Linguistic Intermezzo

Philipp Koehn

15 October 2024

A Naive View of Language

- Language needs to name
  - nouns: objects in the world (*dog*)
  - verbs: actions (*jump*)
  - adjectives and adverbs: properties of objects and actions (*brown*, *quickly*)

A Naive View of Language

- Language needs to name
  - nouns: objects in the world (*dog*)
  - verbs: actions (*jump*)
  - adjectives and adverbs: properties of objects and actions (*brown*, *quickly*)

- Relationship between these have to specified
  - word order
  - morphology
  - function words

morphology

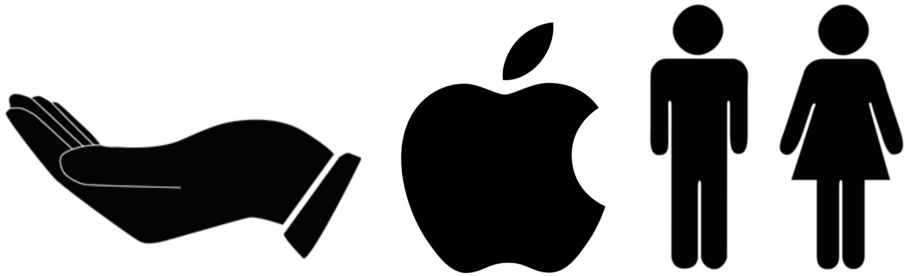
Marking of Relationships: Agreement

- From Catullus, First Book, first verse (Latin):
- Gender (and case) agreement links adjectives to nouns

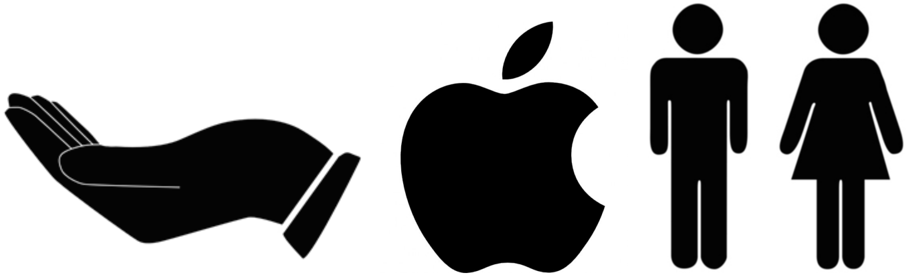
Cui	*dono*	*lepidum*	*novum*	*libellum*	*arida*	*modo*	*pumice*	*expolitum*	*?*
Whom	*I-present*	*lovely*	*new*	*little-book*	*dry*	*manner*	*pumice*	*polished*	*?*

(To whom do I present this lovely new little book now polished with a dry pumice?)

Marking of Relationships to Verb: Case



Marking of Relationships to Verb: Case



- German:

Die Frau	*gibt*	*dem Mann*	*den Apfel*
The woman	*gives*	*the man*	*the apple*
subject		indirect object	object

- Case inflection indicates role of noun phrases

Writingwordstogether

- Definition of word boundaries purely an artifact of writing system

Writingwordstogether

- Definition of word boundaries purely an artifact of writing system
- Differences between languages
  - Agglutinative compounding
    Informatikseminar vs. *computer science seminar*
  - Function word vs. affix

Writingwordstogether

- Definition of word boundaries purely an artifact of writing system

- Differences between languages
  - Agglutinative compounding

    Informatikseminar vs. *computer science seminar*

  - Function word vs. affix

- Border cases
  - *Joe's* — one token or two?
  - Morphology of affixes often depends on phonetics / spelling conventions

    dog+s → *dogs* vs. *pony* → *ponies*

    ... but note the English function word *a*:

    a donkey vs. *an aardvark*

Changing Part-of-Speech

- Derivational morphology allows changing part of speech of words

Changing Part-of-Speech

- Derivational morphology allows changing part of speech of words

- Example:
  - base: *nation*, noun

$\rightarrow$ *national*, adjective
$\rightarrow$ *nationally*, adverb
$\rightarrow$ *nationalist*, noun
$\rightarrow$ *nationalism*, noun
$\rightarrow$ *nationalize*, verb

Changing Part-of-Speech

- Derivational morphology allows changing part of speech of words

- Example:
  - base: *nation*, noun

  $\rightarrow$ *national*, adjective
  $\rightarrow$ *nationally*, adverb
  $\rightarrow$ *nationalist*, noun
  $\rightarrow$ *nationalism*, noun
  $\rightarrow$ *nationalize*, verb

- Sometimes distinctions between POS quite fluid (enabled by morphology)
  - *I want to integrate morphology*
  - *I want the integration of morphology*

Meaning Altering Affixes

- English

undo

redo

hypergraph

Meaning Altering Affixes

- English

undo

redo

hypergraph

- German: *zer-* implies action causes destruction

Er **zer***redet das Thema* → *He talks the topic* **to death**

Meaning Altering Affixes

- English

undo

redo

hypergraph

- German: *zer-* implies action causes destruction

Er **zer***redet das Thema* → *He talks the topic* **to death**

- Spanish: *-ito* means object is small

burro → *burrito*

Adding Subtle Meaning

- Morphology allows adding subtle meaning
  - verb tenses: time action is occurring, if still ongoing, etc.
  - count (singular, plural): how many instances of an object are involved
  - definiteness (*the cat* vs. *a cat*): relation to previously mentioned objects
  - grammatical gender: helps with co-reference and other disambiguation
- Sometimes redundant: same information repeated many times

Unknown Words

- Ratio of unknown words in WMT 2013 test set:

Source language	Ratio unknown
Russian	2.0%
Czech	1.5%
German	1.2%
French	0.5%
English (to French)	0.5%

- Caveats:
  - corpus sizes differ
  - not clear which unknown words have known morphological variants

Differently Encoded Information

- Languages with different sentence structure

das	*behaupten*	*sie*	*wenigstens*
this	*claim*	*they*	*at least*
the		*she*	

- Convert from inflected language into configuration language (and vice versa)

- Ambiguities can be resolved through syntactic analysis
  - the meaning *the* of *das* not possible (not a noun phrase)
  - the meaning *she* of *sie* not possible (subject-verb agreement)

Non-Local Information

- Pronominal anaphora

I saw the movie and it is good.

- How to translate *it* into German (or French)?
  - *it* refers to *movie*
  - *movie* translates to *Film*
  - *Film* has masculine gender
  - ergo: *it* must be translated into masculine pronoun *er*

- We are not handling pronouns very well

Complex Semantic Inference

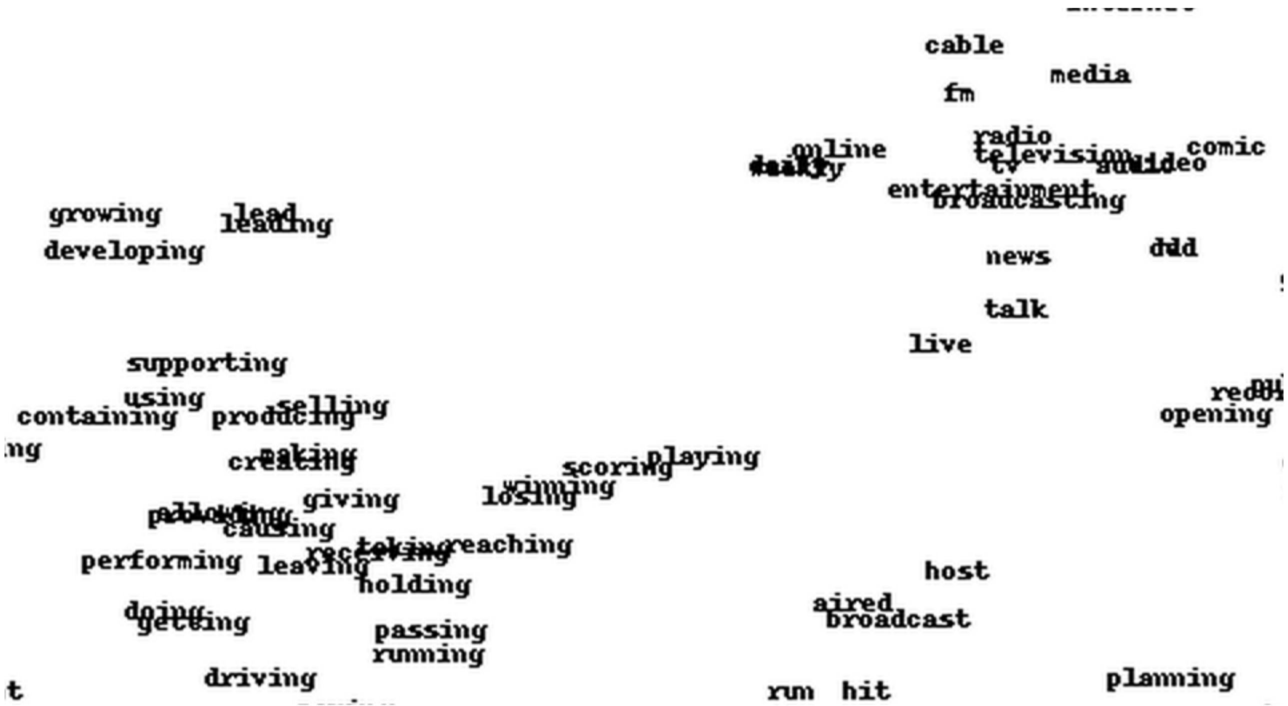
- Example

Whenever I visit my uncle and his daughters,
I can't decide who is my favorite cousin.

