
Syntax and Semantics

Philipp Koehn

25 October 2018



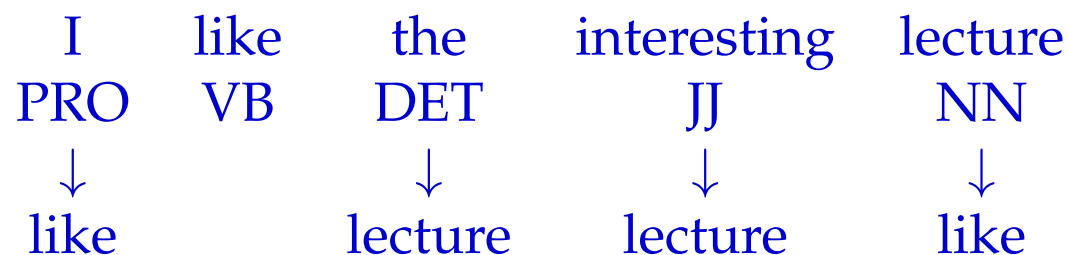
syntax

- 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

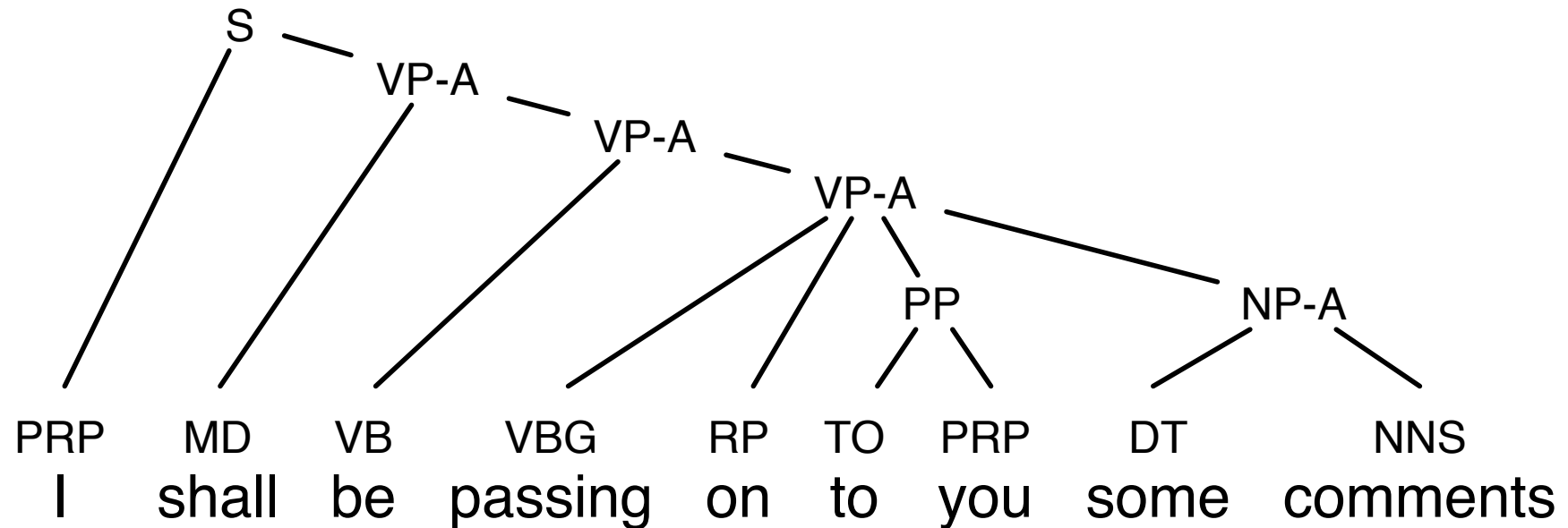
Dependency Structure



- 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
 - noun phrases: *the big man, a house, ...*
 - prepositional phrases: *at 5 o'clock, in Edinburgh, ...*
 - 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

