



NUI Galway
OÉ Gaillimh

Introduction to NLP

Linguistic Concepts

Dr. Paul Buitelaar
Data Science Institute, NUI Galway



Learning Outcomes of This Lecture

Insight into linguistic foundations of Natural Language Processing

Understand relation between different linguistic levels of analysis

Understand core concepts in morphological, syntactic and semantic analysis

Understand different types of language data



Overview

Foundations in linguistic theory

Linguistic levels of analysis

- Morphology
- Syntax
- Semantics

Types of language data

- Corpus
- Lexicon

Overview

Foundations in linguistic theory

Linguistic levels of analysis

- Morphology
- Syntax
- Semantics

Types of language data

- Corpus
- Lexicon

Linguistic Theory

Linguistic study develops a theory that explains

use of language - 'language performance'

knowledge and acquisition of language - 'language competence'

NLP is concerned with

use of language - **data**

knowledge and acquisition of language - **models**

Linguistic Theory – Some History

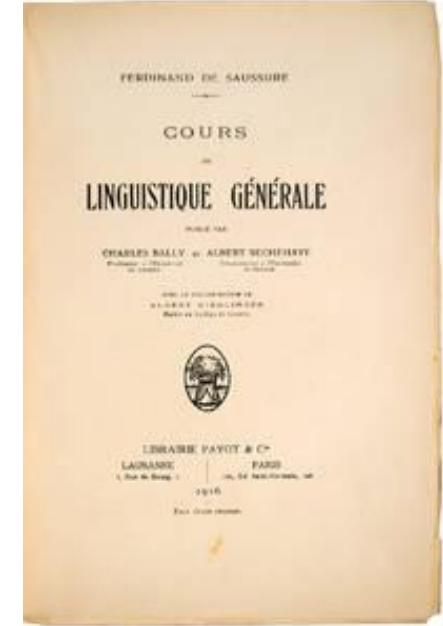
Structural Linguistics (de Saussure, 1916)

Paradigmatic Relations

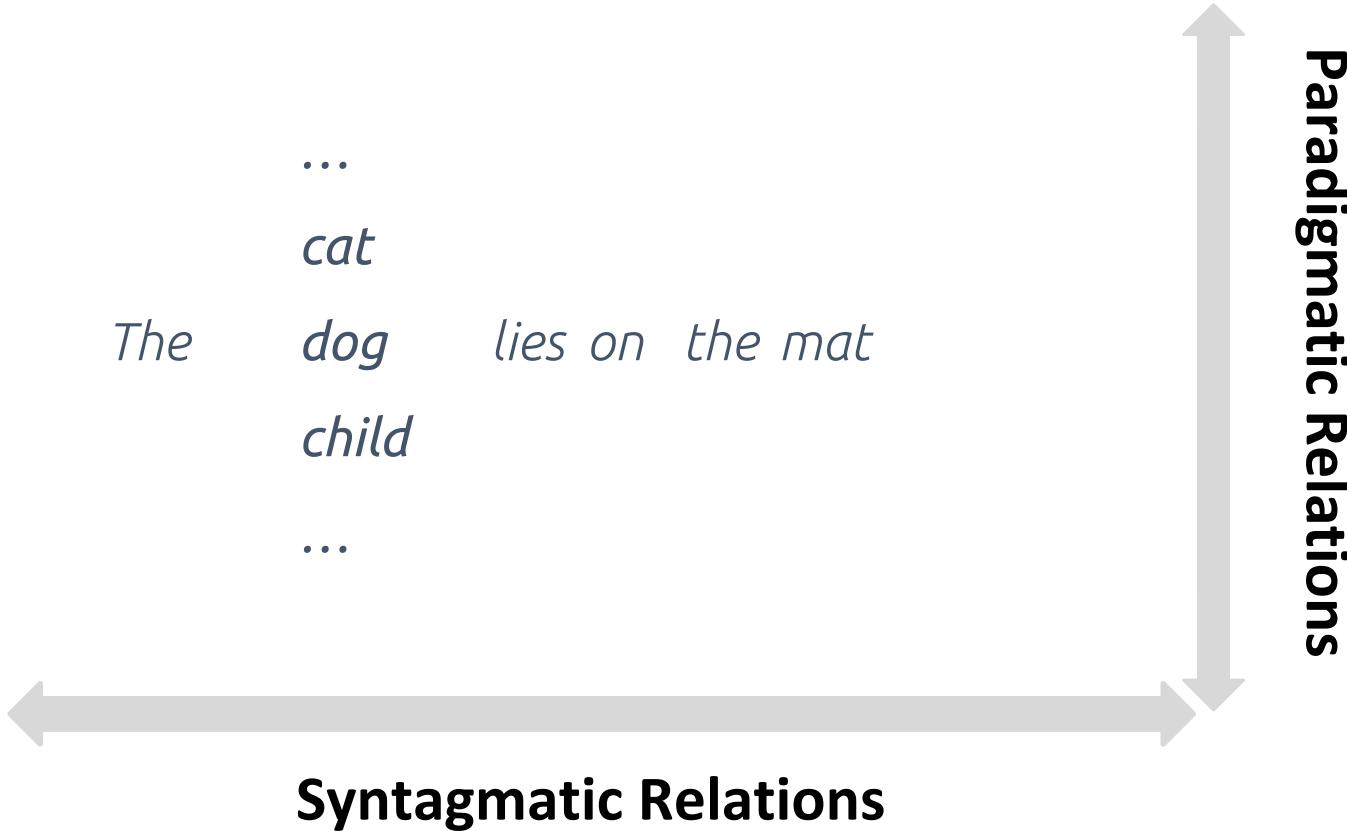
relations between **words that may replace each other** in a sequence