- How to translate *cousin* into German? Male or female?
- Google Translate is getting this wrong (checked October 2024)

word embeddings

Word Embeddings



- In neural translation models words are mapped into, say, 500-dimensional continuous space
- Contextualized in encoder layers

Latent Semantic Analysis

- Word embeddings not a new idea
- Representing words based on their context has long tradition in natural language processing

Latent Semantic Analysis

- Word embeddings not a new idea
- Representing words based on their context has long tradition in natural language processing
- Co-occurence statistics

word	context			
	cute	*fluffy*	*dangerous*	*of*
dog	231	76	15	5767
cat	191	21	3	2463
lion	5	1	79	796

- But: large counts of function words misleading

Pointwise Mutual Information

- Pointwise mutual information

$$\mathrm{PMI}(x;y) = \log \frac{p(x,y)}{p(x)p(y)}$$

- Intuition: measures how much more frequent than chance

word	context			
	cute	*fluffy*	*dangerous*	*of*
dog	9.4	6.3	0.2	1.1
cat	8.3	3.1	0.1	1.0
lion	0.1	0.0	12.1	1.0

- Similar words have similar vectors

Singular Value Decomposition

- Raw co-occurence statistics matrix very sparse

$\Rightarrow$ Reduce into lower dimensional matrix

Singular Value Decomposition

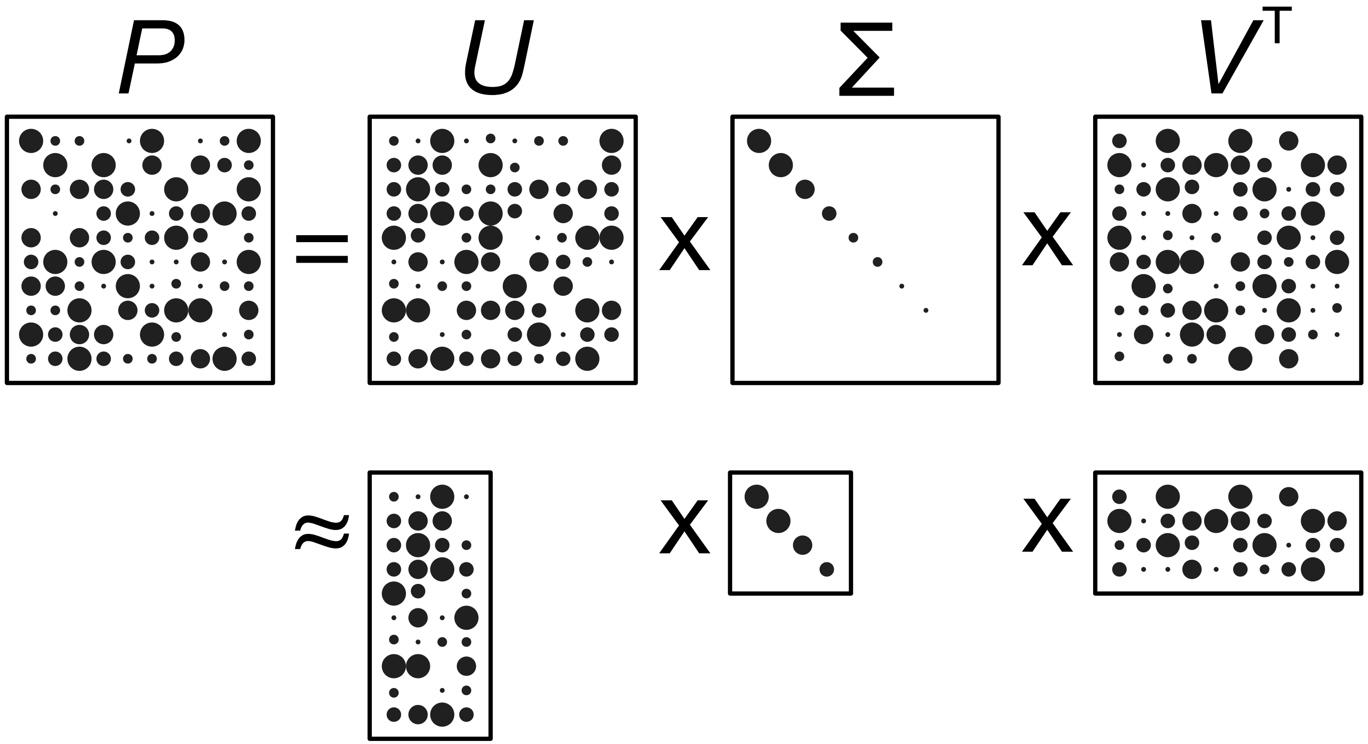
- Raw co-occurence statistics matrix very sparse

⇒ Reduce into lower dimensional matrix

- Factorize the PMI matrix P into
  - two orthogonal matrices U and V
    (i.e. UU^T and VV^T are an identity matrix)
  - diagonal matrix Σ
    (i.e., it only has non-zero values on the diagonal)

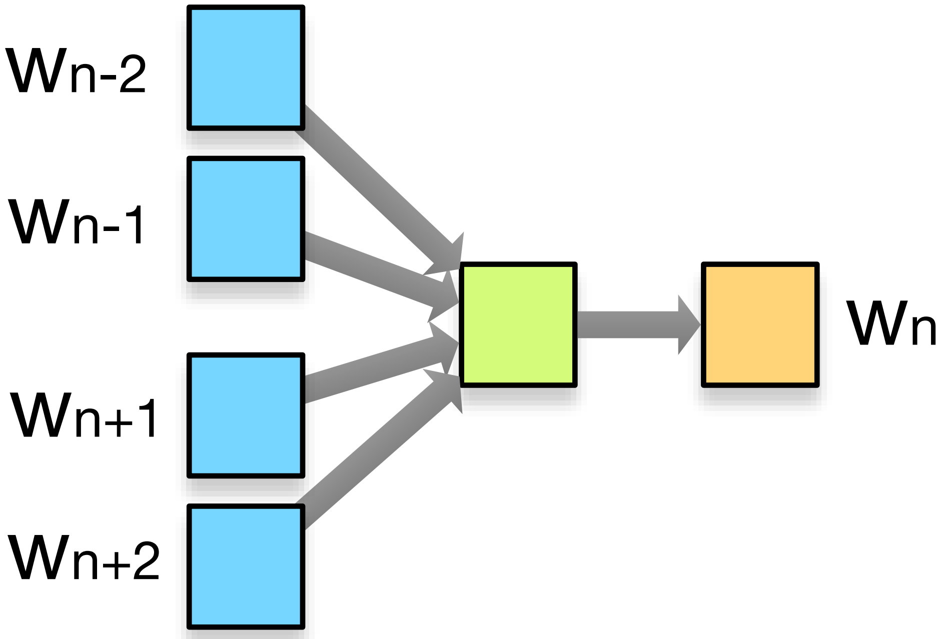
$$P = U\Sigma V^T$$

Singular Value Decomposition



- Not going into details how to compute this
- Geometric interpretation: rotation U, a stretching Σ, and another rotation V^T
- Matrices U and V^T play similar role as embedding matrices

Continuous Bag of Words (CBOW)



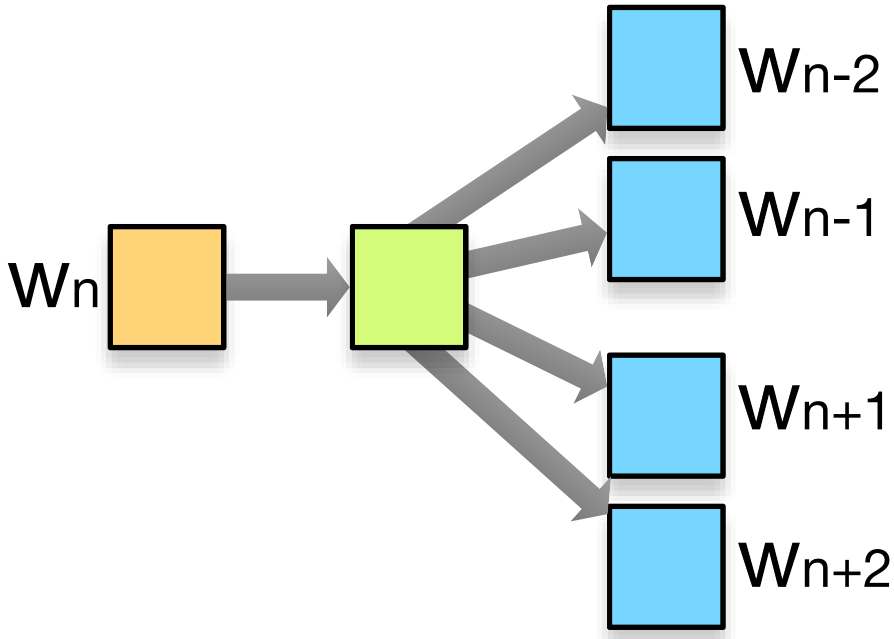
- Predict word from context

$$h_t = \frac{1}{2n} \sum_{j \in \{-n,...,-1,1,...,n\}} C w_{t+j}$$

$$y_t = \text{softmax}(U h_t)$$

- Similar to n-gram language model

Skip Gram



- Predict context from word

$$y_t = \text{softmax}(UCw_t)$$

- C input word embedding matrix, U output word embedding matrix

GloVe

- Global Vectors: use co-occurrence statistics

word	context			
	cute	*fluffy*	*dangerous*	*of*
dog	231	76	15	5767
cat	191	21	3	2463
lion	5	1	79	796