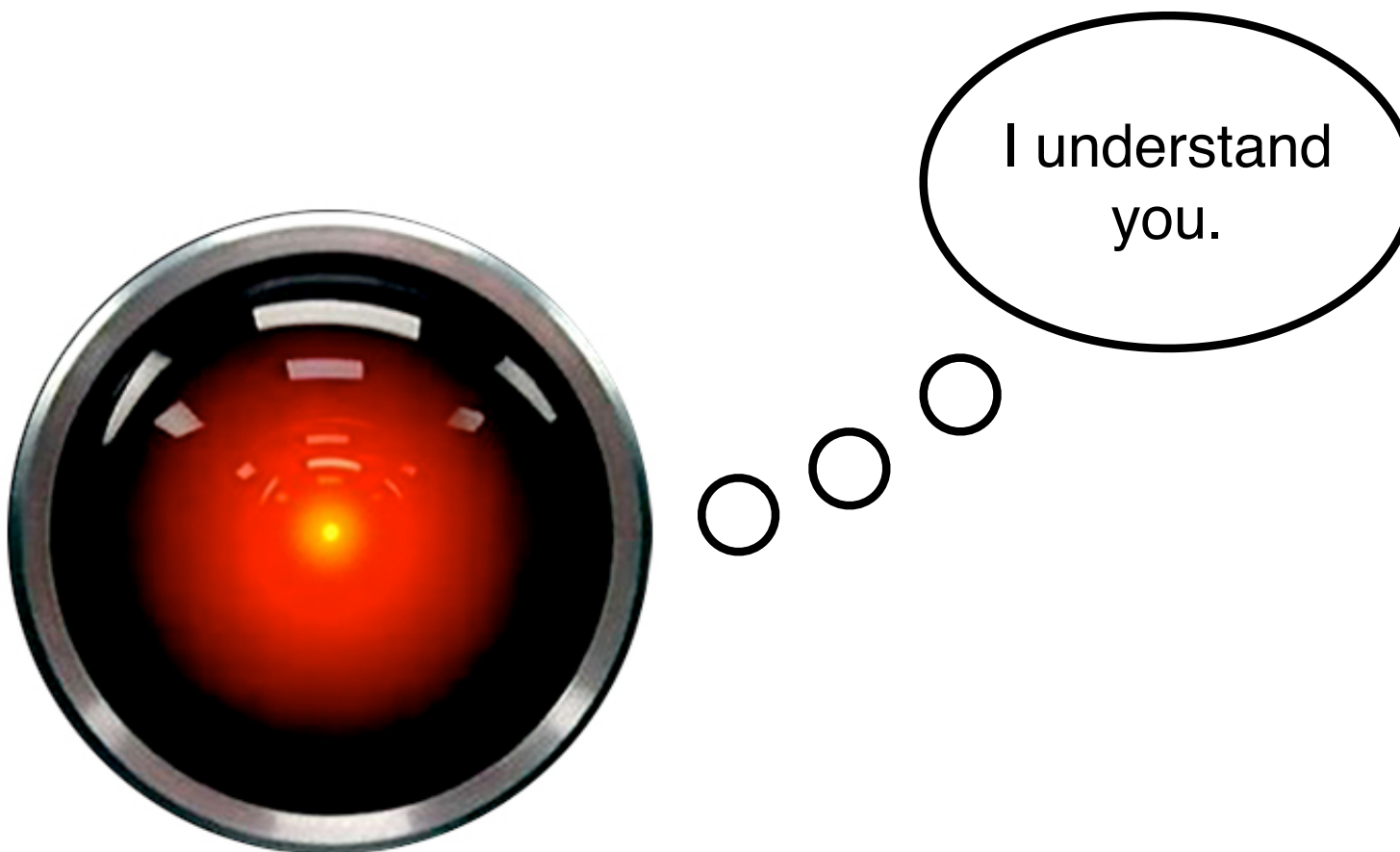


Phrase structure grammar tree for an English sentence
(as produced Collins' parser)

semantics

- 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



knowledge



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

- 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

- 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, 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*

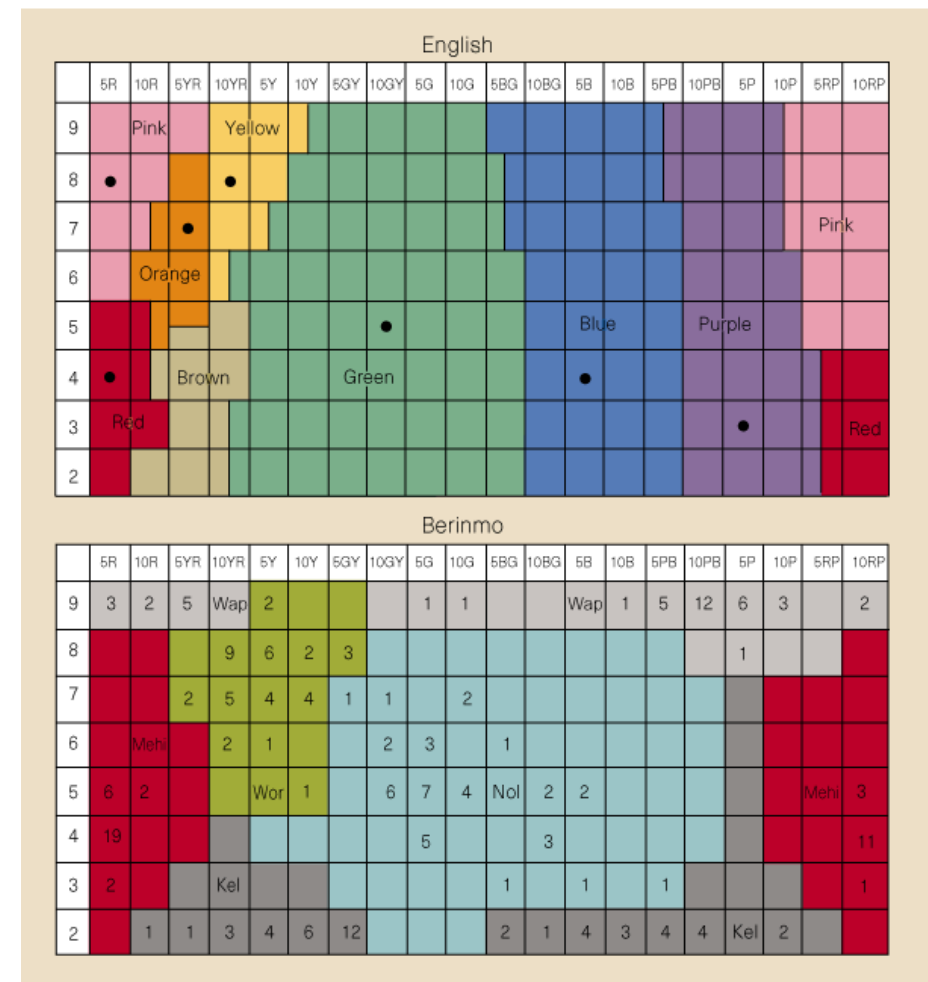
- Most relevant for machine translation:
different translations → different sense
- Example *interest* translated into German
 - *Zins*: financial charge paid for loan (Wordnet sense 4)
 - *Anteil*: stake in a company (Wordnet sense 6)
 - *Interesse*: all other senses