Syntagmatic Relations

relations between **words that may combine with each other** in a sequence



Structural Linguistics - Example



Paradigmatic Relations

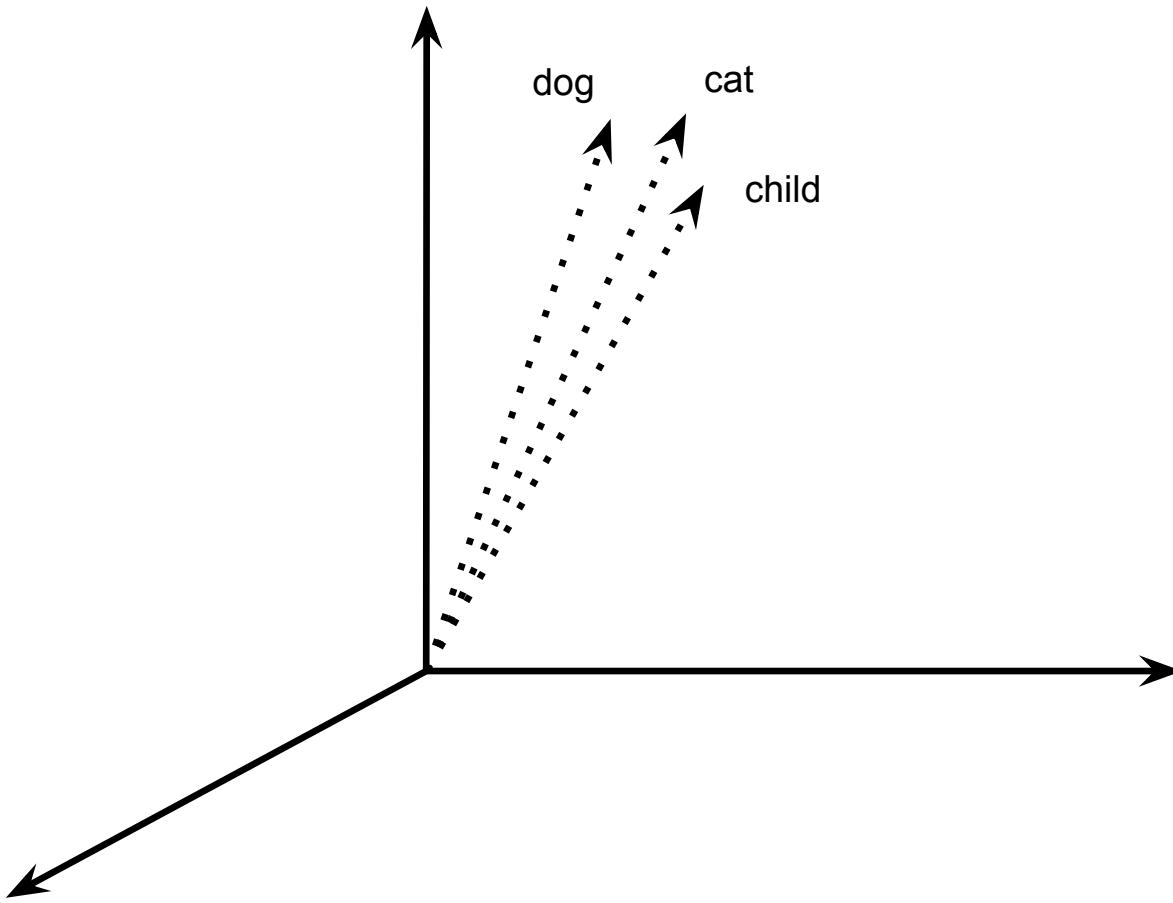


Distributional Hypothesis (Firth 1957) - “you shall know a word by the company it keeps”

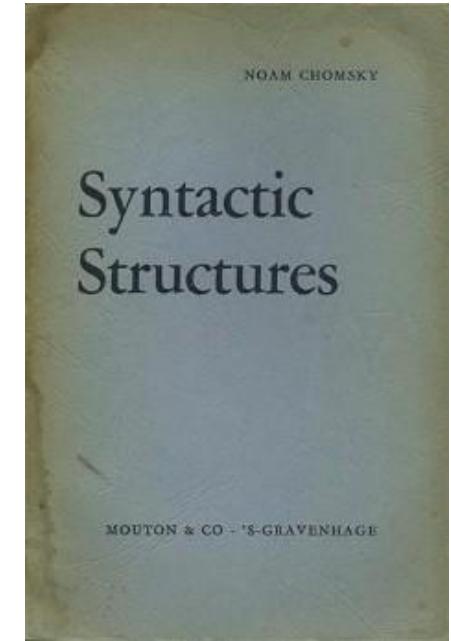
For example, ‘cat’, ‘dog’, ‘child’ can occur in similar contexts and are therefore similar in meaning (distributional semantics)

Distributional Hypothesis is the basis for the **Vector Space Model**

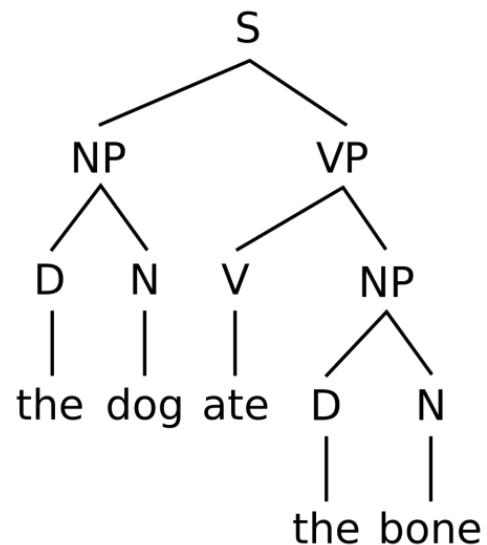
Vector Space Model



Syntagmatic Relations



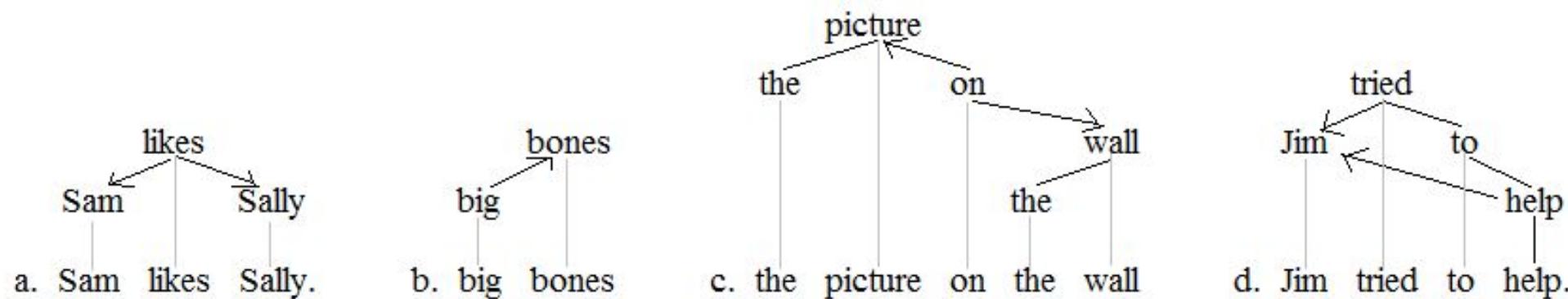
Constituency or Phrase Structure (Chomsky, 1957)



Syntagmatic Relations

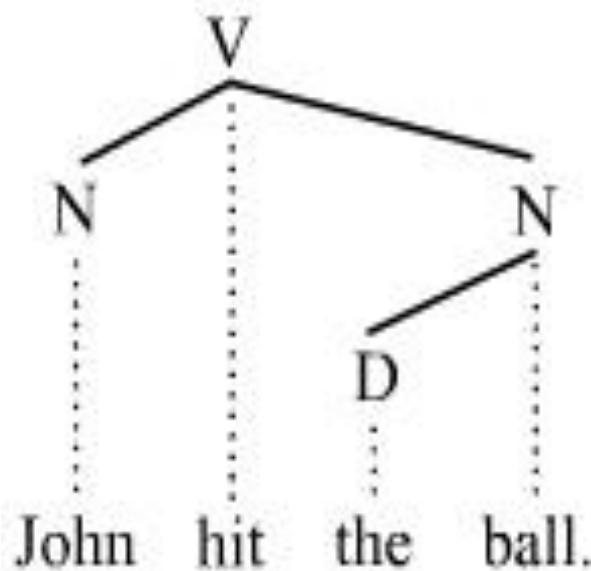
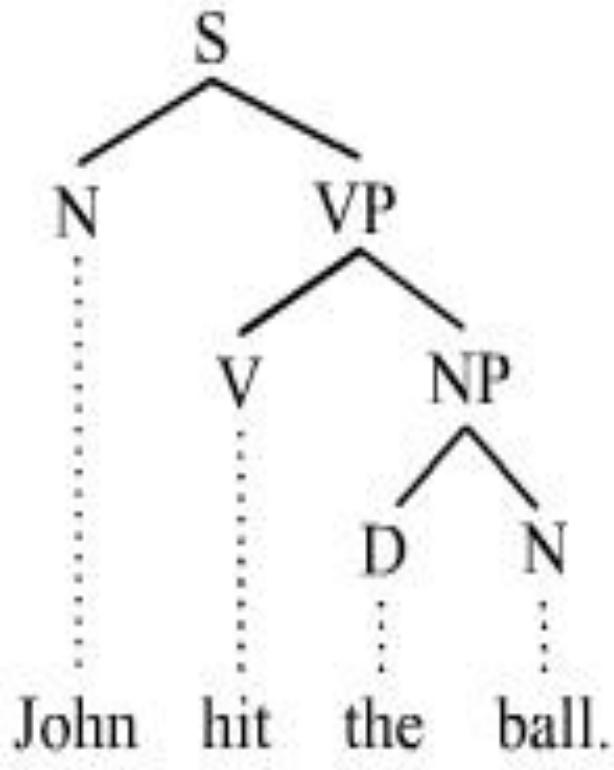


Dependency Structure (Tesnière, 1959)