- Predict the values in this matrix X, using target word embeddings v_i and context word embeddings $\tilde{v}_j$

$$\text{cost} = \sum_i \sum_j \tilde{v}_j^T |v_i - \log X_{ij}|$$

- Training: loop over all words, and their context words

Refinements

- Bias terms b and $\tilde{b}$

$$\text{cost} = \sum_i \sum_j |b_i + \tilde{b}_j + \tilde{v}_j^T v_i - \log X_{ij}|$$

Refinements

- Bias terms b and $\tilde{b}$

$$\text{cost} = \sum_i \sum_j |b_i + \tilde{b}_j + \tilde{v}_j^T v_i - \log X_{ij}|$$

- Most word pairs (i, j) meaningless, especially for rare words

Refinements

- Bias terms b and $\tilde{b}$

$$\text{cost} = \sum_i \sum_j |b_i + \tilde{b}_j + \tilde{v}_j^T v_i - \log X_{ij}|$$

- Most word pairs (i, j) meaningless, especially for rare words

- Discount them with a scaling function

$$f(x) = \min(1, (x/x_{\max})^\alpha)$$

hyper parameter values, e.g., $\alpha = \frac{3}{4}$ and $x_{\max} = 200$

Refinements

- Bias terms b and $\tilde{b}$

$$\text{cost} = \sum_i \sum_j |b_i + \tilde{b}_j + \tilde{v}_j^T v_i - \log X_{ij}|$$

- Most word pairs (i, j) meaningless, especially for rare words

- Discount them with a scaling function

$$f(x) = \min(1, (x/x_{\text{max}})^\alpha)$$

hyper parameter values, e.g., $\alpha = \frac{3}{4}$ and $x_{\text{max}} = 200$

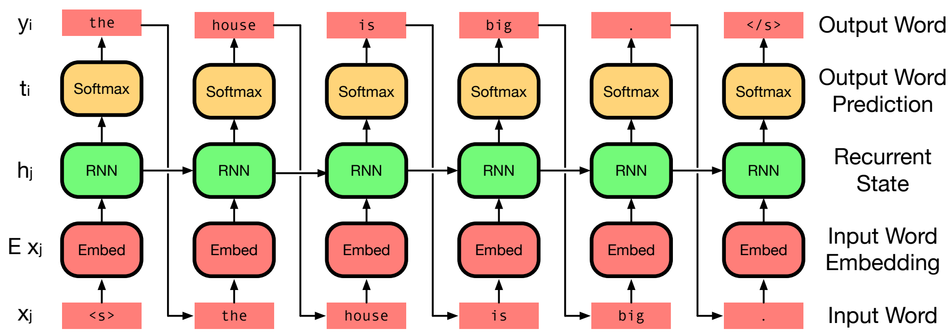
- Complete refined cost function

$$\text{cost} = \sum_i \sum_j f(X_{ij})(b_i + \tilde{b}_j + \tilde{v}_j^T v_i - \log X_{ij})^2$$

ELMo

- Word embeddings widely used in natural language processing
- But: better refine them in the sentence context

⇒ *Embeddings from language models* (ELMo)
(we have always done this in the encoder of our neural translation models)



- Several layers, use weighted sum of representations at different layers
  - syntactic information is better represented in early layers
  - semantic information is better represented in deeper layers.

BERT

- Contextualized word embeddings with Transformer model

- Masked training

```
The quick brown fox jumps over the lazy dog.
                     ⇑
 The quick MASK fox MASK over the lazy dog.
```

- Next sentence prediction

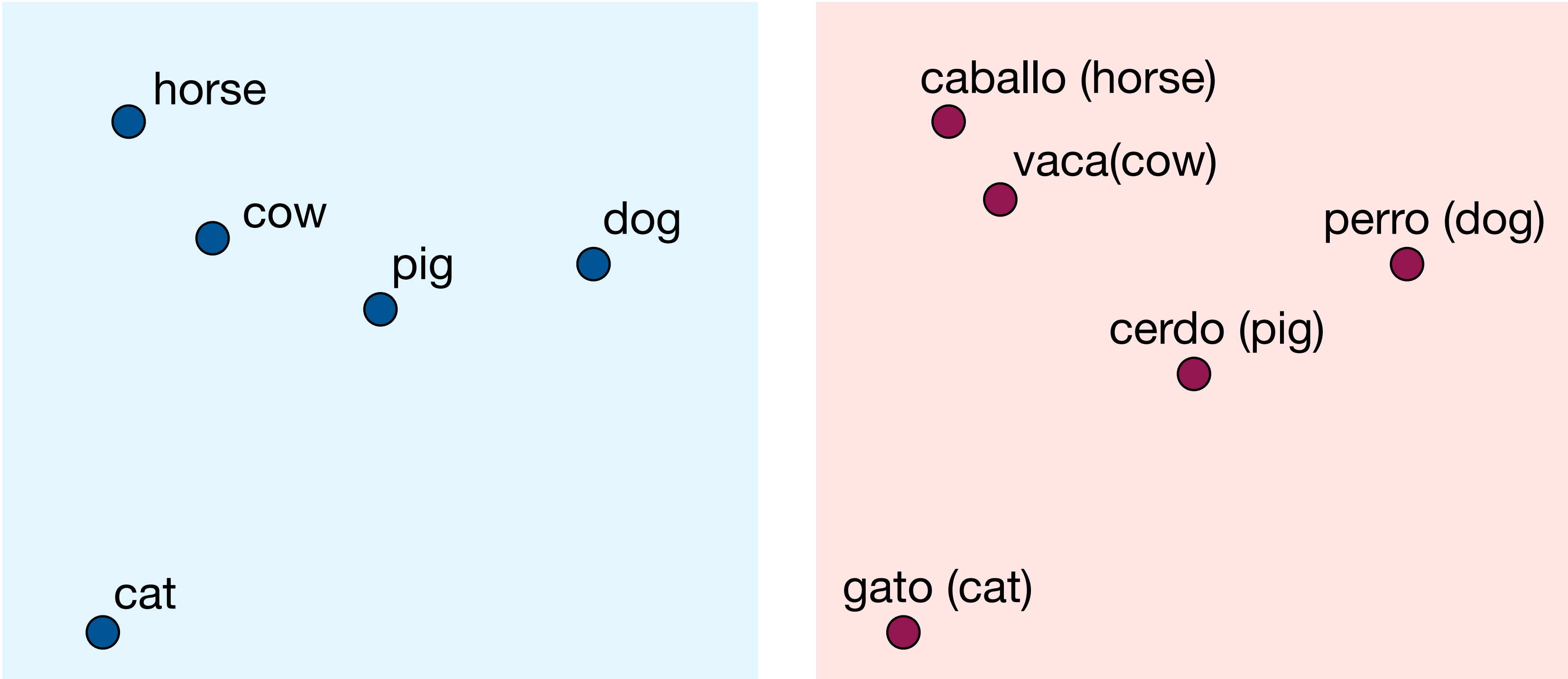
```
Each unhappy family is unhappy in its own way.
                     ⇑
        All happy families are alike.
```

multi-lingual word embeddings

Multi-Lingual Word Embeddings

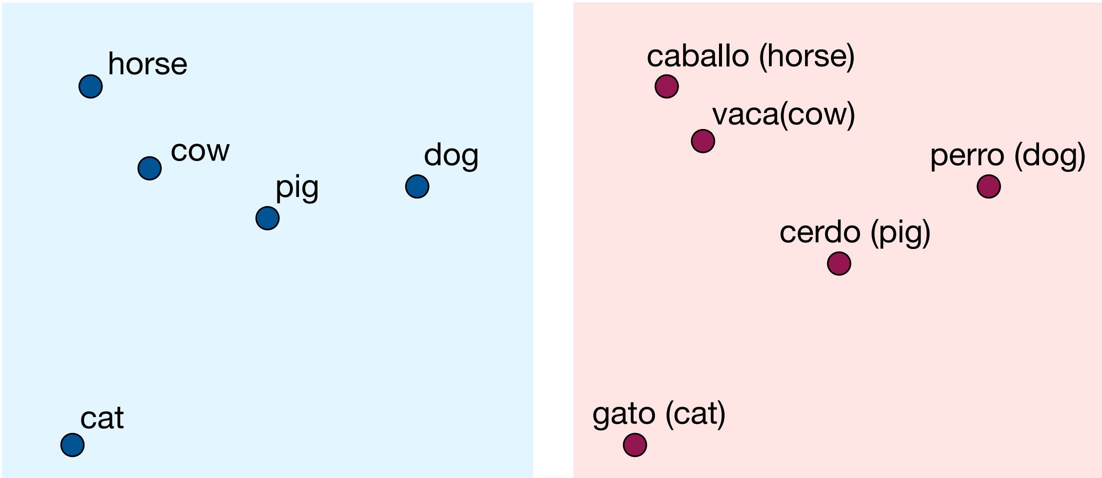
- Word embeddings often viewed as semantic representations of words
- Tempting to view embedding spaces as language-independent
  cat (English), *gato* (Spanish) and *Katze* (German) are mapped to same vector
- Common semantic space for words in all languages?

