# CMPE 409 Machine Translation
## Words & Sentence

Murat ORHUN

Istanbul Bilgi University

March 23, 2022

## Lexical

The meaning of a word

- English
    - band (music group)
    - band (material)
    - band (wavelength)
    - House of Windsor

- Turkish
    - savaş( war)
    - Savaş (name of a perseon)
    - at (horse, name)
    - at (throw, verb)
    - Çay ocağı
    - Türk Ocakları

## Polysem & Homonymy &Word Senses)

- Polysem: Word may have different meanings: get, bank, book,interest
- Homonymy: words that are completely unrelated, but are spelled the same way: can, beans, ring, right,
- Word Senses: Words with different meanings (Polysems)
- interest: has meanings indicating curiosity (*interest in football*), a stake (*a 5% interest in Google*), or the fee paid for a loan (*interest rate of 4.9%*).
- Hom many senses?
- *the national interest*

# Meas of interst-1

## Noun

- S: (n) **interest**, involvement (a sense of concern with and curiosity about someone or something) *"an interest in music"*
- S: (n) sake, **interest** (a reason for wanting something done) *"for your sake"; "died for the sake of his country"; "in the interest of safety"; "in the common interest"*
- S: (n) **interest**, interestingness (the power of attracting or holding one's attention (because it is unusual or exciting etc.)) *"they said nothing of great interest"; "primary colors can add interest to a room"*
- S: (n) **interest** (a fixed charge for borrowing money; usually a percentage of the amount borrowed) *"how much interest do you pay on your mortgage?"*
- S: (n) **interest**, stake ((law) a right or legal share of something; a financial involvement with something) *"they have interests all over the world"; "a stake in the company's future"*
- S: (n) **interest**, interest group ((usually plural) a social group whose members control some field of activity and who have common aims) *"the iron interests stepped up production"*
- S: (n) pastime, **interest**, pursuit (a diversion that occupies one's time and thoughts (usually pleasantly)) *"sailing is her favorite pastime"; "his main pastime is gambling"; "he counts reading among his interests"; "they criticized the boy for his limited pursuits"*

## Meas of interst-2

Verb

- <u>S:</u> (v) **interest** (excite the curiosity of; engage the interest of)
- <u>S:</u> (v) <u>concern</u>, **interest**, <u>occupy</u>, <u>worry</u> (be on the mind of) *"I worry about the second Germanic consonant shift"*
- <u>S:</u> (v) <u>matter to</u>, **interest** (be of importance or consequence) *"This matters to me!"*

# English to Turkish Translation

How do you traslate "interst" to Turkish?

- ilgi
- faiz
- merak
- alaka
- ...

https://www.seslisozluk.net/interest-nedir-ne-demek/

## English to German Translation

How do you traslate "interst" to German?

- Interesse (curiosity sense)
- Anteil (stake sense)
- Zins (money sense)

## Wordnet

WordNet®️ is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. https://wordnet.princeton.edu/

# Wordnet

About 50 diffent means of "fall"

## WordNet Search - 3.1
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for: | fall | | Search WordNet |

Display Options: | (Select option to change) ▾ | | Change |

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

### Noun

- **S:** (n) **fall**, autumn (the season when the leaves fall from the trees) *"in the fall of 1973"*
- **S:** (n) spill, tumble, **fall** (a sudden drop from an upright position) *"he had a nasty spill on the ice"*
- **S:** (n) **Fall** (the lapse of mankind into sinfulness because of the sin of Adam and Eve) *"women have been blamed ever since the Fall"*