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

Phrase Structure vs. Dependency Structure



Overview

Foundations in linguistic theory

Linguistic levels of analysis

- Morphology
- Syntax
- Semantics

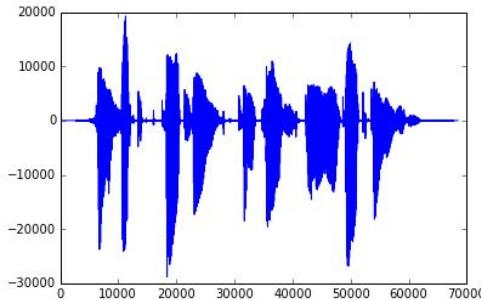
Types of language data

- Corpus
- Lexicon



Linguistic Levels of Analysis

Phonology - acoustic signal



Morphology – word structure

Syntax – sentence and phrase structure

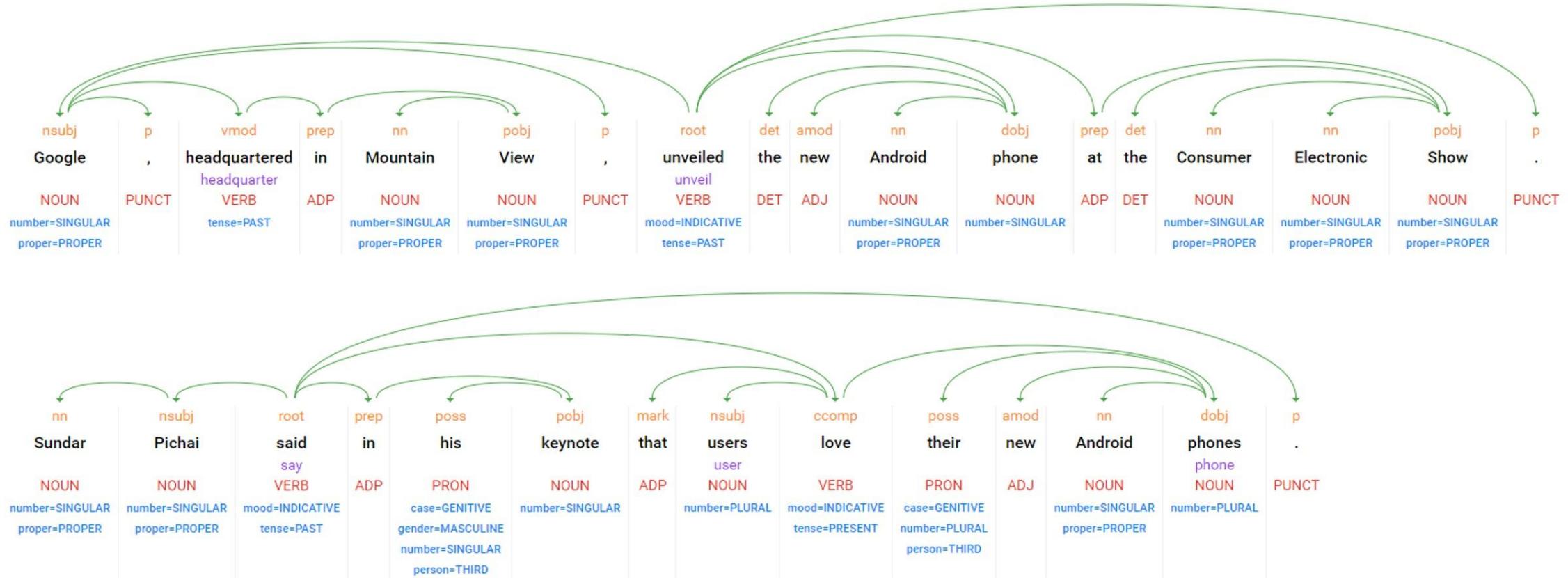
Semantics – meaning and reference

Pragmatics - communication



Linguistic Levels of Analysis

Dependency Parse Label Part of Speech Lemma Morphology



Overview

Foundations in linguistic theory

Linguistic levels of analysis

- **Morphology**
- Syntax
- Semantics

Types of language data

- Corpus
- Lexicon

Morphology – Stemming

Non-linguistic (simplified) analysis of word structure by removing endings/beginnings of words - leaving a common **stem**

For example:

learns > *learn-s*

learning > *learn-ing*

learned > *learn-ed*

Morphology – Stemming

Problematic for instance for irregular verbs

For example:

teaches > ***teach-es***

teaching > ***teach-ing***

taught > **??**

Morphology – Lemmatization

Linguistic analysis of word structure by a *transformation* of morphologically related words to a common **lemma**

For example:

studies *ies* > *y* ***study***

studying *ing* > - ***study***

studied *ied* > *y* ***study***

teaches *es* > - ***teach***

teaching *ing* > - ***teach***

taught > ***teach***

Morphology – Inflection

Linguistic analysis of internal structure of words into
***lemma* with inflectional features**

For example:

writes ***write*** + Verb + 3rd Person + Singular + Present

wrote ***write*** + Verb + 3rd Person + Singular + Past

writing ***write*** + Verb + Progressive

Morphology – Inflection

Inflectional patterns can be shared across word classes

For example, many English verbs share this pattern:

write s

Verb + 3rd Person + Singular + Present

read s

Verb + 3rd Person + Singular + Present

work s

Verb + 3rd Person + Singular + Present

Morphology – Inflection

But note that there are many **irregular inflectional patterns**

For example:

wr o te

Verb + 3rd Person + Singular + Past

read

Verb + 3rd Person + Singular + Past

work ed

Verb + 3rd Person + Singular + Past



Morphology – Inflection

Also note ambiguity between syntactic classes ('part of speech')

For example, identical but **ambiguous inflection** of nouns and verbs:

book s **Verb + 3rd Person + Singular + Present**

book s **Noun + Plural**



Morphology – Derivation

Inflection: grammatical role and structure of words

Derivation: generation of words from other words

For example:

<i>agreement</i>	Noun	derived from	<i>agree</i>	Verb
<i>friendship</i>	Noun	derived from	<i>friend</i>	Noun
<i>civilize</i>	Verb	derived from	<i>civil</i>	Adjective
<i>civilization</i>	Noun	derived from	<i>civilize</i>	Verb

Morphology – Compounds & Decomposition

Compound: concatenation of simple words into a complex word
(note that analysis of compounds is referred to as **decomposition**)

Compounds are not common in English, but for example:

bookshelves > *book + shelves*

bedroom > *bed + room*

policeman > *police + man*

Morphology – Compounds in Other Languages

Compounds (morphological decomposition) are very common in German

Flachbildschirm > *flach + Bildschirm* (*flat + screen*)

Flachbildschirm > *Flachbild + Schirm* (*flat view + screen*)

Flachbildschirm > *flach + Bild + Schirm* (*flat + picture + screen*)

also very common in Dutch

bestuurdersaansprakelijkheidsverzekering (*driver + liability + insurance*)

overeenstemmingsbeoordelingsprocedures (*conformity + assessment + procedures*)

and in many other languages

Multi-Word Expression (MWE)

Compounds are rare in English but correspond instead to
multi-word expressions

For example, technical terms such as in the health domain

cardiac arrest

or names of people, locations, organisations, etc.

Barack Obama

San Francisco

United Nations

Types & Tokens

The number of **tokens** is the total occurrence of individual words

The number of **types** is the total occurrence of unique words

For example, in the sentence '*the cat sat on the mat*' there are:

6 tokens *the, cat, sat, on, the, mat*

5 types *the, cat, sat, on, mat*



Morphology - Types & Tokens

Morphological and MWE analysis effects **nature and number of types and tokens** in a data set through inflection, derivation, decomposition

For example:

Barack Obama's bookshelves had books about ancient civilizations

can be analyzed as:

Barack_Obama book shelf have book about ancient civil|civilize|civilization

Non-Linguistic Tokens

Often **non-linguistic tokens** are used as well, e.g. in tweets:

IRIS, an SMT System developed [@unlp_insight](#), mentioned in [#nuigalway](#) newsletter

Overview

Foundations in linguistic theory

Linguistic levels of analysis

- Morphology
- **Syntax**
- Semantics

Types of language data

- Corpus
- Lexicon

Part of Speech

Syntactic category of a word (noun, verb, etc.), aka ‘Part of Speech (PoS) tag’

Useful for subsequent syntactic parsing - analysing phrase / dependency structure

PoS tag set for English has 36 tags – see next slide for 25 most common tags

PoS tagging example

John saw the saw and decided to take it to the table .