Language-Specific Word Embeddings



- Train English word embeddings C_E and Spanish word embeddings C_S

Mapping Word Embedding Spaces



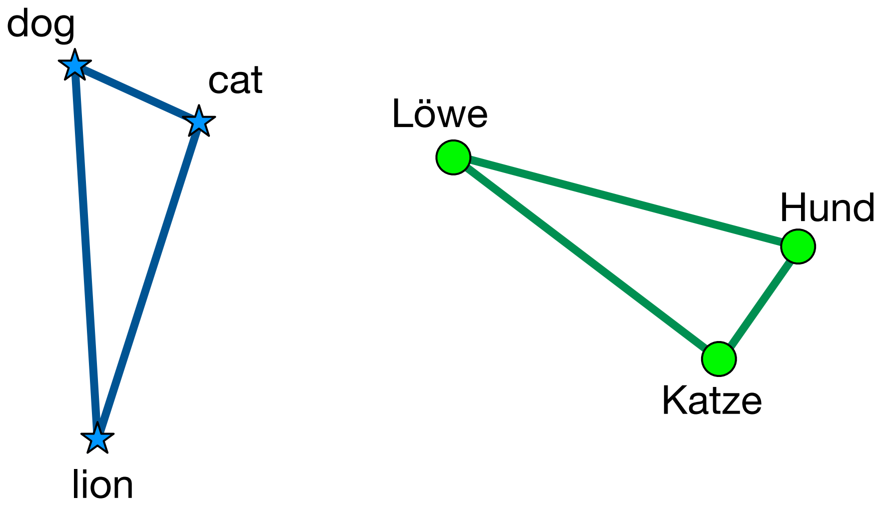
- Learn mapping matrix $W_{S \to E}$ to minimize Euclidean distance between each word and its translation

$$\text{cost} = \sum_i ||W_{S \to E}\; c_i^S - c_i^E||$$

- Needed: Seed lexicon of word translations (may be based on cognates)

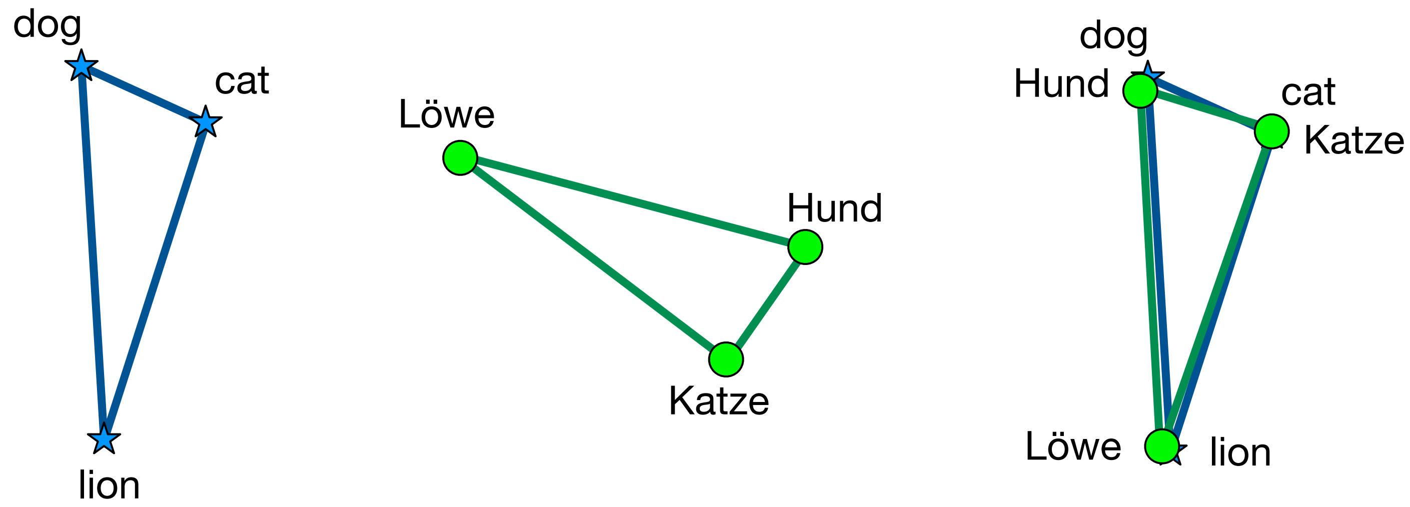
- Hubness problem: some words being the nearest neighbor of many words

Using only Monolingual Data



- Learn transformation matrix $W_{S \rightarrow E}$ without seed lexicon?
- Intuition: relationship between *dog*, *cat*, and *lion*, independent of language
- How can we rotate the triangle to match up?

Using only Monolingual Data



- One idea: learn transformation matrix $W_{\text{German}\rightarrow\text{English}}$ so that words match up

Adversarial Training

- Another idea: adversarial training
  - points in the German and English space do not match up

  → adversary can classify them as either German and English

Adversarial Training

- Another idea: adversarial training
  - points in the German and English space do not match up

  $\rightarrow$ adversary can classify them as either German and English

- Training objective of adversary to learn classifier P

$$\text{cost}_D(P|W) = -\frac{1}{n}\sum_{i=1}^{n} \log P(\text{German}|Wg_i) - \frac{1}{m}\sum_{j=1}^{m} \log P(\text{English}|e_j)$$

- Training objective of unsupervised learner

$$\text{cost}_L(W|P) = -\frac{1}{n}\sum_{i=1}^{n} \log P(\text{English}|Wg_i) - \frac{1}{m}\sum_{j=1}^{m} \log P(\text{German}|ej)$$

large vocabularies

Large Vocabularies

- Zipf's law tells us that words in a language are very unevenly distributed.
  - large tail of rare words
    (e.g., new words *retweeting, website, woke, lit*)
  - large inventory of names, e.g., *eBay, Yahoo, Microsoft*

Large Vocabularies

- Zipf's law tells us that words in a language are very unevenly distributed.
  - large tail of rare words
    (e.g., new words *retweeting, website, woke, lit*)
  - large inventory of names, e.g., *eBay, Yahoo, Microsoft*

- Neural methods not well equipped to deal with such large vocabularies
  (ideal representations are continuous space vectors $\rightarrow$ word embeddings)

Large Vocabularies

- Zipf's law tells us that words in a language are very unevenly distributed.
  - large tail of rare words
    (e.g., new words *retweeting, website, woke, lit*)
  - large inventory of names, e.g., *eBay, Yahoo, Microsoft*

- Neural methods not well equipped to deal with such large vocabularies
  (ideal representations are continuous space vectors $\rightarrow$ word embeddings)

- Large vocabulary
  - large embedding matrices for input and output words
  - prediction and softmax over large number of words

- Computationally expensive, both in terms of memory and speed

Special Treatment for Rare Words

- Limit vocabulary to 20,000 to 80,000 words
- First idea
  - map other words to unknown word token (UNK)
  - model learns to map input UNK to output UNK
  - replace with translation from backup dictionary
- Not used anymore, except for numbers and units
  - numbers: English *540,000*, Chinese *54* TENTHOUSAND, Indian *5.4 lakh*
  - units: map *25cm* to *10 inches*

Some Causes for Large Vocabularies

- Morphology

  tweet, *tweets*, *tweeted*, *tweeting*, *retweet*, ...

  → morphological analysis?

Some Causes for Large Vocabularies

- Morphology

  tweet, *tweets*, *tweeted*, *tweeting*, *retweet*, ...

  → morphological analysis?

- Compounding

  homework, *website*, ...

  → compound splitting?

Some Causes for Large Vocabularies

- Morphology

  tweet, *tweets*, *tweeted*, *tweeting*, *retweet*, ...

  → morphological analysis?

- Compounding

  homework, *website*, ...

  → compound splitting?

- Names

  Netanyahu, *Jones*, *Macron*, *Hoboken*, ...

  → transliteration?

Some Causes for Large Vocabularies

- Morphology

  tweet, *tweets*, *tweeted*, *tweeting*, *retweet*, ...

  → morphological analysis?

- Compounding

  homework, *website*, ...

  → compound splitting?

- Names

  Netanyahu, *Jones*, *Macron*, *Hoboken*, ...

  → transliteration?

