART-WHOLE** (i.e leg-chair)
    3. *Antonym*: **OPPOSITED** (i.e good-bar)

- **How complete is WordNet?**
  - contains 118k unique synsets
  - however, it misses:
    - **multiword expressions**: "take a break", "pay attention"
    - **neologisms**: "hoodie", "facepalm"
    - **names**: "Microsoft"
    - predictable word uses, albeit uncommon: "Badger is a delicacy in Mongolia"

3 Word Sense Disambiguation

3.1 Defining Word Sense Disambiguation

- **What is Word Sense Disambiguation?**
  - given a word token and its context, determine the **sense** in which the word is used
  - useful for **machine translation**, **question-answering**, **information retrieval** and **text classification**
- **What are the 2 types of WSD?**
  - **Lexical Sample**: use **supervised machine learning**, and treat as **classification** - given a list of words, their senses, and "golden labels"; train by using context for each word

    She pays 3% **interest/INTEREST-MONEY** on the loan.
    He showed a lot of **interest/INTEREST-CURIOSITY** in the painting.
    Playing chess is one of my **interests/INTEREST-HOBBY**.

  - **All-Words**: given an entire corpus, and word senses, ask to annotate each content word (similar to POS tagging

3.2 Naive Bayes for WSD

- **How can NB be used for WSD?**
  - consider a word, with a set of senses S
  - from the word, we extract a **feature vector**
  - NB tells us that:

$$\begin{aligned}\hat{s} &= \underset{s\in S}{argmax} P(s \mid \underline{f}) \\ &= \underset{s\in S}{argmax} \frac{P(\underline{f} \mid s)P(s)}{P(\underline{f})} \\ &= \underset{s\in S}{argmax} P(\underline{f} \mid s)P(s) \\ &= \underset{s\in S}{argmax} P(s)\prod_{j=1}^{n} P(f_j \mid s)\end{aligned}$$

- **Which features can be used for NB (and other supervised task)**
  - **direct neighbours** (aka **collocational features)** (i.e "interest paid", "rising interest) - provide local context information
  - **content** words in a 50 word window (i.e "financial", "pursued")
  - text topic
  - POS tag, POS tag of context
  - syntactically related words
  - BOW of most common words in text
- **What other methods are available?**

  - decision lists
  - decision trees
  - neural networks

- **How can we evaluate WSD?**
  - **Extrinsic**: evaluate in downstream applications (question-answering, machine translation)
    * hard and time-consuming to implement
    * might indicate performance only in context of application (i.e not generalisable)
  - **Intrinsic**: evaluate against gold labels (i.e accuracy/precision/recall)
  - **Baseline**: compare with a "naive" WSD model (i.e picks most common sense always)

- **What are general issues with WSD?**
  - **Fine-Grain**
    * how coarse should the gold-standard be? (i.e how many senses to consider)
  - **Expensive**
    * very hard to annotate corpora with word senses, particularly if ver find-grained
  - **Number of Classifiers**
    * train separate classifier to disambiguate individual words
    * hard for infrequent words
    * motivation for unsupervised/semi-supervised methods

- **What are alternative approaches to WSD?**
  - can define coarse **semantic categories**
  - disambiguating is simpler:

$$\textit{apple} \rightarrow \textit{food}$$

$$\textit{apple} \not\rightarrow \textit{organisation}$$

  - applicable even if word not in lexicon
  - alternatively, can use **supersenses** like in **WordNet**

N:Tops	N:OBJECT	V:COGNITION
N:ACT	N:PERSON	V:COMMUNICATION
N:ANIMAL	N:PHENOMENON	V:COMPETITION
N:ARTIFACT	N:PLANT	V:CONSUMPTION
N:ATTRIBUTE	N:POSSESSION	V:CONTACT
N:BODY	N:PROCESS	V:CREATION
N:COGNITION	N:QUANTITY	V:EMOTION
N:COMMUNICATION	N:RELATION	V:MOTION
N:EVENT	N:SHAPE	V:PERCEPTION
N:FEELING	N:STATE	V:POSSESSION
N:FOOD	N:SUBSTANCE	V:SOCIAL
N:GROUP	N:TIME	V:STATIVE
N:LOCATION	V:BODY	V:WEATHER
N:MOTIVE	V:CHANGE	

4 Distributional Semantics: Categorising Semantic Similarity

4.1 Motivation: The Distributional Hypothesis

- **What is the distributional hypothesis?**
  - **meaning** can be inferred based on **context words**
  - hence:

$$\textit{similar context} \implies \textit{similar meaning}$$

a bottle of ***tezgüino*** is on the table
everybody likes ***tezgüino***
tezgüino makes you drunk
we make ***tezgüino*** out of corn

Figure 17: We can infer that "tezgüino", despite never having read such a thing, is an alcoholic beverage.