Languages Differ

- Foreign language may make finer distinctions
- Translations of *river* into French
 - *fleuve*: river that flows into the sea
 - *rivière*: smaller river
- 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)



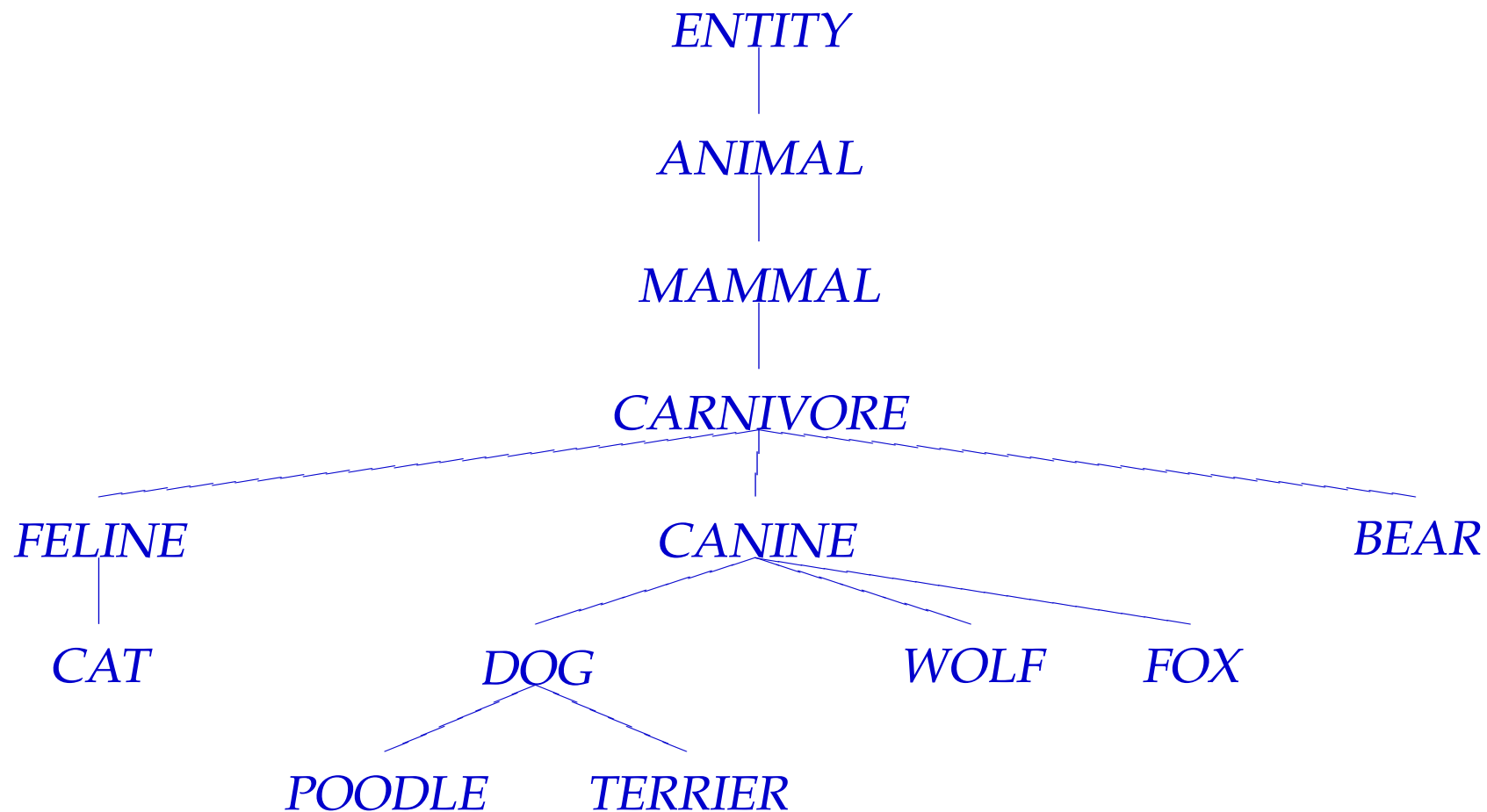
Color names in English and Berinomo (Papua New Guinea)