⇒ Breaking up words into **subwords** may be a good idea

Byte Pair Encoding

- Start by breaking up words into characters

```
t h e ␣ f a t ␣ c a t ␣ i s ␣ i n ␣ t h e ␣ t h i n ␣ b a g
```

- Merge frequent pairs

```
t h→th    th e ␣ f a t ␣ c a t ␣ i s ␣ i n ␣ th e ␣ th i n ␣ b a g
a t→at    th e ␣ f at ␣ c at ␣ i s ␣ i n ␣ th e ␣ th i n ␣ b a g
i n→in    th e ␣ f at ␣ c at ␣ i s ␣ in ␣ th e ␣ th in ␣ b a g
th e→the  the ␣ f at ␣ c at ␣ i s ␣ in ␣ the ␣ th in ␣ b a g
```

- Each merge operation increases the vocabulary size
  - starting with the size of the character set (maybe 100 for Latin script)
  - stopping after, say, 50,000 operations

Byte Pair Encoding

Obama receives Net@@ any@@ ahu

the relationship between Obama and Net@@ any@@ ahu is not exactly friendly . the two wanted to talk about the implementation of the international agreement and about Teheran 's destabil@@ ising activities in the Middle East . the meeting was also planned to cover the conflict with the Palestinians and the disputed two state solution . relations between Obama and Net@@ any@@ ahu have been stra@@ ined for years . Washington critic@@ ises the continuous building of settlements in Israel and acc@@ uses Net@@ any@@ ahu of a lack of initiative in the peace process . the relationship between the two has further deteriorated because of the deal that Obama negotiated on Iran 's atomic programme . in March , at the invitation of the Republic@@ ans , Net@@ any@@ ahu made a controversial speech to the US Congress , which was partly seen as an aff@@ ront to Obama . the speech had not been agreed with Obama , who had rejected a meeting with reference to the election that was at that time im@@ pending in Israel .

Subwords

- Byte pair encoding induces subwords

- But: only accidentally along linguistic concepts of morphology
  - morphological: `critic@@ ises`, `im@@ pending`
  - not morphological: `aff@@ ront`, `Net@@ any@@ ahu`

- Still: Similar to unsupervised morphology (frequent suffixes, etc.)

Sentence Piece

▁Obama ▁receives ▁Net any ahu

▁the ▁relationship ▁between ▁Obama ▁and ▁Net any ahu ▁is ▁not ▁exactly ▁friendly ▁. ▁the ▁two ▁wanted ▁to ▁talk ▁about ▁the ▁implementation ▁of ▁the ▁international ▁agreement ▁and ▁about ▁Teheran ▁'s ▁destabil ising ▁activities ▁in ▁the ▁Middle ▁East ▁. ▁the ▁meeting ▁was ▁also ▁planned ▁to ▁cover ▁the ▁conflict ▁with ▁the ▁Palestinians ▁and ▁the ▁disputed ▁two ▁state ▁solution ▁. ▁relations ▁between ▁Obama ▁and Net ▁any ▁ahu ▁have ▁been ▁stra ined ▁for ▁years ▁. ▁Washington ▁critic ises ▁the ▁continuous ▁building ▁of ▁settlements ▁in ▁Israel ▁and ▁acc uses ▁Net any ahu ▁of ▁a ▁lack ▁of ▁initiative ▁in ▁the ▁peace ▁process ▁. ▁the ▁relationship ▁between ▁the ▁two ▁has ▁further ▁deteriorated ▁because ▁of ▁the ▁deal ▁that ▁Obama ▁negotiated ▁on ▁Iran ▁'s ▁atomic ▁programme ▁. ▁in ▁March ▁, ▁at ▁the ▁invitation ▁of ▁the ▁Republic ans ▁, ▁Net any ahu ▁made ▁a ▁controversial ▁speech ▁to ▁the ▁US ▁Congress ▁, ▁which ▁was ▁partly ▁seen ▁as ▁an ▁aff ront ▁to ▁Obama ▁. ▁the ▁speech ▁had ▁not ▁been ▁agreed ▁with ▁Obama ▁, ▁who ▁had ▁rejected ▁a ▁meeting ▁with ▁reference ▁to ▁the ▁election ▁that ▁was ▁at ▁that ▁time ▁im pending ▁in ▁Israel .

character-based models

Character-Based Models

- Explicit word models that yield word embeddings

- Standard methods for frequent words
  - distribution of `beautiful` in the data

  → embedding for `beautiful`

- Character-based models
  - create sequence embedding for character string `b e a u t i f u l`
  - training objective: match word embedding for `beautiful`

- Induce embeddings for unseen morphological variants
  - character string `b e a u t i f u l l y`

  → embedding for `beautifully`

- Hope that this learns morphological principles

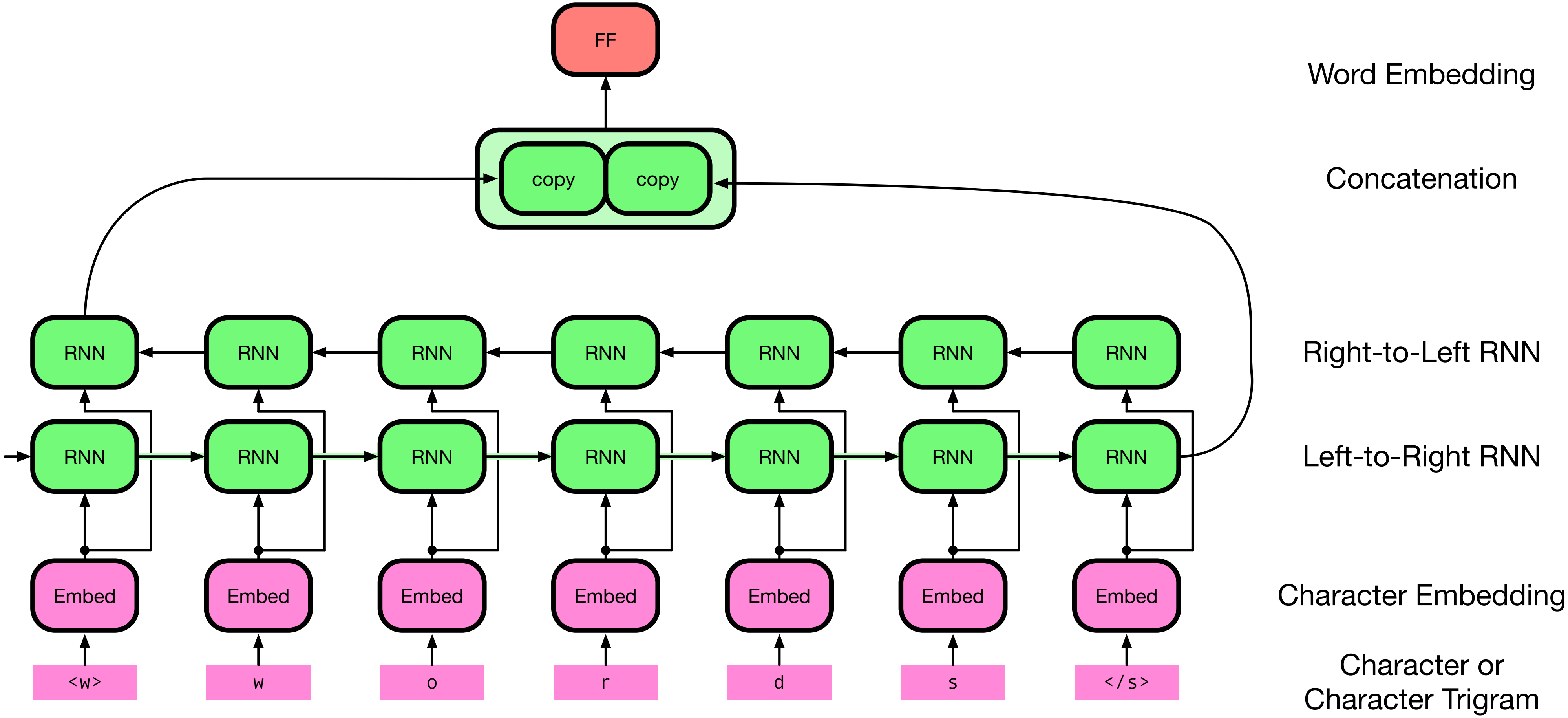
Character Sequence Models

- Same model as for words
- Tokens = single characters, incl. special space symbol
- But: generally poor performance
- With some refinements, use in output shown competitive