- **Can't we use a thesaurus to gauge similarity?**
  1. Every language doesn't have a thesaurus
  2. Many missing words/phrases
  3. Don't work well with verbs/adjectives

4.2 The Idea of Distributional Semantics

- **What are vector semantics?**
  - encode **word meaning** via **vectors**
  - known as **embeddings**
- **What are distributional semantic models?**
  - model **distributional semantics** by means of **vectors**
  - each word defined as a **vector** based on its **context**
  - also known as **vector space model**
- **What considerations should be made when defining the vectors?**
  1. What type of **context** should be considered?
  2. How should context words be **weighted**?
  3. How can **similarity** be measured?
  4. How do we **evaluate** the resulting vector representations?
- **What is a naive distributional semantic representation?**
  - consider a vocabulary V
  - build a matrix of size $|V| \times |V|$ with entry (i, j) denoting the number of times that word V_j appears in the context of word V_i

– the ith row of the matrix is the **vector** representation of V_i:

	aardvark	...	computer	data	result	pie	sugar	...
cherry	0	...	2	8	9	442	25	...
strawberry	0	...	0	0	1	60	19	...
digital	0	...	1670	1683	85	5	4	...
information	0	...	3325	3982	378	5	13	...

– this approach has many flaws, including the fact that it leads to **sparse** vectors

4.3 Defining Context

- **What are word co-occurrences?**
  - can lead to notions of similarity
  - **First-Order Co-Occurrence** (aka *syntagmatic association*): words occurring close to each other (i.e "wrote" and "book")
  - **Second-Order Co-Occurrence** (aka *paradigmatic association*): words occurring in similar contexts (i.e "wrote" and "spoke")
- **What types of contexts can be considered?**
  - **large** windows around the **target word**
    * typically used to gauge **topic** similarity
    * at the limit, consider the number of times a word occurs in the document
  - **small** windows around the **target word**
    * better captures individual word similarity
    * can lead to relations beyond co-occurrence (i.e dependency relation between words)
  - typically ignore **stopwords** (very common + uninformative)

4.4 Weighting Vector Representations

- **Why are frequencies not good when defining vector components?**
  - better than using binary indicators (1 if word appears in context, 0 otherwise)
  - frequencies still improvable: **skewed** and **misrepresentative**
    * frequent words everywhere, **independent** of **target word**
- **What are collocations?**
  - word pairs which appear **frequently** together, but **infrequently** in other contexts
- **What is pointwise mutual information?**
  - gauges the idea of collocation
  - the higher the PMI, the more dependent 2 words are of each other:

$$PMI(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)}$$

    where:
    * $P(x, y)$: probability of x, y appearing together

* $P(x)P(y)$: probability of x, y appearing together **if they are sampled independently**

- **What are issues associated with PMI?**

  1. Computed using **Counts**
     - oversensitive when infrequent words co-occur

I_{1000}	w^1	w^2	$w^1 w^2$	Bigram
16.95	5	1	1	Schwartz eschews
15.02	1	19	1	fewest visits
13.78	5	9	1	FIND GARDEN
12.00	5	31	1	Indonesian pieces
9.82	26	27	1	Reds survived
9.21	13	82	1	marijuana growing
7.37	24	159	1	doubt whether
6.68	687	9	1	new converts
6.00	661	15	1	like offensive
3.81	159	283	1	must think

  2. **Negative** PMI is unreliable
     - implies words co-occur less often than expected by chance
     - unreliable, since need large corpus for this to happen (i.e if $P(x) = P(y) = 10^{-6}$, would require $P(x, y) < 10^{-12}$, so large corpus required)
     - can use PPMI alternatively: any negative PMI becomes 0

- **What alternatives exist to PMI for finding collocations?**
  - Student t-test
  - Perason's χ^2 statistic
  - likelihood ratio test

4.5 Measuring Word Similarity

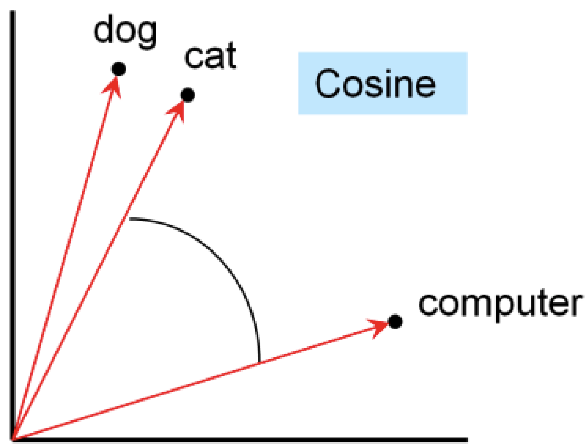
- **Why is distance not a good measure for similarity?**
  - if a dimension has extreme value, distance very large
  - however, the vectors could be "on top" of each other, so very similar