NNP VBD DT NN CC VBD TO VB PRP IN DT NN .

Penn Treebank PoS Tag Set - for English

Nouns	Singular	NN	<i>dog, fork</i>
	Plural	NNS	<i>dogs, forks</i>
	Proper noun - singular, plural	NNP, NNPS	<i>John, Smiths, ...</i>
	Personal pronoun	PRP	<i>I, you, he, she, it</i>
	Wh-pronoun	WP	<i>who, what, ...</i>
Verbs	Base, infinitive	VB	<i>eat</i>
	Past tense	VBD	<i>ate</i>
	Gerund	VBG	<i>eating</i>
	Past participle	VBN	<i>eaten</i>
	Non 3 rd person singular present tense	VBP	<i>eat</i>
	3 rd person singular present tense	VBZ	<i>eats</i>
	Modal	MD	<i>should, can</i>
	To Particle	TO	<i>to eat</i>
Adjectives	Basic	JJ	<i>red, tall</i>
	Comparative	JJR	<i>redder, taller</i>
	Superlative	JJS	<i>reddest, tallest</i>
Adverbs	Basic	RB	<i>quickly</i>
	Comparative	RBR	<i>quicker</i>
	Superlative	RBS	<i>quickest</i>
Prepositions	Preposition	IN	<i>on, in, by, to, with, ...</i>
Determiners	Basic determiner	DT	<i>a, an, the</i>
	WH-determiner	WDT	<i>which, that, ...</i>
Conjunctions	Coordinating Conjunction	CC	<i>and, but, or</i>
Particles	Particle	RP	<i>off (took off), up (put up), ...</i>

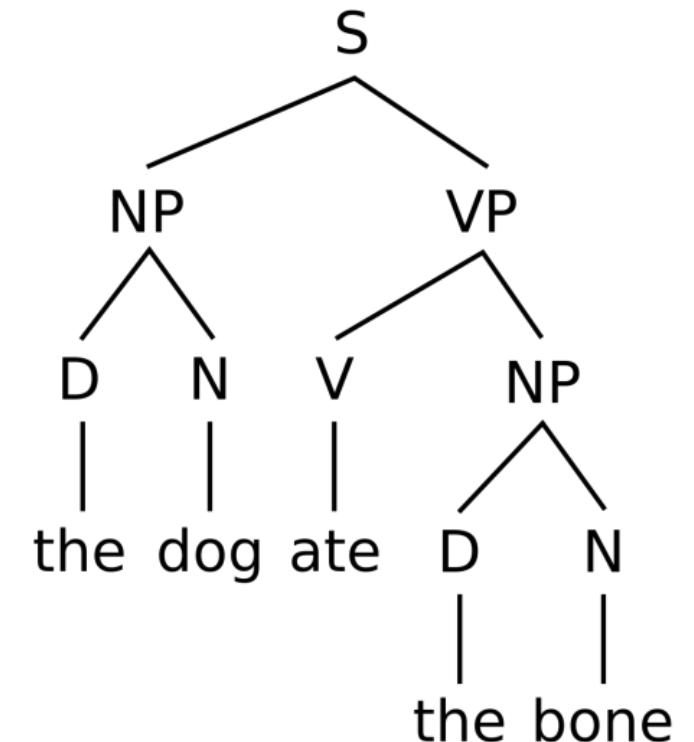
Phrase Structure

Overall structure of groups of words that function as identifiable **constituents (phrases)**

For example

*the dog, the bone
ate the bone*

Noun Phrase (NP)
Verb Phrase (VP)



Phrase Structure Grammar

A (context-free) Phrase Structure Grammar $\mathbf{G}=(N,\Sigma,P,S)$ consists of:

a set of **non-terminal symbols** N , e.g.

N, V, NP, VP, S

a set of **terminal symbols** Σ , e.g.

astronomer, cat, see, the

a set of **productions** P , e.g.

$S \rightarrow NP\ VP$

a **start symbol**, normally 'S'

Phrase Structure Grammar - Example

$G=(N, \Sigma, P, S)$

N : Det, Noun, Prep, Verb, NP, PP, VP, S

Σ : *cat, mat, on, sits, the*

P : $S \rightarrow NP\ VP$

$[NP\ the\ cat]\ [VP\ sits\ on\ the\ mat]$

$NP \rightarrow Det\ Noun$

$[DET\ the]\ [N\ cat] \mid [DET\ the]\ [N\ mat]$

$VP \rightarrow Verb\ PP$

$[V\ is]\ [PP\ on\ the\ mat]$

$PP \rightarrow Prep\ NP$

$[P\ on]\ [NP\ the\ mat]$

Start symbol S

$[S\ the\ cat\ sits\ on\ the\ mat]$

Lexical Rules (Lexicon)