# Wordnet

- <u>S:</u> (v) **fall** (descend in free fall under the influence of gravity) *"The branch fell from the tree"; "The unfortunate hiker fell into a crevasse"*
- <u>S:</u> (v) <u>descend</u>, **fall**, <u>go down</u>, <u>come down</u> (move downward and lower, but not necessarily all the way) *"The temperature is going down"; "The barometer is falling"; "The curtain fell on the diva"; "Her hand went up and then fell again"*
- <u>S:</u> (v) **fall** (pass suddenly and passively into a state of body or mind) *"fall into a trap"; "She fell ill"; "They fell out of favor"; "Fall in love"; "fall asleep"; "fall prey to an imposter"; "fall into a strange way of thinking"; "she fell to pieces after she lost her work"*
- <u>S:</u> (v) **fall**, <u>come</u> (come under, be classified or included) *"fall into a category"; "This comes under a new heading"*
- <u>S:</u> (v) <u>precipitate</u>, <u>come down</u>, **fall** (fall from clouds) *"rain, snow and sleet were falling"; "Vesuvius precipitated its fiery, destructive rage on Herculaneum"*
- <u>S:</u> (v) **fall** (suffer defeat, failure, or ruin) *"We must stand or fall"; "fall by the wayside"*
- <u>S:</u> (v) **fall** (die, as in battle or in a hunt) *"Many soldiers fell at Verdun"; "Several deer have fallen to the same gun"; "The shooting victim fell dead"*
- <u>S:</u> (v) **fall**, <u>shine</u>, <u>strike</u> (touch or seem as if touching visually or audibly) *"Light fell on her face"; "The sun shone on the fields"; "The light struck the golden necklace"; "A strange sound struck my ears"*

# Word Sense Disambiguation

The task of determining the right word sense for a word in a given context is called word sense disambiguation

- Word context
- Neighboring words
- Content words
- We will discuss more again...

# Word Sense Disambiguation(WSD)

- In computational linguistics, word-sense disambiguation (WSD) is an open problem concerned with identifying which sense of a word is used in a sentence.
- *I am going to* bank *today.*
- *Savaş artık bitti (Turkish)*

## Hypernym Relationship

- Word senses, may be organized according to their relationship to each other.
- For instance: A *dog* is a *mammal*, which is an *animal*.
- This is: **is - a**, or **hypernym** relation
- **hypernyms**: abstract (parent)
- **hyponyms**: specific (child)

# Hypernym Relationship in Wordnet
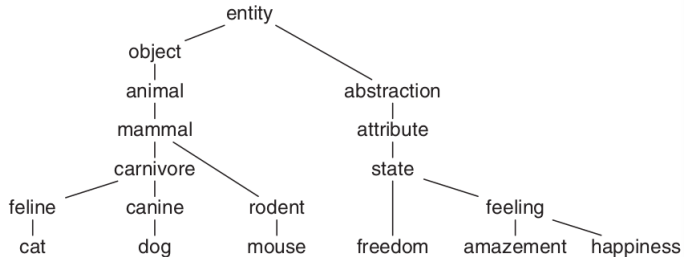
## Part of Relationship

- a *wheel* is part of a *car*
- a *wolf* is part of a *pack*
- *trunk,crown* are parts of a *tree*
- A collection of *trees* forms a *forest*
- **meronym** (components, part-of)
- **holonyms** (contationed, has-a)

# Wordnet

- Turkish
- Swedish
- Chinese
- and More

# Online Dictionary-Sesli Sözklük



fall 🔊 UK🔊 US🔊 AU🔊 ☆

İngilizce - Türkçe

**fall** teriminin İngilizce Türkçe sözlükte anlamı

1. sonbahar
   - *Sonbaharda Paris'e gidiyorum.* - *I'm going to Paris in the fall.*
   - *Gariptir, bu sonbaharda çok sayıda pembe dizi izliyorum.* - *Oddly, I've been watching a lot of telenovelas this fall.*
2. düşüş
   - *O, düşüşten sonra topal oldu.* - *He was lame after the fall.*
   - *Gurur bir düşüşten önce gider.* - *Pride goes before a fall.*
3. düşmek
   - *Ağaç düşmek üzereydi.* - *The tree was ready to fall down.*
   - *Buz üzerinde düşmek incitir.* - *Falling on ice hurts.*
4. {i} düşme
   - *O, bir düşmede yaralandı.* - *She was injured in a fall.*
   - *Kötü alışkanlıklara düşmek kolaydır.* - *It's easy to fall into bad habits.*
5. {i} dökülme
   - *Saçınız dökülmeye başlayacaktır.* - *Your hair will start to fall out.*
6. {f} düş