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*, ...■
- 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.*

Ontology

22



Representing Meaning

- The meaning of *dog* is *DOG* or *dog*(x)
Not much gained here■
- Words that have similar meaning should have similar representations■
- Composition of meaning

$\text{meaning}(\textit{daughter}) = \text{meaning}(\textit{child}) + \text{meaning}(\textit{female})$ ■

- Analogy

$\text{meaning}(\textit{king}) + \text{meaning}(\textit{woman}) - \text{meaning}(\textit{man}) = \text{meaning}(\textit{queen})$

- Contexts may be represented by a vector of word counts

Example:

*Then he grabbed his new mitt and **bat**, and headed back to the dugout for another turn at **bat**. Hulet isn't your average baseball player. "It might have been doctoring up a **bat**, grooving a **bat** with pennies or putting a little pine tar on the baseball. All the players were sitting around the dugout laughing at me."*

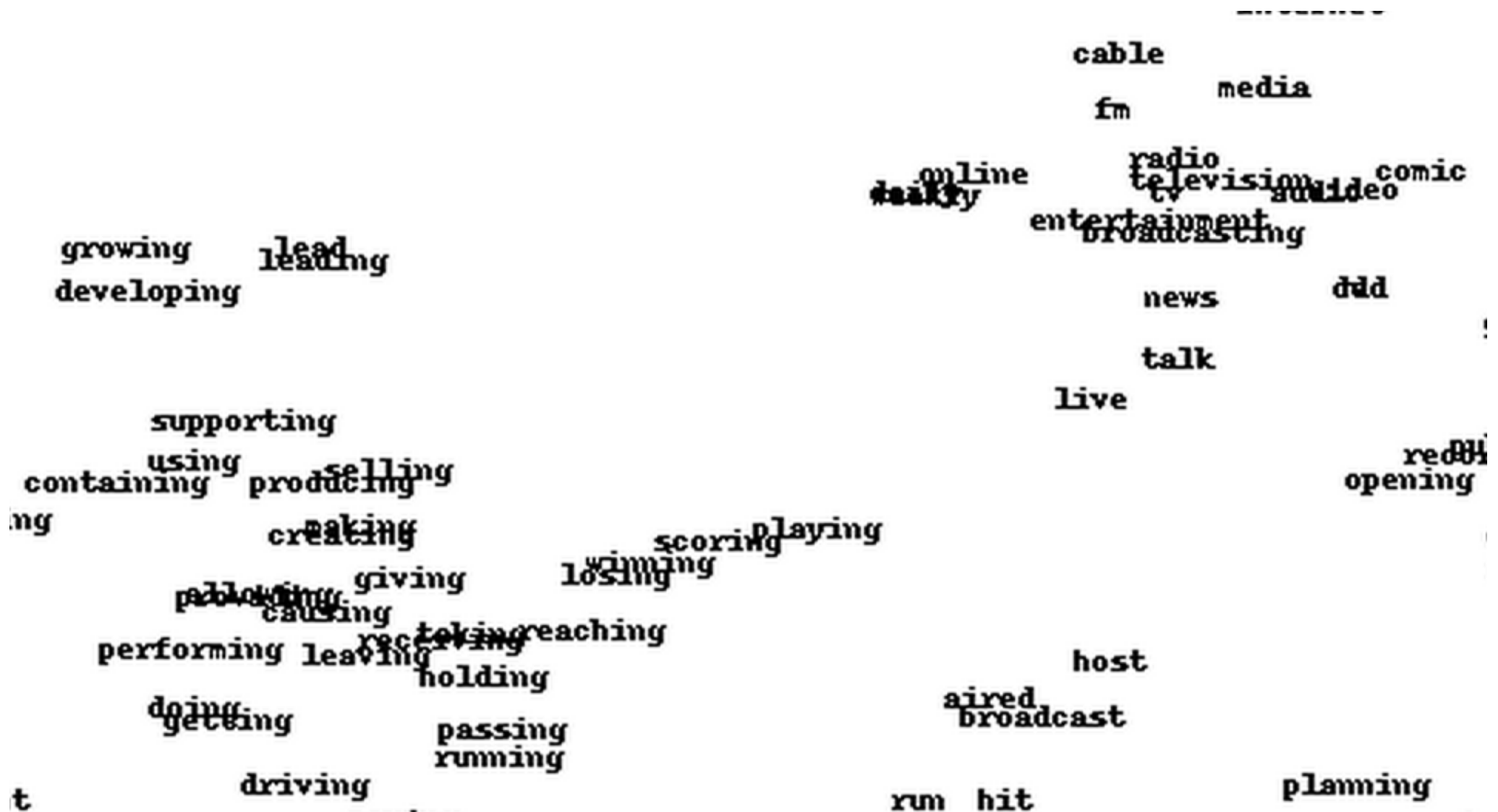
The word counts normalized, so all the vector components add up to one.