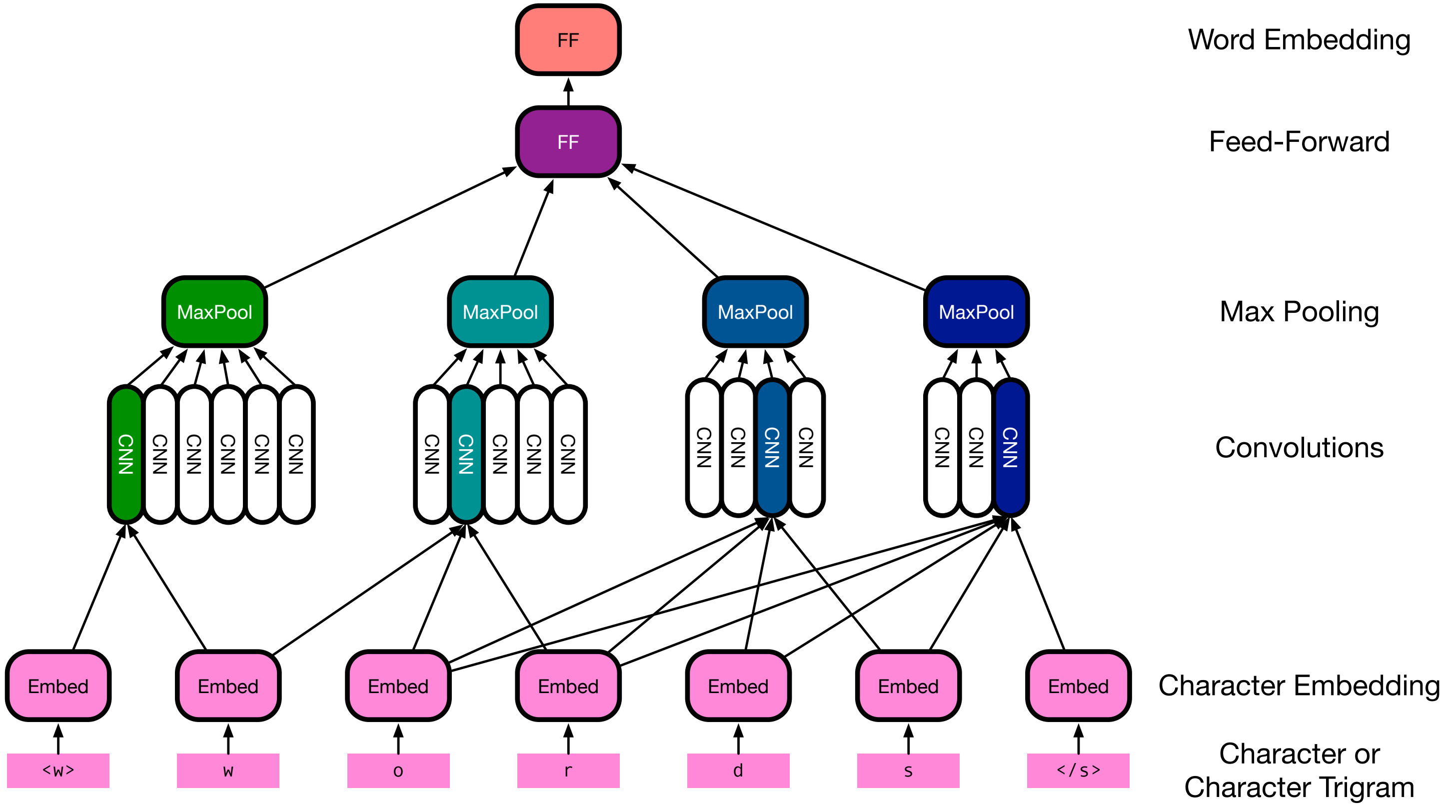
Character Based Word Models

- Word embeddings as before
- Compute word embeddings based on character sequence
- Typically, interpolated with traditional word embeddings

Recurrent Neural Networks



Convolutional Neural Networks



- Convolutions of different size: 2 characters, 3 characters, ..., 7 characters
- May be based on letter n-grams (trigrams shown)

syntax

Tree-Based Models

- Traditional statistical models operate on sequences of words

Tree-Based Models

- Traditional statistical models operate on sequences of words

- Many translation problems can be best explained by pointing to syntax
  - reordering, e.g., verb movement in German–English translation
  - long distance agreement (e.g., subject-verb) in output

Tree-Based Models

- Traditional statistical models operate on sequences of words

- Many translation problems can be best explained by pointing to syntax
  - reordering, e.g., verb movement in German–English translation
  - long distance agreement (e.g., subject-verb) in output

⇒ Translation models based on tree representation of language
  - successful for statistical machine translation
  - open research challenge for neural models

Dependency Structure

I	like	the	interesting	lecture
PRO	VB	DET	JJ	NN
$\downarrow$		$\downarrow$	$\downarrow$	$\downarrow$
like		lecture	lecture	like

- Center of a sentence is the verb
- Its dependents are its arguments (e.g., subject noun)
- These may have further dependents (adjective of noun)

Phrase Structure Grammar

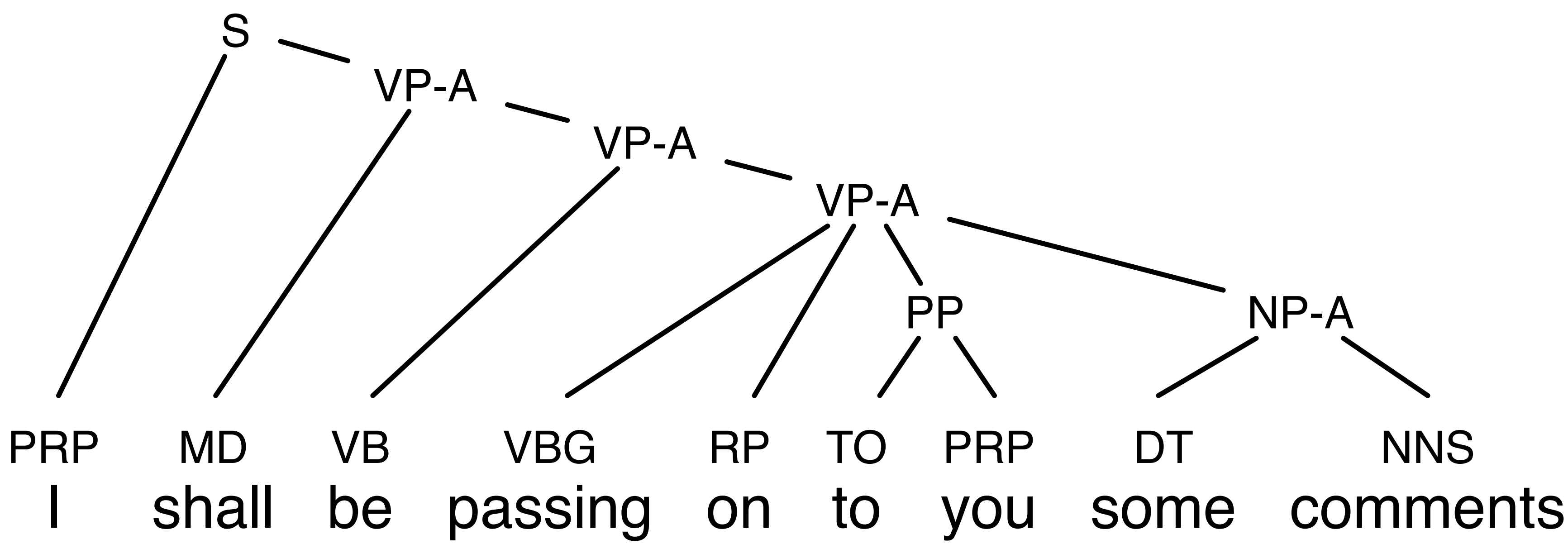
- Phrase structure
  - noun phrases: *the big man*, *a house*, ...
  - prepositional phrases: *at 5 o'clock*, *in Baltimore*, ...
  - verb phrases: *going out of business*, *eat chicken*, ...
  - adjective phrases, ...

Phrase Structure Grammar

- Phrase structure
  - noun phrases: *the big man*, *a house*, ...
  - prepositional phrases: *at 5 o'clock*, *in Baltimore*, ...
  - verb phrases: *going out of business*, *eat chicken*, ...
  - adjective phrases, ...

- Context-free Grammars (CFG)
  - non-terminal symbols: phrase structure labels, part-of-speech tags
  - terminal symbols: words
  - production rules: NT $\rightarrow$ [NT,T]+
    example: NP $\rightarrow$ DET NN