https://www.seslisozluk.net

# Online Dictionary-LEO (Link Everything Online)

LEO GmbH is a leading provider of online dictionaries for multiple language pairs.

| Substantive | | |
|---|---|---|
| 🖫 ⓘ ⏵ **fall** | 📖 ⏵ der **Fall**  *Pl.: die Fälle* | |
| 🖫 ⓘ ⏵ **fall** | 📖 ⏵ der **Sturz**  *Pl.: die Stürze* | |
| 🖫 ⓘ ⏵ **fall** *(Amer.)* | 📖 ⏵ der **Herbst**  *Pl.: die Herbste* | |
| 🖫 ⓘ ⏵ **fall** | 📖 ⏵ der **Rückgang**  *Pl.: die Rückgänge* | |
| 🖫 ⓘ ⏵ **case**  *auch*  [JURA][LING.][MED.] | 📖 ⏵ der **Fall**  *Pl.: die Fälle* | |
| 🖫 ⓘ ⏵ **drop** | 📖 ⏵ der **Fall**  *Pl.: die Fälle* | |
| 🖫 ⓘ ⏵ **fall** *Sg., meist im Plural:* falls  - waterfall | 📖 ⏵ der **Wasserfall**  *Pl.: die Wasserfälle* | |
| 🖫 ⓘ ⏵ **fall** | 📖 ⏵ der **Absturz**  *Pl.: die Abstürze* | |
| 🖫 ⓘ ⏵ **fall** | 📖 ⏵ das **Gefälle**  *Pl.: die Gefälle* | |
| 🖫 ⓘ ⏵ **fall** | 📖 ⏵ die **Abnahme**  *Pl.: die Abnahmen* | |
| 🖫 ⓘ ⏵ **fall** | 📖 ⏵ das **Druckgefälle**  *Pl.: die Druckgefälle* | |
| 🖫 ⓘ ⏵ **fall** | 📖 ⏵ die **Fallhöhe**  *Pl.: die Fallhöhen* | |
| | ⏵ die **Neigung**  *Pl.: die Neigungen* | |
| Weitere Substantive anzeigen *(13 / 303)* ⌄ | der **Niedergang**  *kein Pl.* | |

https://www.leo.org

# Online Dictionary-LEO (Link Everything Online)



https://www.leo.org

Words
**Sentences**
Corpora
References

**Sentence structure**
Grammar
Translation of Sentence Structure
Discussion

## Components of Sentence

- A sentence is a textual unit consisting of one or more words that are grammatically linked.

- Verb is the central element of a sentence
  Example:

  ```
  Jane bought the house
  Ben kitap aldım (Turkish)
  ```

- Verb, Subject, Object

## Components of Sentence

- Example:

        Jane swims    (no object)
        Jane gave Joe the book (two or more...)

- Valency: Objects required by a verb

Words
Sentences
Corpora
References

**Sentence structure**
Grammar
Translation of Sentence Structure
Discussion

## Components of Sentence

- arguments: Objects required by a verb
- adjuncts: Additional information
- Example:

  ```
  Jane bought the house.
  Jane bought the house from Jim.
  Jane gave Joe the book.
  Jane bought the beautiful house.
  Jane bought the beautiful house in the city center
  ```

- arguments and adjuncts is not always straight- forward.