grabbed	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
mitt	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
headed	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
dugout	$\begin{pmatrix} 2 \end{pmatrix}$	$\begin{pmatrix} 0.10 \end{pmatrix}$
turn	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
average	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
baseball	$\begin{pmatrix} 2 \end{pmatrix}$	$\begin{pmatrix} 0.10 \end{pmatrix}$
player	$\begin{pmatrix} 2 \end{pmatrix}$	$\begin{pmatrix} 0.10 \end{pmatrix}$
doctoring	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
grooving	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
pennies	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
pine	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
tar	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
sitting	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$
laughing	$\begin{pmatrix} 1 \end{pmatrix}$	$\begin{pmatrix} 0.05 \end{pmatrix}$

- Average over all occurrences of word
- Context may also just focus on directly neighboring words

Word Embeddings

25



Word Sense Disambiguation

- For many applications, we would like to disambiguate senses
- Supervised learning problem *plant* → *PLANT-FACTORY*
- Features
 - Directly neighboring words
 - * *plant life*
 - * *manufacturing plant*
 - * *assembly plant*
 - * *plant closure*
 - * *plant species*
 - Any content words in a 50 word window
 - Syntactically related words
 - Syntactic role in sense
 - Topic of the text
 - Part-of-speech tag, surrounding part-of-speech tags

subcategorization frames

- Example

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

- Propbank

Arg0-PAG: rejecter (vnrole: 77-agent)

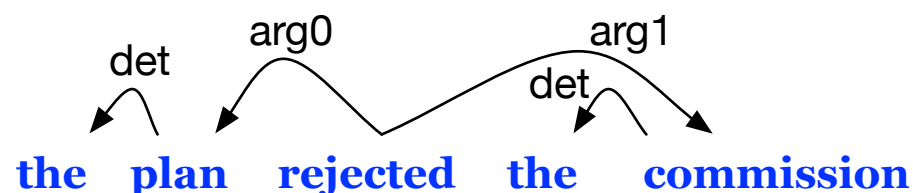
Arg1-PPT: thing rejected (vnrole: 77-theme)

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
- Are dependency relations enough?
 - *reject* — subj → *plan* ⇒ bad
 - *reject* — subj → *commission* ⇒ good

logical form

- Classical example

Every farmer has a donkey

- Ambiguous, two readings
- Each farmer as its own donkey

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

- There is only one donkey

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

- Does this matter for translation? (typically not)

- Input sentence

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

- Facts from input sentence

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

- World knowledge

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

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

- Inference

$\text{female}(d) \rightarrow \text{female}(c)$

- 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?

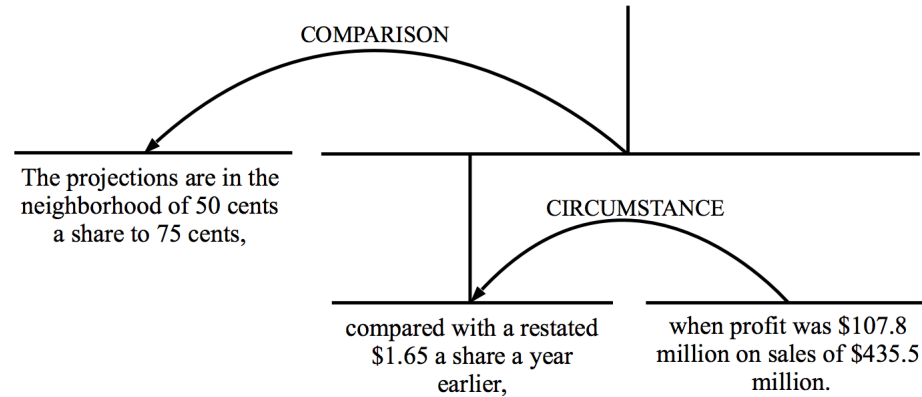
Implicit Discourse Relationships

- English syntactic structure may imply causation

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

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

- 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))))
```


- Abstract meaning representation

```
(1 / 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.*



adding linguistic annotation

Adding Linguistic Annotation

- Improving neural models with linguistic information
 - linguistic annotation to the input sentence
 - linguistic annotation to the output sentence,
 - build linguistically structured models.

- Neural models good with rich context
 - prediction conditioned on entire input and all previously output words
 - good at generalizing and draw from relevant knowledge
- Adding more information to conditioning context straightforward
- Relevant linguistic information
 - part-of-speech tags
 - lemmas
 - morphological properties of words
 - syntactic phrase structure
 - syntactic dependencies
 - semantics