Phrase Structure Grammar



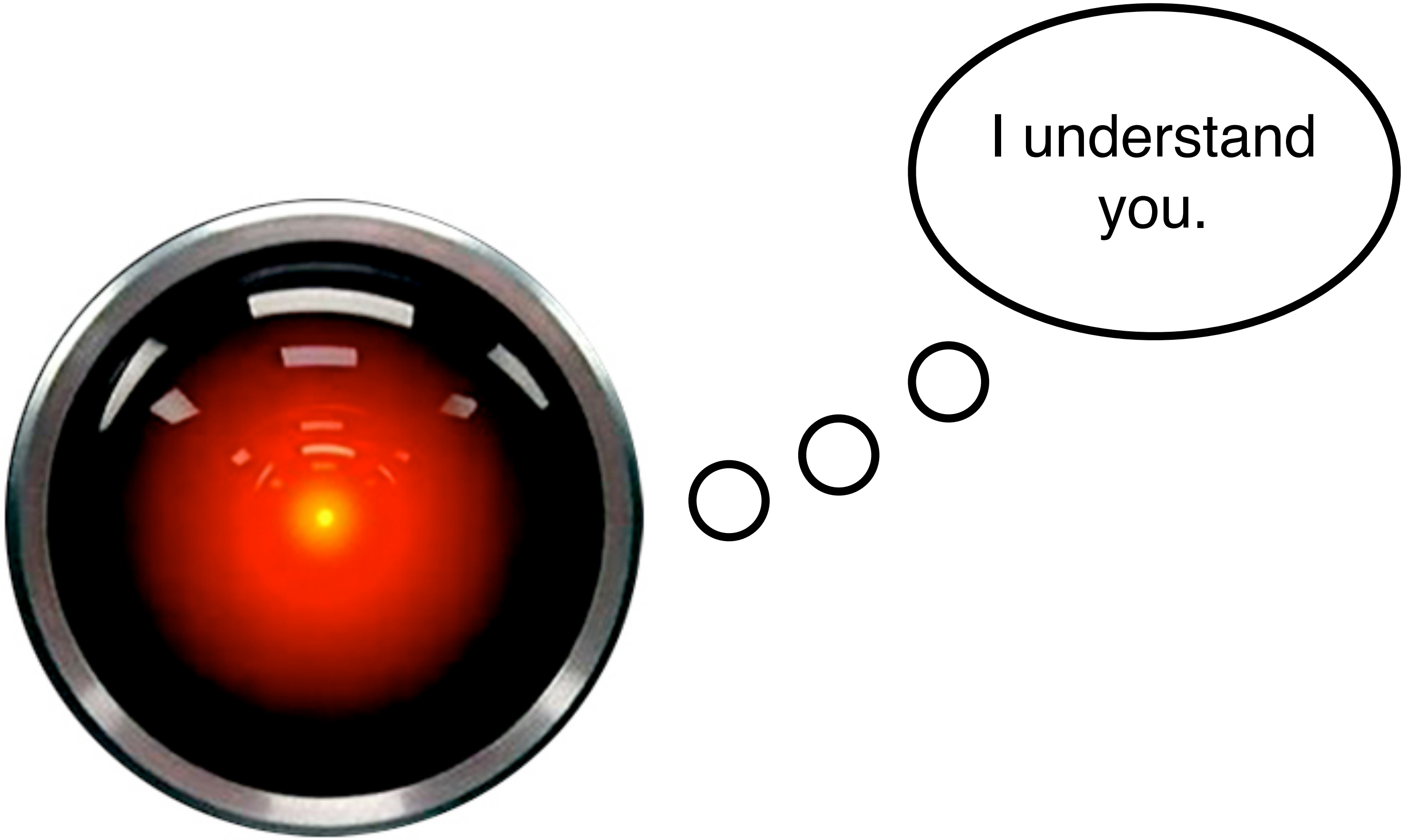
Phrase structure grammar tree for an English sentence
(as produced Collins' parser)

semantics

Meaning

- The grand goal of **artificial intelligence**
  - machines that do not mindlessly process data
  - ... but that ultimately **understand** its meaning
- But what is meaning?

Meaningful Machines



A Scale of Understanding?

wisdom

$\Uparrow$

knowledge

$\Uparrow$

data

Eliza

Joseph Weizenbaum wrote the computer program Eliza in 1969 to demonstrate how easily people can be fooled into thinking that machines have some deep understanding.

young woman: *Men are all alike.*
eliza: *In what way?*
young woman: *They're always bugging us about something specific or other.*
eliza: *Can you think of a specific example?*
young woman: *Well, my boyfriend made me come here.*
eliza: *Your boyfriend made you come here?*
young woman: *He says I'm depressed much of the time.*
eliza: *I am sorry to hear you are depressed.*
young woman: *It's true. I'm unhappy.*
eliza: *Do you think coming here will help you not to be unhappy?*

Online demo: `http://www-ai.ijs.si/eliza-cgi-bin/eliza_script`

semantic translation problems

Semantic Translation Problems

- Syntactic analysis may be ambiguous

Das	*Vorhaben*	*verwarf*	*die*	*Kommission*	*.*
the	*plan*	*rejected*	*the*	*commission*	*.*

- Both readings (SVO and OSV) are syntactically possible

- But: OSV reading is semantically much more plausible

⇒ Need for semantic model to produce semantically plausible output

lexical semantics

Word Senses

- Some words have multiple meanings
- This is called polysemy
- Example: *bank*
  - financial institution: *I put my money in the bank.*
  - river shore: *He rested at the bank of the river.*
- How could a computer tell these senses apart?

Homonym

- Sometimes two completely different words are spelled the same
- This is called a homonym
- Example: *can*
  - modal verb: *You can do it!*
  - container: *She bought a can of soda.*
- Distinction between polysemy and homonymy not always clear

How Many Senses?

- How many senses does the word *interest* have?
  - *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.*
- Are these seven different senses? Four? Three?

Wordnet

- Wordnet, a hierarchical database of senses, defines synsets
- According to Wordnet, *interest* is in 7 synsets
  - Sense 1: *a sense of concern with and curiosity about someone or something*, Synonym: *involvement*
  - Sense 2: *the power of attracting or holding one's interest (because it is unusual or exciting etc.)*, Synonym: *interestingness*
  - Sense 3: *a reason for wanting something done*, Synonym: *sake*
  - Sense 4: *a fixed charge for borrowing money; usually a percentage of the amount borrowed*
  - Sense 5: *a diversion that occupies one's time and thoughts (usually pleasantly)*, Synonyms: *pastime, pursuit*
  - Sense 6: *a right or legal share of something; a financial involvement with something*, Synonym: *stake*
  - Sense 7: *(usually plural) a social group whose members control some field of activity and who have common aims*, Synonym: *interest group*

Sense and Translation

- Most relevant for machine translation:
  different translations $\rightarrow$ different sense

Sense and Translation

- Most relevant for machine translation:
  different translations → different sense

- Example *interest* translated into German
  - *Zins*: financial charge paid for load (Wordnet sense 4)
  - *Anteil*: stake in a company (Wordnet sense 6)
  - *Interesse*: all other senses

Languages Differ

- Foreign language may make finer distinctions

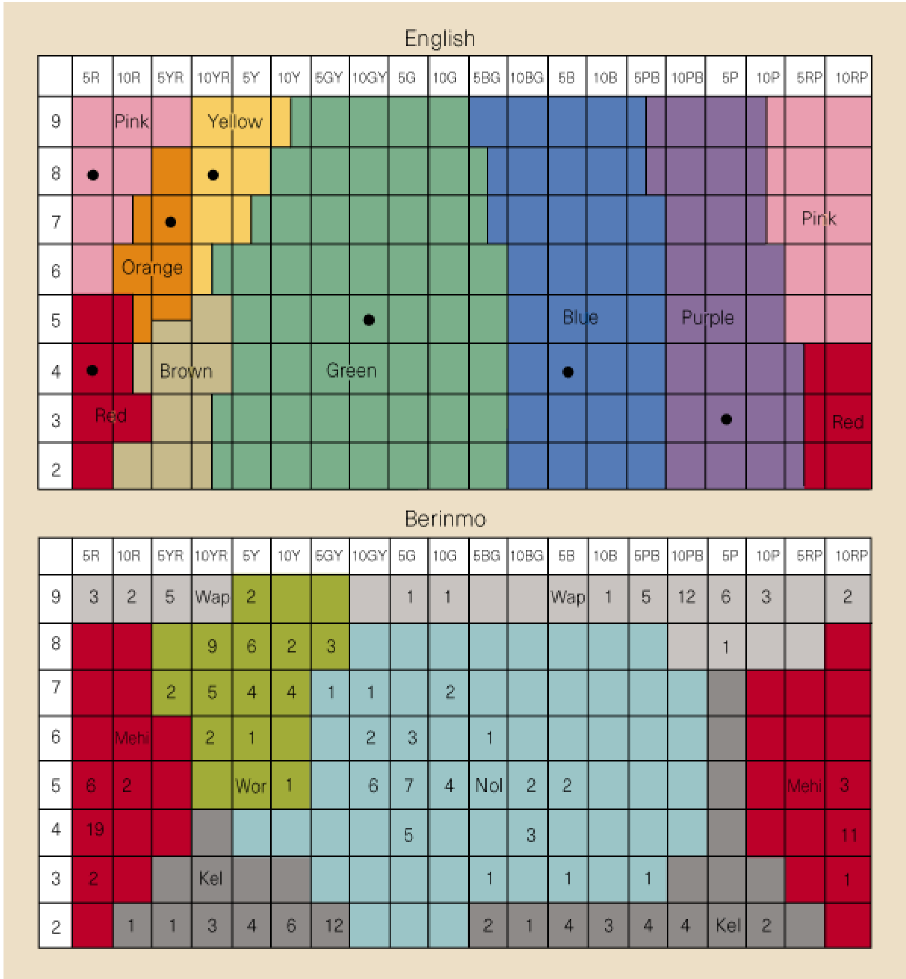
- Translations of *river* into French
  - *fleuve*: river that flows into the sea
  - *rivière*: smaller river

Languages Differ

- Foreign language may make finer distinctions
- Translations of *river* into French
  - *fleuve*: river that flows into the sea
  - *rivière*: smaller river
- English may make finer distinctions than a foreign language
- Translations of German *Sicherheit* into English
  - *security*
  - *safety*
  - *confidence*

Overlapping Senses

- Color names may differ between languages
- Many languages have one word for blue and green
- Japanese: *ao*
  change early 20th century:
  midori (*green*) and *ao* (*blue*)
- But still:
  - vegetables are *greens* in English, *ao-mono* (blue things) in Japanese
  - "go" traffic light is *ao* (blue)



Color names in English and Berinomo (Papua New Guinea)

One Last Word on Senses

- Lot of research in word sense disambiguation is focused on polysemous words with clearly distinct meanings, e.g. *bank*, *plant*, *bat*, ...