- **What are the issues associated with using the standard dot product?**
  - dot product is a good measure of similarity:

$$\underline{v} \cdot \underline{w} = \sum_{i=1}^{n} v_i w_i = \|\underline{v}\| \|\underline{w}\| \cos(\theta)$$

  - however, very conditioned by **vector magnitude** $\implies$ high frequency words skew towards similarity, which is undesirable (so 2 very large vectors could be thought as similar, even if unrelated

- **How can we fix the issue with the dot product?**
  - **normalise** the vector to have unit length
  - then:
$$\underline{v} \cdot \underline{w} = \cos(\theta)$$
  - hence, similarity as a measure of **angle**



- **What other measures of similarity are there?**
  - Jaccard measure
  - Dice measure
  - Jenson-Shannon divergence

4.6 Evaluating Distributional Representations

- **How can we evaluate the word vector embeddings?**
  1. **Extrinsic**
     - evaluate performance in downstream application (i.e question answering, automatic essay marking)
     - hard, not general purpose
  2. **Intrinsic**
     - evaluate by comparing with **psycholinguistic data**:
       * **Relatedness Judgement**: ask humans to rate the degree of similarity between concepts (1-10), based on some scale:
$$Rate(Lemon, Truth) = 1 \land Rate(Lemon, Orange) = 10 \implies Rate(Lemon, Flower) = ?$$
         Very person dependent, very question dependent (how is the question asked?)
       * **Word Association**: given a word, count how many times another word comes to mind; use it to compute probabilities

		ORANGE	0.16
		SOUR	0.11
		TREE	0.09
LEMON	$\Rightarrow$	YELLOW	0.08
		TEA	0.07
		JUICE	0.05
		...	

∗ **Human vs Machine**: have humans and computer rank create a ranked list of words related to w; use Spearman rank correlation to see how well the rankings match

4.7 Dealing with Sparsity

- **How can we convert the word representations into a less sparse vector?**
  - we are considering vectors in $\mathbb{R}^{|V|}$ - extremely sparse
  - apply **Latent Semantic Analysis** for dimensionality reduction:

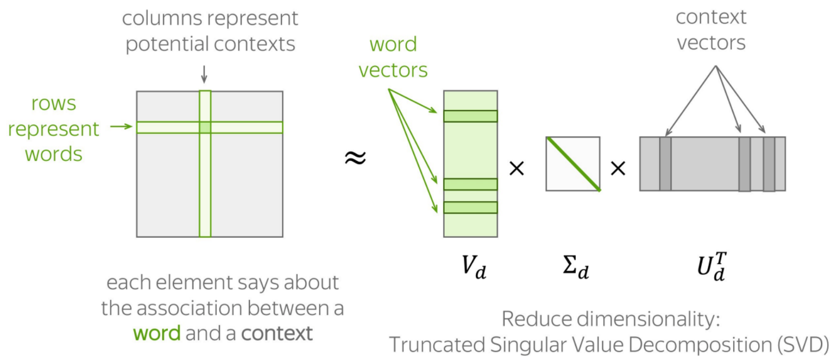


Figure 18: We can create a matrix using our word vector representations.
If we apply **Singular Value Decomposition**, it allows us to decompose a matrix into a product of 3 matrices.
If we **truncate** the SVD, we pick only the first d columns of V, Σ, U^T.
The matrices V_d, U_d^T can then be used as our compressed word representations (technically, their rows/columns)

- **Can we learn compressed representations directly?**
  - mainly using **neural networks**
  - for example, hidden layers when using a NN to predict context words when using an input word

4.8 Compositionality

- **What is compositionality?**
  - meaning derived by **composing words**
  - for example, “red barn” means that there is a “barn” that is “red”

- not all language is compositional (i.e "the White House")

- **How can compositionality be added to a vector space?**
  - define operator $\oplus$ such that:

$$meaning(w_1 w_2) = meaning(w_1) meaning(w_2)$$

  - possibly:
    * vector addition (empirically not good)
    * tensor product (multiply entries together - not too meaningful)
    * train NN to learn a non-linear operation (currently)