Enriched Input

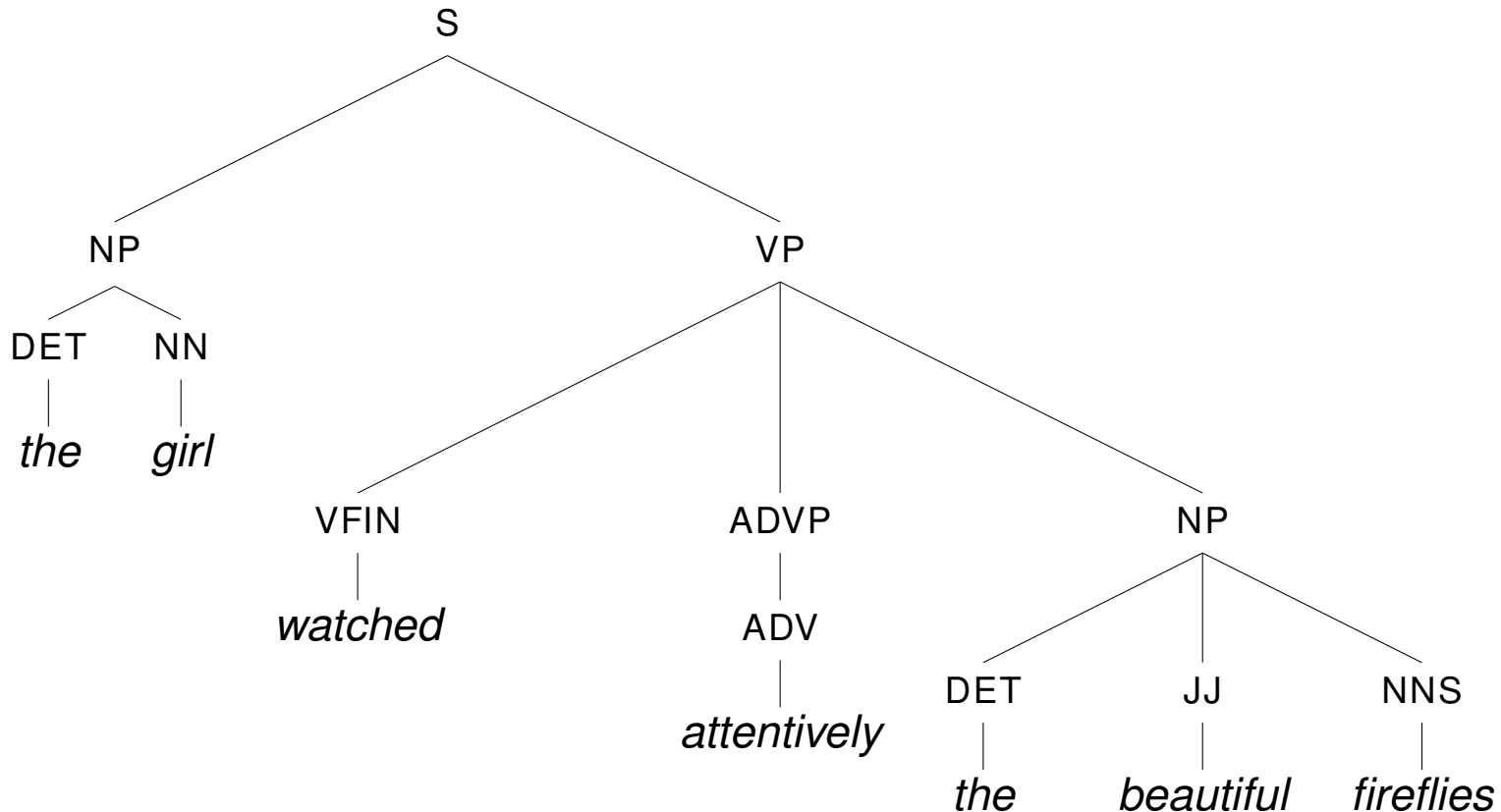
Words	<i>the</i>	<i>girl</i>	<i>watched</i>	<i>attentively</i>	<i>the</i>	<i>beautiful</i>	<i>fireflies</i>
Part of speech	DET	NN	VFIN	ADV	DET	JJ	NNS
Lemma	<i>the</i>	<i>girl</i>	<i>watch</i>	<i>attentive</i>	<i>the</i>	<i>beautiful</i>	<i>firefly</i>
Morphology	-	SING.	PAST	-	-	-	PLURAL
Noun phrase	BEGIN	CONT	OTHER	OTHER	BEGIN	CONT	CONT
Verb phrase	OTHER	OTHER	BEGIN	CONT	CONT	CONT	CONT
Synt. dependency	<i>girl</i>	<i>watched</i>	-	<i>watched</i>	<i>fireflies</i>	<i>fireflies</i>	<i>watched</i>
Depend. relation	DET	SUBJ	-	ADV	DET	ADJ	OBJ
Semantic role	-	ACTOR	-	MANNER	-	MOD	PATIENT
Semantic type	-	HUMAN	VIEW	-	-	-	ANIMATE

- Each property encoded as 1-hot vector
- Note: phrasal annotation: BEGIN, CONTINUE, OTHER
- Can all this be discovered by machine learning instead?