One Last Word on Senses

- Lot of research in word sense disambiguation is focused on polysemous words with clearly distinct meanings, e.g. *bank*, *plant*, *bat*, ...

- Often meanings are close and hard to tell apart, e.g. *area*, *field*, *domain*, *part*, *member*, ...
  - *She is a part of the team.*
  - *She is a member of the team.*
  - *The wheel is a part of the car.*
  - * *The wheel is a member of the car.*

subcategorization frames

Verb Subcategorization

- Example

Das	Vorhaben	verwarf	die	Kommission	.
the	*plan*	*rejected*	*the*	*commission*	.

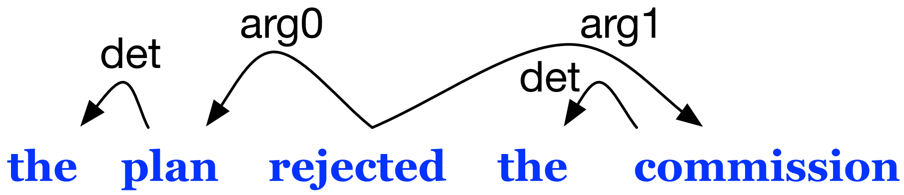
- Propbank

```
Arg0-PAG: rejecter (vnrole:  77-agent)
Arg1-PPT: thing rejected (vnrole:  77-theme)
Arg3-PRD: attribute
```

- Is *plan* a typical `Arg0` of *reject*?

Dependency Parsing

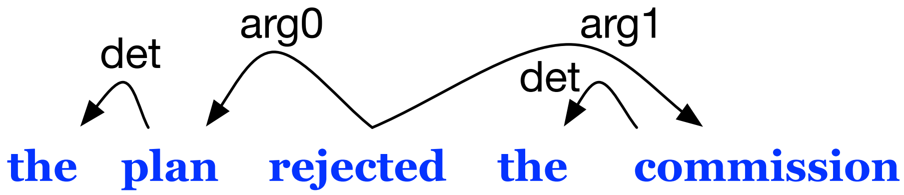
- Dependencies between words



- Can be obtained by
  - dedicated dependency parser
  - CFG grammar with head word rules

Dependency Parsing

- Dependencies between words



- Can be obtained by
  - dedicated dependency parser
  - CFG grammar with head word rules

- Are dependency relations enough?
  - *reject* — subj $\rightarrow$ *plan* $\Rightarrow$ bad
  - *reject* — subj $\rightarrow$ *commission* $\Rightarrow$ good

logical form

First Order Logic

- Classical example

Every farmer has a donkey

- Ambiguous, two readings

First Order Logic

- Classical example

Every farmer has a donkey

- Ambiguous, two readings

- Each farmer as its own donkey

$$\forall\ x\text{: } farmer(x)\ \exists\ y\text{: } donkey(y) \wedge owns(x,y)$$

First Order Logic

- Classical example

Every farmer has a donkey

- Ambiguous, two readings

- Each farmer as its own donkey

$$\forall\ x{:}\ farmer(x)\ \exists\ y{:}\ donkey(y) \wedge owns(x,y)$$

- There is only one donkey

$$\exists\ y{:}\ donkey(y) \wedge \forall\ x{:}\ farmer(x) \wedge owns(x,y)$$

- Does this matter for translation? (typically not)

Logical Form and Inference

- Input sentence

Whenever I visit my uncle and his daughters,
I can't decide who is my favorite cousin.

Logical Form and Inference

- Input sentence

Whenever I visit my uncle and his daughters,
I can't decide who is my favorite cousin.

- Facts from input sentence

$\exists$ *d: female(d)*
$\exists$ *u: father(u,d)*
$\exists$ *i: uncle(u,i)*
$\exists$ *c: cousin(i,c)*

Logical Form and Inference

- Input sentence

Whenever I visit my uncle and his daughters,
I can't decide who is my favorite cousin.

- Facts from input sentence

$$\exists\, d: female(d)$$
$$\exists\, u: father(u,d)$$
$$\exists\, i: uncle(u,i)$$
$$\exists\, c: cousin(i,c)$$

- World knowledge

$$\forall\, i,u,c: uncle(u,i) \land father(u,c) \rightarrow cousin(i,c)$$

Logical Form and Inference

- Input sentence

Whenever I visit my uncle and his daughters,
I can't decide who is my favorite cousin.

- Facts from input sentence

$$\exists\ d\text{: } female(d)$$
$$\exists\ u\text{: } father(u,d)$$
$$\exists\ i\text{: } uncle(u,i)$$
$$\exists\ c\text{: } cousin(i,c)$$

- World knowledge

$$\forall\ i,u,c\text{: } uncle(u,i) \wedge father(u,c) \rightarrow cousin(i,c)$$

- Hypothesis that $c = d$ is consistent with given facts and world knowledge

- Inference

$$female(d) \rightarrow female(c)$$

Scope

- Example (Knight and Langkilde, 2000)

green eggs and ham

  - Only eggs are green

    (green eggs) and ham

  - Both are green

    green (eggs and ham)

Scope

- Example (Knight and Langkilde, 2000)

green eggs and ham

  - Only eggs are green

(green eggs) and ham

  - Both are green

green (eggs and ham)

- Spanish translations
  - Only eggs are green

huevos verdes y jamón

  - Also ambiguous

jamón y huevos verdes

- Machine translation should preserve ambiguity

discourse

Ambiguous Discourse Markers

- Example

Since you brought it up, I do not agree with you.

Since you brought it up, we have been working on it.

- How to translated *since*? Temporal or conditional?

Implicit Discourse Relationships

- English syntactic structure may imply causation

Wanting to go to the other side, the chicken crossed the road.

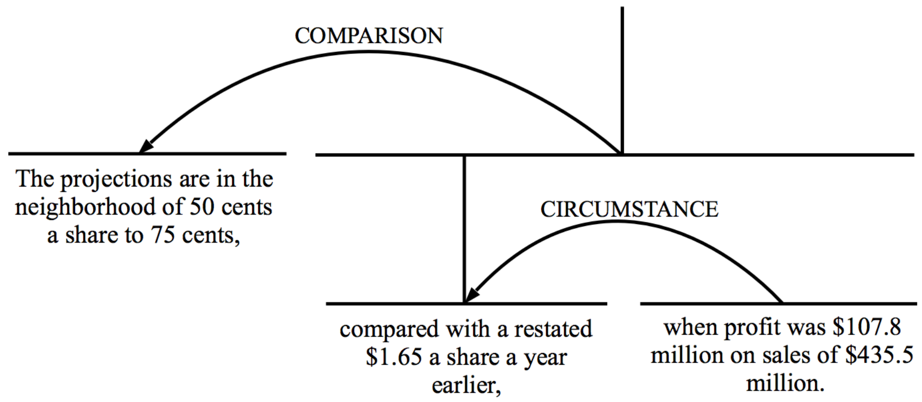
- This discourse relationship may have to made explicit in another language

Discourse Parsing

- Discourse relationships,

  e.g., Circumstance, Antithesis, Concession, Solutionhood, Elaboration, Background, Enablement, Motivation, Condition, Interpretation, Evaluation, Purpose, Evidence, Cause, Restatement, Summary, ...

- Hierarchical structure



- There is a discourse treebank, but inter-annotator agreement is low

abstract meaning representations

Example

He looked at me very gravely , and put his arms around my neck .

```
(a / and
      :op1 (l / look-01
            :ARG0 (h / he)
            :ARG1 (i / i)
            :manner (g / grave
                  :degree (v / very)))
      :op2 (p / put-01
            :ARG0 h
            :ARG1 (a2 / arm
                  :part-of h)
            :ARG2 (a3 / around
                  :op1 (n / neck
                        :part-of i))))
```

Abstracts from Syntax

- Abstract meaning representation

```
(l / look-01
   :ARG0 (h / he)
   :ARG1 (i / i)
   :manner (g / grave
      :degree (v / very)))
```

- Possible English sentences
  - *He looks at me gravely.*
  - *I am looked at by him very gravely.*
  - *He gave me a very grave look.*

Feature Engineering vs Machine Learning

- Engineering approach
  - identify weak points of current system
  - develop changes that address them

Feature Engineering vs Machine Learning

- Engineering approach
  - identify weak points of current system
  - develop changes that address them

- Machine learning
  - deeper models
  - more robust estimation techniques
  - fight over-fitting or under-fitting
  - other adjustments

Feature Engineering vs Machine Learning

- Engineering approach
  - identify weak points of current system
  - develop changes that address them
- Machine learning
  - deeper models
  - more robust estimation techniques
  - fight over-fitting or under-fitting
  - other adjustments
- Difficult to analyze neural models $\rightarrow$ engineering hard to do