$G = (N, \Sigma, P, S)$

N : Det, Noun, Prep, Verb, NP, PP, VP, S

Σ : *cat, mat, on, sits, the*

P : Det \rightarrow *the*

Noun \rightarrow *cat | mat | dots*

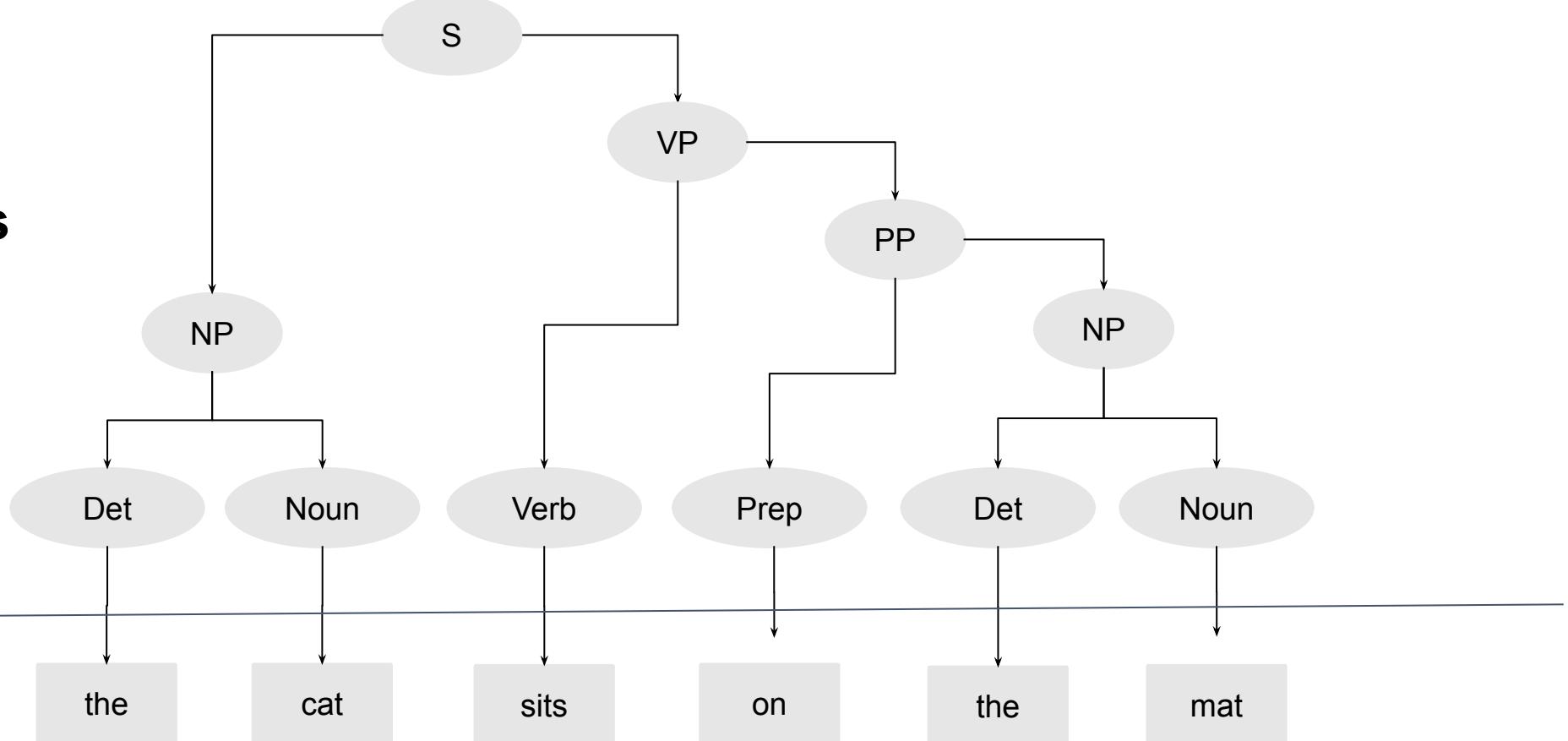
Prep \rightarrow *on | with*

Verb \rightarrow *sits*

Start symbol S

Parse Tree based on this Grammar & Lexicon

non-terminals



terminals

Phrase Structure Grammar with Recursion

$G=(N, \Sigma, P, S)$ consisting of:

N : Det, Noun, Prep, Verb, NP, PP, VP, S

Σ : *cat, dots, mat, on, sits, the, with*

P : $S \rightarrow NP\ VP$

$NP \rightarrow Det\ Noun$

$VP \rightarrow Verb\ PP$

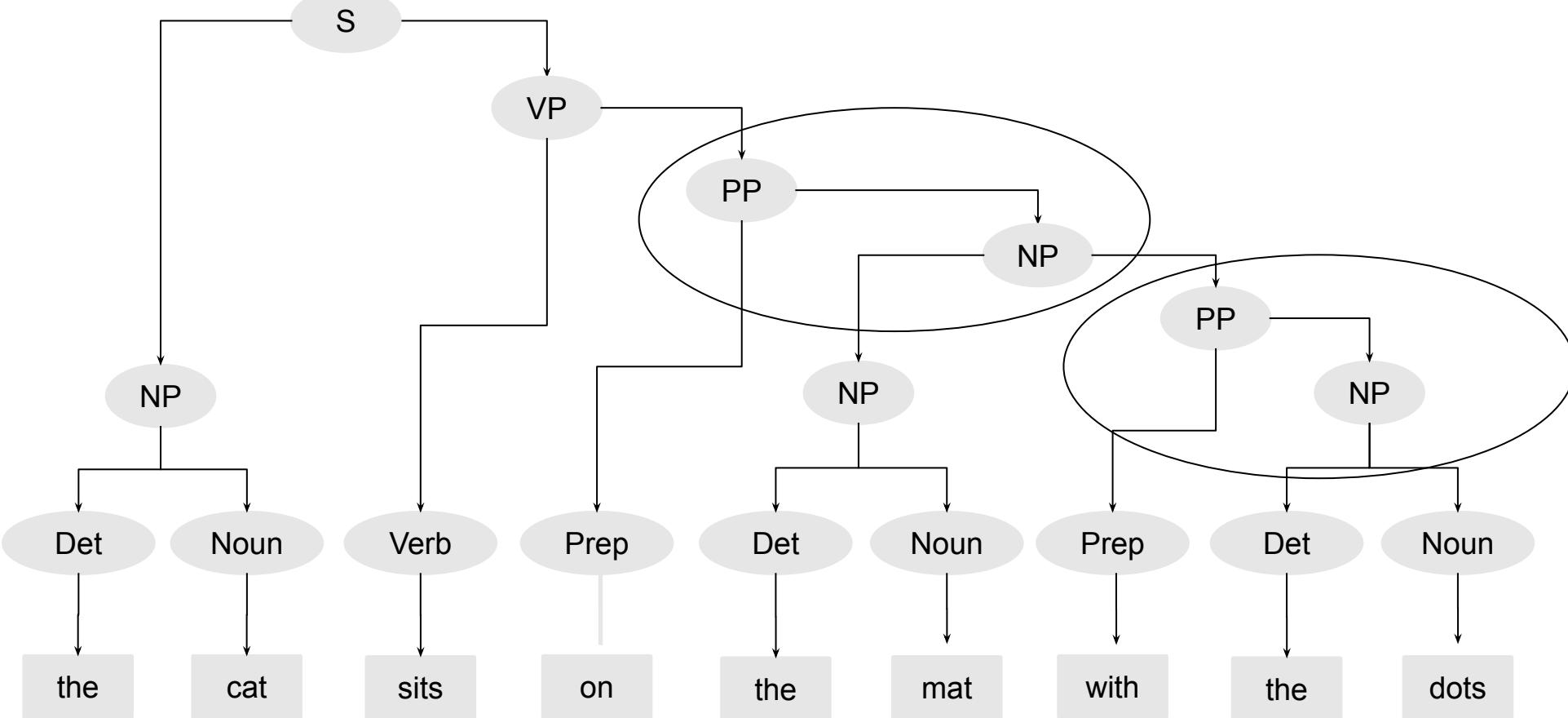
$PP \rightarrow Prep\ NP$ **< Recursion**

$NP \rightarrow NP\ PP$ **< Recursion**

Start symbol S

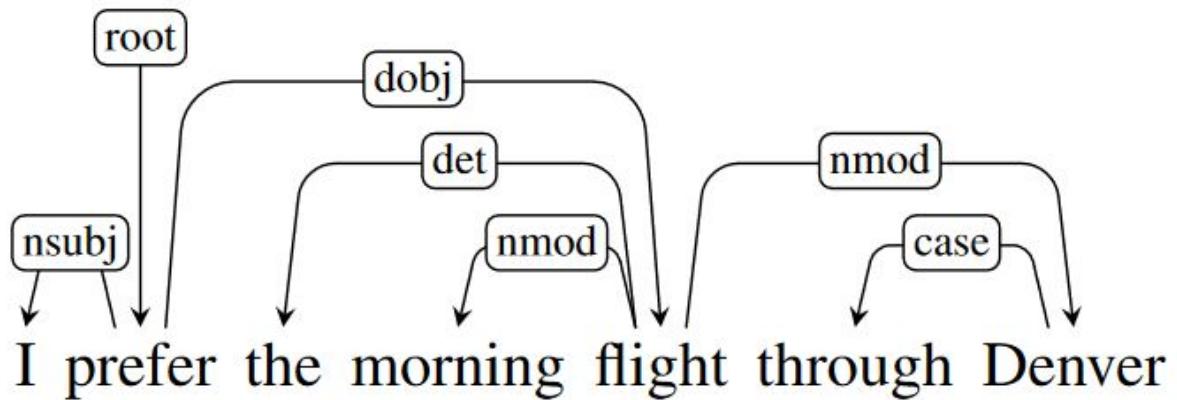
the cat sits on the mat with the dots

Parse Tree with Recursion



Dependency Structure

Syntactic analysis by use of a set of dependency relations between words, e.g.