Words
**Sentences**
Corpora
References

**Sentence structure**
Grammar
Translation of Sentence Structure
Discussion

## Components of Sentence

- See extensions
- Example:

  the house
  the beautiful house
  the very beautiful house
  the houese in the city center (prepostion)
  the house in the city center next ot super market

- Clauses: consists of a verb with arguments and adjuncts

Words
Sentences
Corpora
References

**Sentence structure**
Grammar
Translation of Sentence Structure
Discussion

## Components of Sentence

- Sentences can include Recursion
- The recursive expansion of sentences into nested construction causes a lot of problems
- Example:

    Joe eats steak with a knife.
    Jim eats steak with ketchup.
    Jane watches the man with the telescope. (discuss)

Words
**Sentences**
Corpora
References

**Sentence structure**
Grammar
Translation of Sentence Structure
Discussion

## Structural Ambiguity

- Prepositional phrase
- Connectives
- Example:

  ```
  Jim washes the dishes and watches televi-
  sion with Jane (discuss)
  ```

Words
Sentences
Corpora
References

Sentence structure
Grammar
Translation of Sentence Structure
Discussion

## Solving Ambiguity

- Ambiguity is often resolved semantically.
- Human are not confused
- Computers?
- Does a computers taste a steak?

Words     Sentence structure
**Sentences**     **Grammar**
Corpora     Translation of Sentence Structure
References     Discussion

## Grammar

- In formal language theory, a grammar (when the context is not given, often called a formal grammar for clarity) describes how to form strings from a language's alphabet that are valid according to the language's syntax

- A grammar does not describe the meaning of the strings or what can be done with them in whatever context—only their form.

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

Words    Sentence structure
**Sentences**    **Grammar**
Corpora    Translation of Sentence Structure
References    Discussion

## Treebank

- In linguistics, a treebank is a parsed text corpus that annotates syntactic or semantic sentence structure
- The construction of parsed corpora in the early 1990s revolutionized computational linguistics, which benefitted from large-scale empirical data
- **The Penn Treebank**: annotates naturally-occuring text for linguistic structure, one million words have been annotated.

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

Words
**Sentences**
Corpora
References

Sentence structure
**Grammar**
Translation of Sentence Structure
Discussion
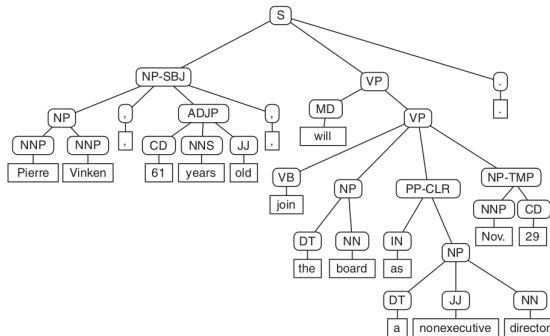
## The Penn Treebank



Figure: Parse tree from the Penn tree bank for the sentence: *Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29*

Words
**Sentences**
Corpora
References

Sentence structure
**Grammar**
Translation of Sentence Structure
Discussion

# Treebank



a. The house at the end of the street is red.

**Constituency structure**

b. The house at the end of the street is red.

**Dependency structure**

Words
**Sentences**
Corpora
References

Sentence structure
**Grammar**
Translation of Sentence Structure
Discussion

## The Penn Treebank

- **The Penn Treebank**: annotates naturally-occuring text for linguistic structure
- one million words have been annotated.

Words
**Sentences**
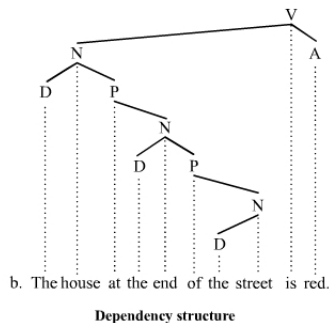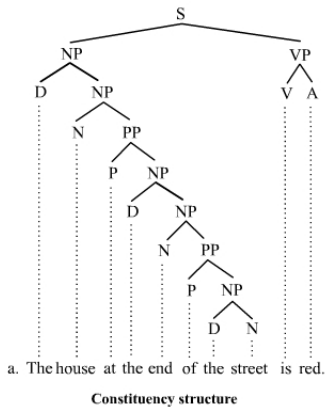Corpora
References

Sentence structure
**Grammar**
Translation of Sentence Structure
Discussion

## Phrase structure grammar

**Phrases**: provides the basis for talking about levels in the parse tree between the sentence node on top and the parts of speech and words at the bottom.