Linguistic Annotation of Output

- Same annotation also be used for output words
- May support more syntactically or semantically coherent output
- Most successful in statistical machine translation: output syntax
 - represented as syntactic tree structures
 - need to convert into sequence

Linguistic Annotation of the Output

Sentence	<i>the girl watched attentively the beautiful fireflies</i>
Syntax tree	 <pre> graph TD S --> NP1[NP] S --> VP[VP] NP1 --> DET1[DET] NP1 --> NN1[NN] DET1 --> the1[the] NN1 --> girl[girl] VP --> VFIN[VFIN] VP --> ADVP[ADVP] VP --> NP2[NP] VFIN --> watched[watched] ADVP --> ADV[ADV] ADV --> attentively[attentively] NP2 --> DET2[DET] NP2 --> JJ[JJ] NP2 --> NNS[NNS] DET2 --> the2[the] JJ --> beautiful[beautiful] NNS --> fireflies[fireflies] </pre>
Linearized	(S (NP (DET <i>the</i>) (NN <i>girl</i>)) (VP (VFIN <i>watched</i>) (ADVP (ADV <i>attentively</i>)) (NP (DET <i>the</i>) (JJ <i>beautiful</i>) (NNS <i>fireflies</i>))))

- Syntactic parsing now also handled by deep learning
- More complex models to build output structure
 - related on left-to-right push-down automata
 - need to maintain stack of opened phrases
 - each step starts, extends, or closes a phrase
- Early work on integrating machine translation and syntactic parsing

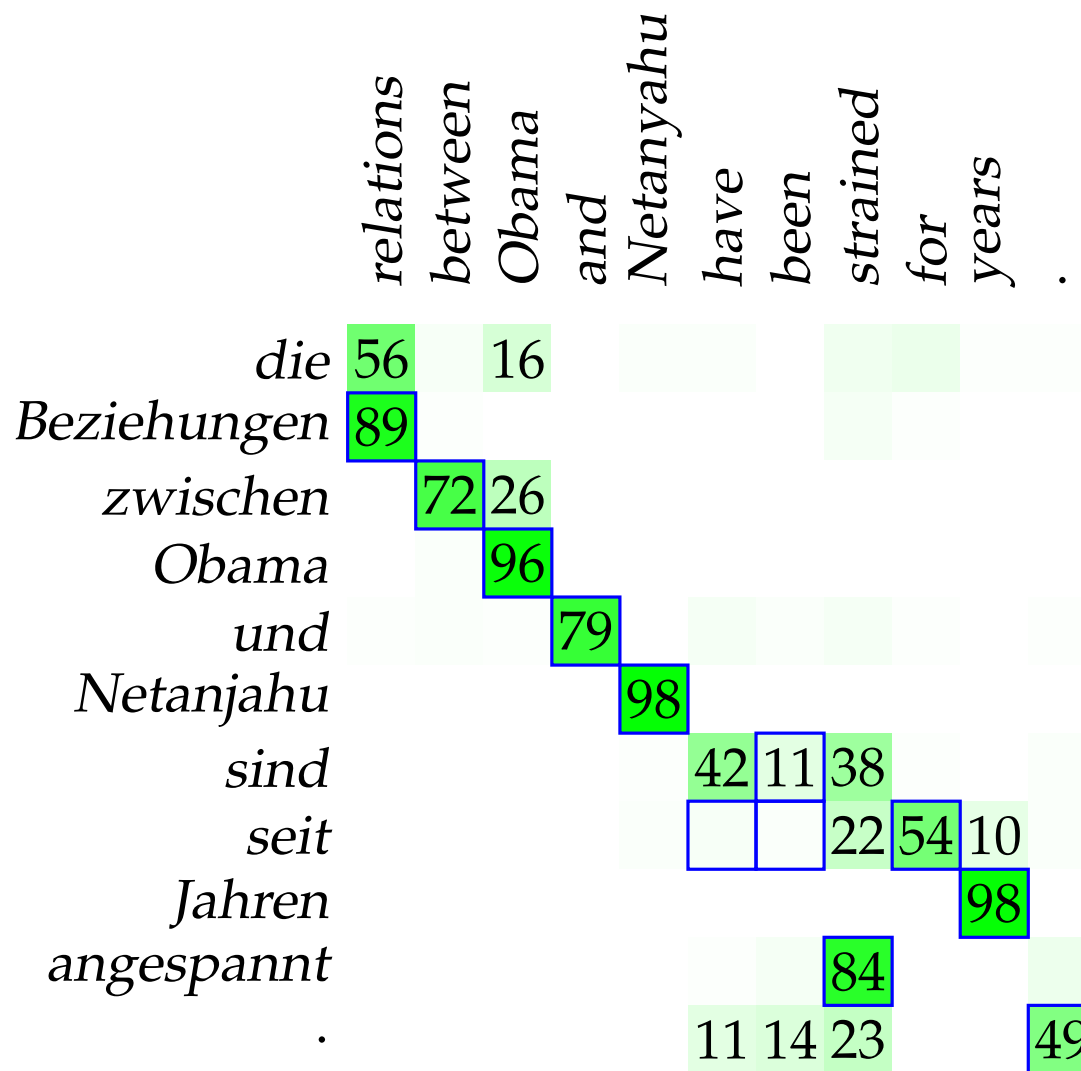
guided alignment training

- Attention mechanism motivated by linguistic fact that each individual output word is often fully explained by a single input word
- Support training with externally generated word alignments
 - generate word alignment with IBM Models
 - bias attention to these alignments
- Added cost function
 - alignment matrix A
 - alignment points A_{ij} between input word j and output word i
 - attention weight of neural model α_{ij}

$$\text{cost}_{\text{MSE}} = -\frac{1}{I} \sum_{i=1}^I \sum_{j=1}^J (A_{ij} - \alpha_{ij})^2$$