Jurafsky & Martin Ch 13

Dependency Relations

Most commonly used are **Universal Dependency** relations

Clausal Argument Relations	Description
NSUBJ	Nominal subject
DOBJ	Direct object
IOBJ	Indirect object
CCOMP	Clausal complement
XCOMP	Open clausal complement
Nominal Modifier Relations	Description
NMOD	Nominal modifier
AMOD	Adjectival modifier
NUMMOD	Numeric modifier
APPOS	Appositional modifier
DET	Determiner
CASE	Prepositions, postpositions and other case markers
Other Notable Relations	Description
CONJ	Conjunct
CC	Coordinating conjunction

Jurafsky & Martin p272



Overview

Foundations in linguistic theory

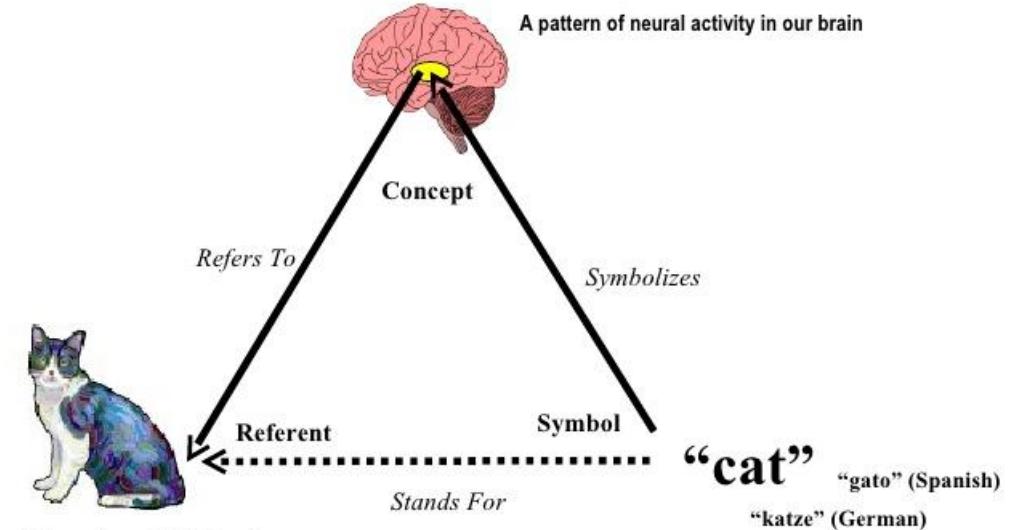
Linguistic levels of analysis

- Morphology
- Syntax
- Semantics

Types of language data

- Corpus
- Lexicon

Semantics



Semantics studies interpretation and reference

Foundations in ‘semiotics’ (study of symbols and their reference)

Semiotic Triangle (Ogden & Richards, 1923) is interface between:

objects in the world (**referents**)

cognitive interpretation of these objects (**concepts**)

symbolic representations for these cognitive interpretations (**symbols**)

Semantics of Language Data

Semantic analysis of language data is concerned with

- **how linguistic symbols** (words, phrases, sentences)
- **refer to concepts** (entities, relations, events)
- **about referents** (objects in the world)

Relation between linguistic symbols and concepts commonly referred to as '**meaning**'

Levels of Semantic Analysis

Lexical Semantics

meaning of a word (*table, large, to book, in, ...*)

Compositional Semantics

combined meaning of several words in a phrase (*large table, book a table*) or sentence (*He booked a large table.*)

Discourse Semantics

combined meaning of several sentences (*He booked a table. It was still available.*)

Note: also in dialog (as in chatbots and other dialog systems)

Semantic Analysis Tasks

Word Sense Disambiguation

identify intended meaning ('sense') of a word in a specific linguistic context

Semantic Role Labeling

identify implied semantic role of an entity expressed by a word or phrase

Coreference Resolution

identify which words or phrases express the same entity

Word Sense Disambiguation (WSD)

Ambiguous words have different meanings due to **homonymy**

table in the context of ‘furniture’

table in the context of ‘accounting’

Related to this is **synonymy** (different words with the same meaning)

table, chart, ...

WSD: identify correct word sense for ambiguous words - for example:

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

Whether on the mantelpiece or on the dining table, vases are a beautiful addition to any home.

Semantic Role Labeling

Verbs introduce **Semantic Roles** for other words in the sentence
For example, *to hit* introduces roles for:

Agent who hits

Recipient who was hit

Instrument what was used to hit with

Temporal when did the hitting take place

Location where did the hitting take place

[*Kristina* AGT] *hit* [*Scott* REC] *with a* [*baseball* INS] [*yesterday* TMP] [*in the park* LOC]

Coreference Resolution

Identify words/phrases referring to the same entity - for example:

Barack Obama nominated Hillary Clinton as his secretary of state on Monday.

He chose her because she had foreign affairs experience as a former First Lady.

Overview

Foundations in linguistic theory

Linguistic levels of analysis

- Morphology
- Syntax
- Semantics

Types of language data

- Corpus
- Lexicon



Types of Language Data

Language Data	Specifics
formal reports	<ul style="list-style-type: none">• complete grammatical sentences
slides	<ul style="list-style-type: none">• incomplete ungrammatical sentences and phrases
text fields in DBs	<ul style="list-style-type: none">• incomplete ungrammatical sentences and phrases• database structure
emails	<ul style="list-style-type: none">• incomplete ungrammatical sentences and phrases• threads with temporal and social network structure
social media	<ul style="list-style-type: none">• incomplete ungrammatical sentences, phrases• threads with temporal and social network structure• multimedia data
...	...

Types of Annotation (Data Labelling)

Annotation	Specifics
part-of-speech	annotation of words with ‘PoS tags’ (N, V, Adj, Prep,...)
word sense	annotation of words with their ‘sense’ (bank_01, bank_02,...)
semantic role	annotation of words with their ‘semantic role’ (Agent, Recipient, Instrument,...)
entity type	annotation of words with a type of ‘entity’ (Person, Location, Organization,...)
translation	words in a ‘corpus’ in language A aligned to words in a ‘corpus’ in language B
...	...

Overview

Foundations in linguistic theory

Linguistic levels of analysis

- Morphology
- Syntax
- Semantics

Types of language data

- **Corpus**
- Lexicon



Corpus

A corpus is an (annotated) language dataset

General corpus vs. domain-specific corpus

Annotated corpus vs. ‘raw data’

Monolingual, bilingual, multilingual corpus

Parallel, comparable corpus

Multimodal corpus

*Please note: **corpus** (singular), **corpora** (plural)*

General Corpus

Representative of a language as a whole

e.g. 'British English from late twentieth century'

Used for training general purpose NLP tools

not tuned to a specific domain

Examples

British National Corpus (BNC)

100 million words, British English text/speech

Open American National Corpus (OANC)

15 million words, American English text/speech

Corpus of Contemporary American English (COCA)

560 million words, American English text

German Reference Corpus

32.85 billion words (Oct. 2017), German text

Domain Corpus

Representative of a domain specific subset of a language

language as used specifically, e.g., in **healthcare, finance, legal, ...**

Used for training domain-specific NLP tools, e.g.

healthcare: corpus of medical documents such as patient records, clinical notes, scientific papers, ...

<https://www.i2b2.org/NLP/DataSets/Main.php>

legal: corpus of legal documents such as national and international laws, legal cases, contracts, , ...

<https://archive.ics.uci.edu/ml/datasets/Legal+Case+Reports>

Annotated Corpus

Corpus with annotations on word, phrase, sentence or document level

Syntax/Morphology PoS, phrase structure, parse tree, ...

Semantics word senses, terms, entities, relations, ...

Other translations, multi-modal references, ...

Annotated corpus is used in NLP for supervised training and/or evaluation

Annotated (labelled) dataset

Corpus Annotation by Crowdsourcing

Data: Given are sentences from suggestion forums for an integrated news feed platform, called 'Feedly'.

Job: Read each sentence, and choose the type of the sentence from the provided options.

1. Suggestion : A suggestion, recommendation, advice, warning, tips to the readers.
2. Non-suggestion: Everything which doesn't fit to the suggestion category as defined above.

You may read the provided *Source Text* from which the sentence is extracted. Please **choose your option only for the given sentence**, and not for the source text.

Sentence:

Love the tagging on the web version.

Source text:

Love the tagging on the web version. This feature needs to be on the mobile versions as well.