- noun phrases
- prepositional phrases
- verb phrases
- adjective phrases

Words
**Sentences**
Corpora
References

Sentence structure
**Grammar**
Translation of Sentence Structure
Discussion

## Context Free Grammar (CFG)

- Nonterminal: part-of-speech tags and phrase categories
- Terminal symbols (words)
- rules

Words
Sentences
Corpora
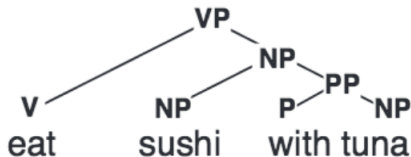References

Sentence structure
Grammar
Translation of Sentence Structure
Discussion

## English Grammar

- S → NP VP

- NP → NNP ADJP

- NP → NNP NNP

- VP → VB NP PP NP

- VB → *join*

Words
Sentences
Corpora
References

Sentence structure
Grammar
Translation of Sentence Structure
Discussion

# English Grammar

- DT → {the, a}
- N → {ball, garden, house, sushi}
- P → {in, behind, with}
- NP → DT N
- NP → NP PP
- PP → P NP

- N: noun
- P: preposition
- NP: noun phrase
- PP: prepositional phrase

Words
Sentences
Corpora
References

Sentence structure
Grammar
Translation of Sentence Structure
Discussion

# English Grammar

- N→ {sushi, tuna}

- P → {with}

- V → {eat}

- NP → N

- NP → NP PP

- PP → P NP

- VP → V NP

Words
**Sentences**
Corpora
References

Sentence structure
**Grammar**
Translation of Sentence Structure
Discussion

## Dependency Tree

- Alternative syntactic structure
- It shows relationships between words
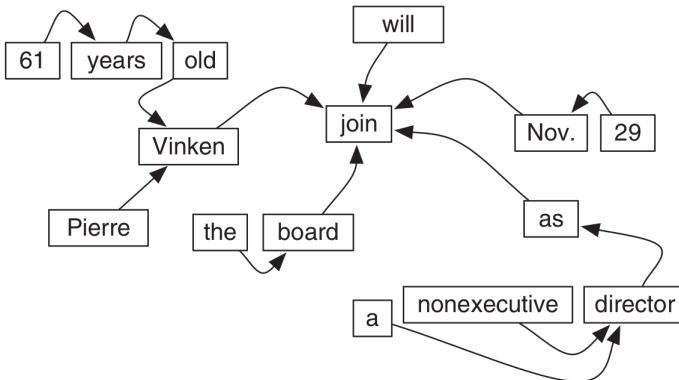
## Dependency Tree

- The central word is the verb; others depend on it, with more words recursively attached.
- Dependencies may be labeled by types such as subject, object, adjunct, etc.

Words    Sentence structure
**Sentences**    **Grammar**
Corpora    Translation of Sentence Structure
References    Discussion

## Dependency Tree
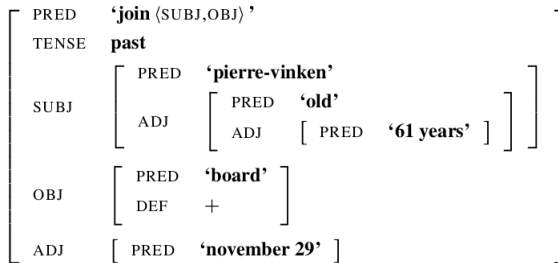
- Head Word: They are words on which others depend
- Dependencies may be labeled by types such as subject, object, adjunct, etc.
- What is the head word of the previous picture?