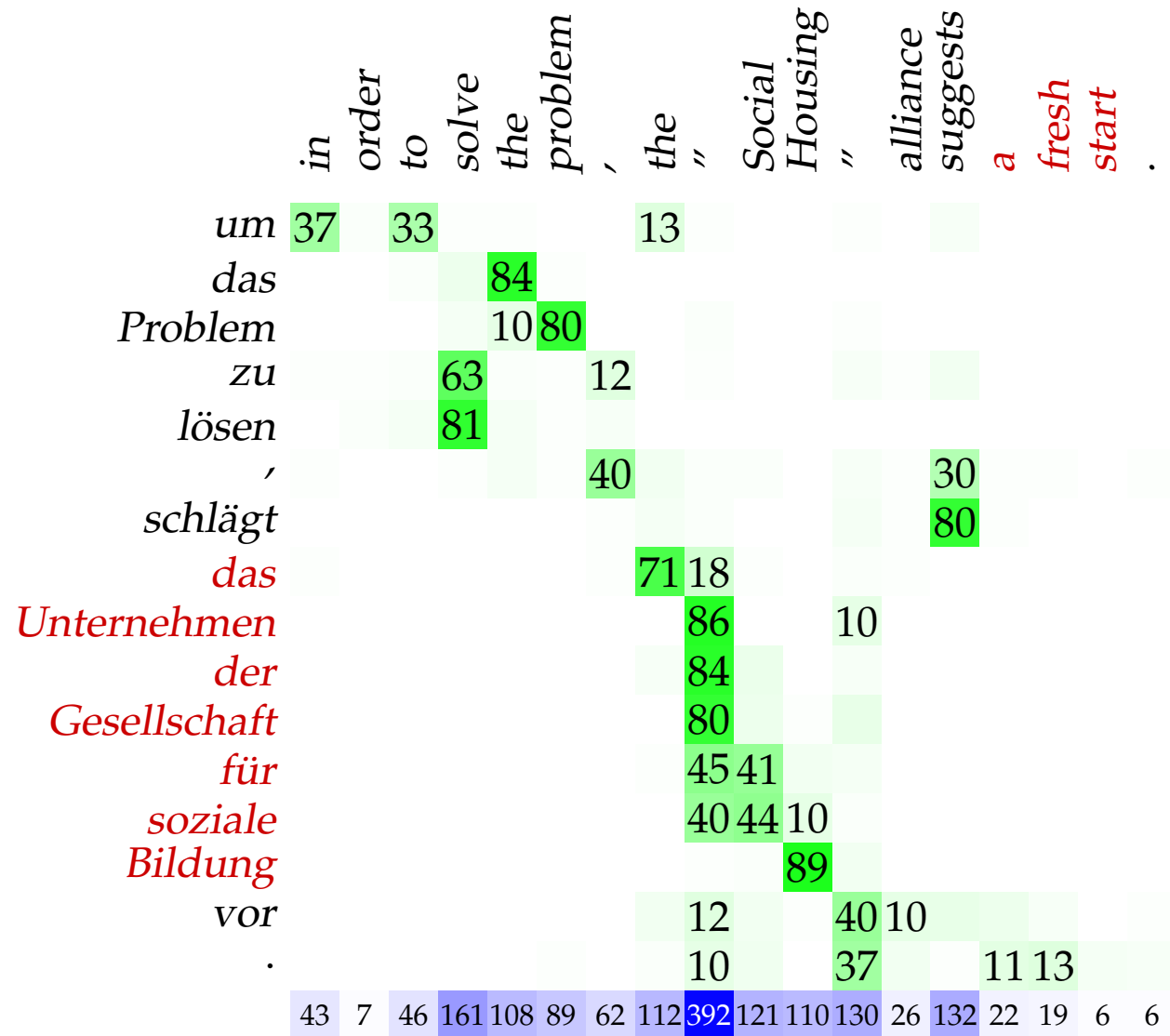
- Word alignment useful by-product of translation

Attention vs. Alignment



modelling coverage

Overgeneration and Undergeneration



Modeling Coverage

- Neural models generally very good at translating all input words
- But: no explicit coverage model, sometimes fails■
- Enforce coverage during decoding
- Integrate coverage model

- Track coverage during decoding

$$\text{coverage}(j) = \sum_i \alpha_{i,j}$$

$$\text{over-generation} = \max\left(0, \sum_j \text{coverage}(j) - 1\right)$$

$$\text{under-generation} = \min\left(1, \sum_j \text{coverage}(j)\right)$$

- Add additional penalty functions to score hypotheses

- Extend translation model
- Use vector that accumulates coverage of input words to inform attention
 - raw attention score $a(s_{i-1}, h_j)$
 - informed by previous decoder state s_{i-1} and input word h_j ■
 - add conditioning on $\text{coverage}(j)$ ■

$$a(s_{i-1}, h_j) = W^a s_{i-1} + U^a h_j + V^a \text{coverage}(j) + b^a \blacksquare$$

- Coverage tracking may also be integrated into the training objective.

$$\log \sum_i P(y_i|x) + \lambda \sum_j (1 - \text{coverage}(j))^2$$



- 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 → engineering hard to do