Choose the information type of the given sentence, from the provided options (required)

- Suggestion
- Non-Suggestion



Monolingual, Bilingual, Multilingual Corpus

Monolingual Corpus

text (or speech data) in **one language**

more readily available (easy to find data in one language)

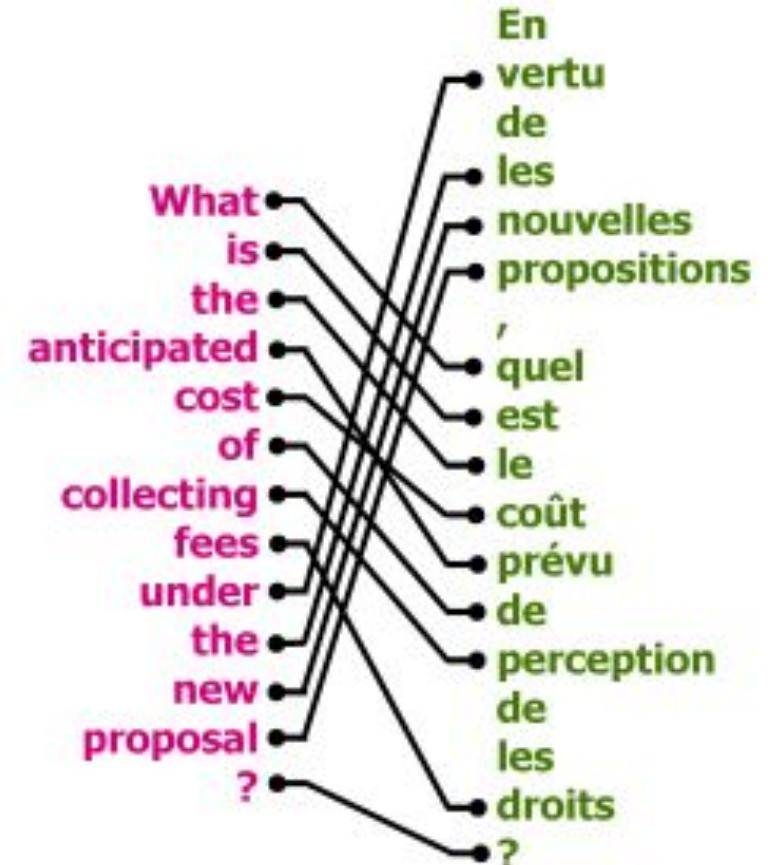
Bilingual and Multilingual Corpus

text (or speech data) in **two or more languages**

consist of **pairs** (bilingual) or **tuples** (multilingual) of translations

mostly used for **training machine translation systems**

not easily available (difficult to find translated data)



Parallel vs. Comparable Corpus

Parallel Corpus

collection of translated documents, e.g. proceedings of the European Parliament with translations between 21 European languages

<http://www.statmt.org/europarl/>

Comparable Corpus

collection of documents on the same topic in different languages, e.g. links between Wikipedia articles in different languages

Parallel Corpus - Example

Linguee

The provision of measures to accommodate the needs of disabled people at the workplace plays an important role in combating discrimination on grounds of disability.

eur-lex.europa.eu

Maatregelen gericht op aanpassing van de werkplek aan de behoeften van personen met een handicap vervullen bij de bestrijding van discriminatie op grond van een handicap een belangrijke rol.

eur-lex.europa.eu

[...] not exceed 48 hours, humane working hours with no forced overtime, and a safe and healthy workplace free from harassment.

fairolympics.org

[...] werkweek, met menswaardige werktijden zonder gedwongen overwerken, en in een veilige en gezonde werkplaats vrij van intimidatie.

fairolympics.org

The end result is that businesses can be more agile in responding to their customers and market needs, and be ready for the workplace of the future.

mds.ricoh.com

Het eindresultaat is dat bedrijven flexibeler kunnen reageren op de markt en de behoeften van hun klanten en klaar zijn voor de werkplek van de toekomst.

ricoh.be

To avoid favoritism or the appearance of favoritism in the workplace in accordance with the policies and procedures adopted by the Company.

colgate.com.ve

Om vriendjespolitiek of de schijn van vriendjespolitiek op de werkplek te vermijden in overeenstemming met de door het bedrijf aangenomen beleidslijnen en procedures.

colgate.com.ve



Comparable Corpus - Example

The **French Revolution** (French: Révolution française [ʁevɔlysjɔ̃ fʁɑ̃sɛz]) was a period of far-reaching social and political upheaval in France that lasted from 1789 until 1799, and was partially carried forward by Napoleon during the later expansion of the French Empire. The Revolution overthrew the monarchy, established a republic, experienced violent periods of political turmoil, and finally culminated in a dictatorship by Napoleon that rapidly brought many of its principles to Western Europe and beyond. Inspired by liberal and radical ideas, the Revolution profoundly altered the course of modern history, triggering the global decline of absolute monarchies while replacing them with republics. Through the Revolutionary Wars, it unleashed a wave of global conflicts that extended from the Caribbean to the Middle East. Historians widely regard the Revolution as one of the most important events in human history.^{[1][2][3]}

English Wikipedia on French Revolution: *Napoleon, war, global view, ...*

Die **Französische Revolution** von 1789 bis 1799 gehört zu den folgenreichsten Ereignissen der neuzeitlichen europäischen Geschichte. Die Abschaffung des feudalabsolutistischen Ständestaats sowie die Propagierung und Umsetzung grundlegender Werte und Ideen der Aufklärung als Ziele der Französischen Revolution – das betrifft insbesondere die **Menschenrechte** – waren mitursächlich für tiefgreifende macht- und gesellschaftspolitische Veränderungen in ganz Europa und haben das moderne Demokratieverständnis entscheidend beeinflusst. Die heutige **Französische Republik** als liberal-demokratische **Verfassungsstaat** westlicher Prägung stützt ihr Selbstverständnis unmittelbar auf die Errungenschaften der Französischen Revolution.

German Wikipedia on French Revolution: *human rights, constitution, ...*

Multimodal Corpus

Consists not only of language data
but also image, audio, video, ...