Words
**Sentences**
Corpora
References

Sentence structure
**Grammar**
Translation of Sentence Structure
Discussion

## Other Grammars..

- Lexical functional grammar (LFG):
- Combinatory categorical grammar (CCG):

Words
**Sentences**
Corpora
References

Sentence structure
**Grammar**
Translation of Sentence Structure
Discussion

# Lexical functional grammar

$$
\begin{bmatrix}
\text{PRED} & \text{`join } \langle \text{SUBJ,OBJ} \rangle \text{'} \\
\text{TENSE} & \text{past} \\
\text{SUBJ} & \begin{bmatrix}
\text{PRED} & \text{`pierre-vinken'} \\
\text{ADJ} & \begin{bmatrix}
\text{PRED} & \text{`old'} \\
\text{ADJ} & \begin{bmatrix} \text{PRED} & \text{`61 years'} \end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}
\text{PRED} & \text{`board'} \\
\text{DEF} & +
\end{bmatrix} \\
\text{ADJ} & \begin{bmatrix} \text{PRED} & \text{`november 29'} \end{bmatrix}
\end{bmatrix}
$$

## Combinatory categorical grammar

- *swim:* S\NP
- *eat:* (S\NP)/NP
- *give:* ((S\NP)/NP)/NP

- The intransitive verb swim requires an NP on the left to form a sen- tence S.
- The transitive verbs eat and give require additional NPs on the right.
- Helps to reduce the number of structural ambiguities.

Words
**Sentences**
Corpora
References

Sentence structure
Grammar
**Translation of Sentence Structure**
Discussion

## Translation of Sentence Structure

- Languages differ in their syntactic structure
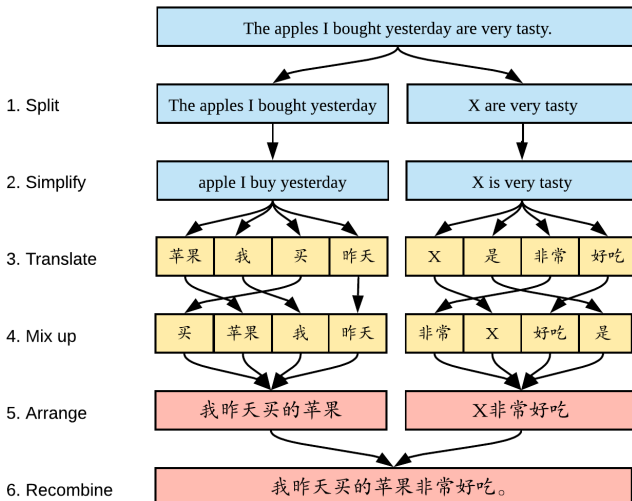- Reordering of words
- Insertion and deletion

Words
Sentences
Corpora
References

Sentence structure
Grammar
**Translation of Sentence Structure**
Discussion

# Translation of Sentence Structure

Words
Sentences
Corpora
References

Sentence structure
Grammar
Translation of Sentence Structure
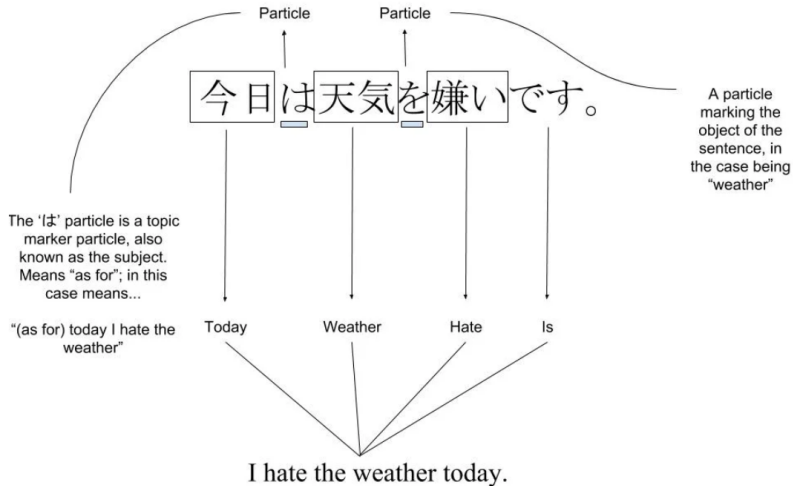Discussion
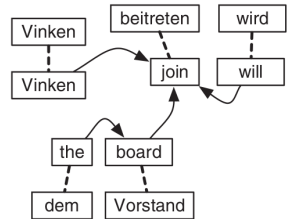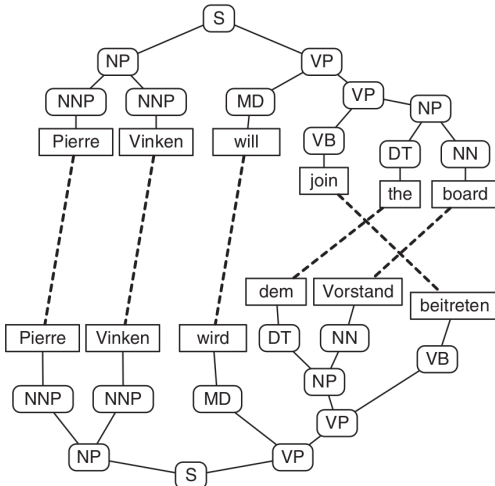
## Translation of Sentence Structure

- German and English differ in their syntactic structure, resulting in changes of word order.
- The movement of the main verb from after the modal (in English) to the end of the sentence (in German) is reflected by one different rule application (VP$\rightarrow$ VP NP in English vs. VP $\rightarrow$ NP VP in German).
- The dependency structure is the same for both languages

Words
Sentences
Corpora
References

Sentence structure
Grammar
**Translation of Sentence Structure**
Discussion

# Translation of Sentence Structure

Words
Sentences
Corpora
References

Sentence structure
Grammar
Translation of Sentence Structure
Discussion

# Translation of Sentence Structure



I hate the weather today.

Words
**Sentences**
Corpora
References

Sentence structure
Grammar
**Translation of Sentence Structure**
Discussion

## Discourse

Co-reference

- Anaphora: denotes the act of referring backwards in a dialog or text
- Cataphora: sees the act of referring forward in a dialog or text

Words
**Sentences**
Corpora
References

Sentence structure
Grammar
**Translation of Sentence Structure**
Discussion

## Discourse

Anaphora

- Susan dropped the **plate**. <u>It</u> shattered loudly.
- The **music stopped**, and <u>that</u> upset everyone.
- Fred was **angry**, and <u>so</u> was I.
- If Sam **buys a new bike**, I will do <u>it</u> as well.

https://en.wikipedia.org/wiki/Anaphora_(linguistics)

Words
**Sentences**
Corpora
References

Sentence structure
Grammar
**Translation of Sentence Structure**
Discussion

## Discourse

Cataphora

- Because **he** was very cold, <u>David</u> put on his coat.
- Although Sam might do **so**, I will not <u>buy a new bike.</u>

https://en.wikipedia.org/wiki/Anaphora (linguistics)

Words
**Sentences**
Corpora
References

Sentence structure
Grammar
Translation of Sentence Structure
**Discussion**

## Discussion about words

- Show Tokenization with Python
- Show Pos-tags
- Wordnet
- Stopword

## Corpora

Reading assignment

## Recourse

- Jurafsky, D. and J. H. Martin. Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Second Edition, Upper Saddle River, NJ: Prentice-Hall, 2008.
- Koehn, P. (2009). Statistical Machine Translation. Cambridge: Cambridge University Press. Chapter-02