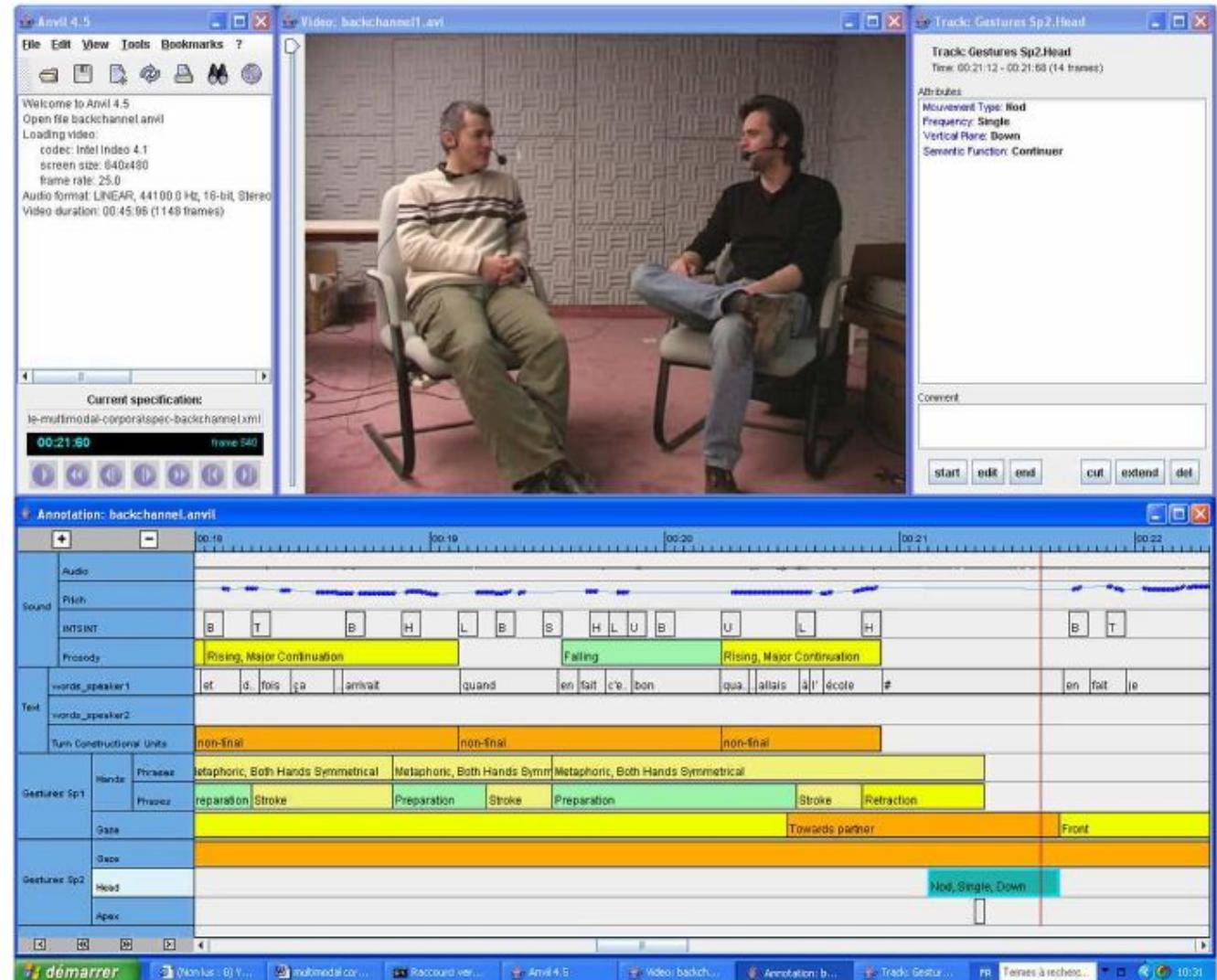


Image from: Blache, P., Bertrand, R. and Ferré, G., 2008, May. Creating and exploiting multimodal annotated corpora: the ToMA project. In *International LREC Workshop on Multimodal Corpora* (pp. 38-53). Springer, Berlin, Heidelberg.



Overview

Foundations in linguistic theory

Linguistic levels of analysis

- Morphology
- Syntax
- Semantics

Types of language data

- Corpus
- Lexicon

Lexicon

Database of words with **lexical properties**
such as spelling, pronunciation,
morphology, PoS, semantics, ...

Originating out of paper dictionaries

work

/wɜ:k/ 

noun

noun: work; plural noun: works; plural noun: the works

1. activity involving mental or physical effort done in order to achieve a result.
"he was tired after a day's work in the fields"
synonyms: labour, toil, exertion, effort, slog, drudgery, the sweat of one's brow; [More](#)
antonyms: leisure, rest
 - work as a means of earning income; employment.
"I'm still looking for work"
synonyms: employment, job, day job, post, position, situation, means of earning one's living, occupation, profession, career, business, trade, line; [More](#)
 - the place where one is employed.
"I was returning home from work on a packed subway"
 - the period of time one spends in paid employment.
"he was going to the theatre after work"
 - **WEST INDIAN**
a job.
"I decided to get a work"
2. a task or tasks to be undertaken.
"they made sure the work was progressing smoothly"
synonyms: tasks, jobs, duties, assignments, commissions, projects; chores
"haven't you got any work to do?"
 - the materials for a task.
"she frequently took work home with her"
 - **BRITISH**
activity involving construction or repair.
"extra costs caused by additional building works"

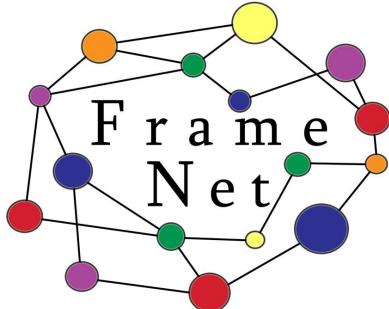


Semantic Lexicon



Organized by ‘synsets’ – represents word meaning by use of sets of **Synonyms**

<https://wordnet.princeton.edu/>



Organized by ‘frames’ – represents word meaning by use of **Semantic Roles**

<https://framenet.icsi.berkeley.edu/fndrupal/about>

WordNet

Word meaning represented in WordNet by use of “synsets” (Miller, 1995)

For example:

{board, plank} represents ‘*board*’ in the meaning of ‘*plank*’

{board, committee} represents ‘*board*’ in the meaning of ‘*committee*’

WordNet - Synsets for the Noun *table*

Noun

- S: (n) table, tabular array (a set of data arranged in rows and columns) "see *table 1*"
- S: (n) table (a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs) "*it was a sturdy table*"
- S: (n) table (a piece of furniture with tableware for a meal laid out on it) "*I reserved a table at my favorite restaurant*"
- S: (n) mesa, table (flat tableland with steep edges) "*the tribe was relatively safe on the mesa but they had to descend into the valley for water*"
- S: (n) table (a company of people assembled at a table for a meal or game) "*he entertained the whole table with his witty remarks*"
- S: (n) board, table (food or meals in general) "*she sets a fine table*"; "*room and board*"



WordNet - Hypernyms & Hyponyms

Synsets are organized hierarchically in WordNet, expressing **generalization** (hypernyms) and **specialization** (hyponyms)

{entity}

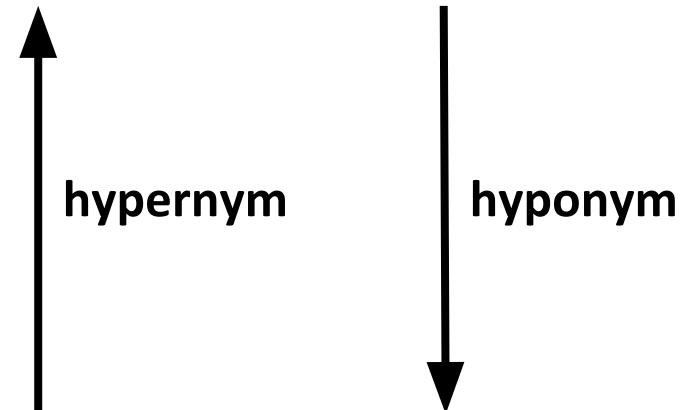
...

{whole, unit}

{building material}

{lumber, timber}

{board, plank}



FrameNet

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

“frame [is] a description of a type of event, relation, or entity and the participants in it”

Frames express Semantic Roles - for example:

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

Frames, Frame Elements & Lexical Units

Frame Apply_Heat

Frame Elements Cook ; Food ; Container ; Heating_Instrument

Lexical Units *bake, boil, cook, simmer, steam, ...*

Note: lexical units within a frame are synonyms



FrameNet

Apply_heat

Definition:

A Cook applies heat to Food, where the Temperature setting of the heat and Duration of application may be specified. A Heating instrument, generally indicated by a locative phrase, may also be expressed. Some cooking methods involve the use of a Medium (e.g. milk or water) by which heat is transferred to the Food. A less semantically prominent Food or Cook is marked Co-participant.

Sally FRIED an egg in butter.

Sally FRIED an egg in a teflon pan.

Ellen FRIED the eggs with chopped tomatoes and garlic.

This frame differs from Cooking_creation in focusing on the process of handling the ingredients, rather than the edible entity that results from the process.

Lexical Unit	Frame	LU Status	Lexical Entry Report	Annotation Report
cook up.v	Coming_up_with	FN1_Sent	LE	Anno
cook up.v	Cooking_creation	Finished_Initial	LE	Anno
cook.n	Cooking_creation	In_Use	LE	Anno
cook.n	People_by_vocation	Created	LE	Anno
cook.v	Apply_heat	Finished_Initial	LE	Anno
cook.v	Cooking_creation	Finished_Initial	LE	Anno
cook.v	Absorb_heat	Needs_SCs	LE	
cookie.n	Food	Finished_Initial	LE	Anno
cooking.n	Cooking_creation	In_Use	LE	Anno
cooking.n	Apply_heat	In_Use	LE	Anno

Lab of this Week

Exercises in linguistic analysis



NUI Galway
OÉ Gaillimh



NUI Galway
OÉ Gaillimh

QA

