

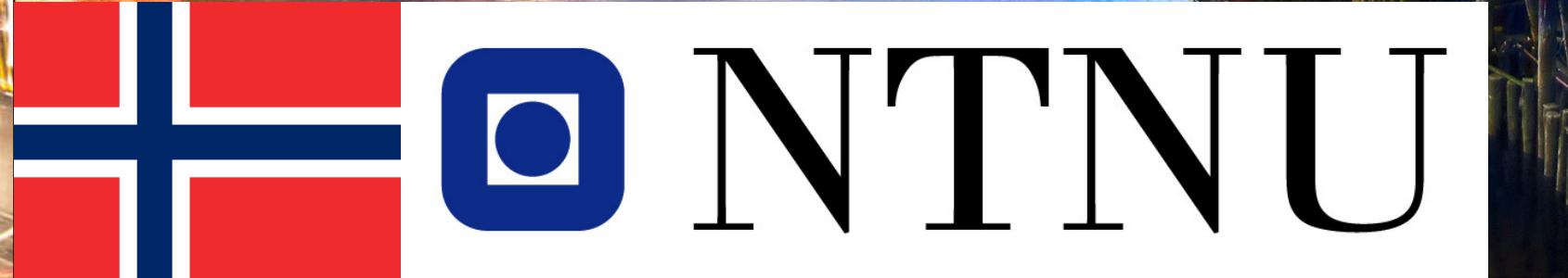
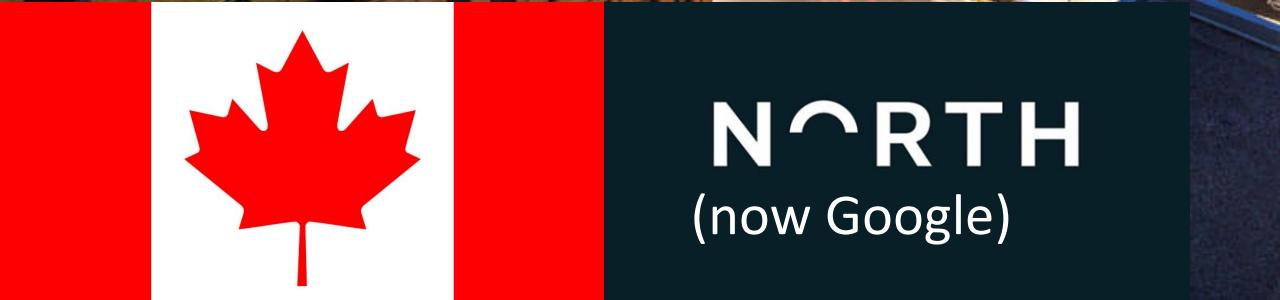
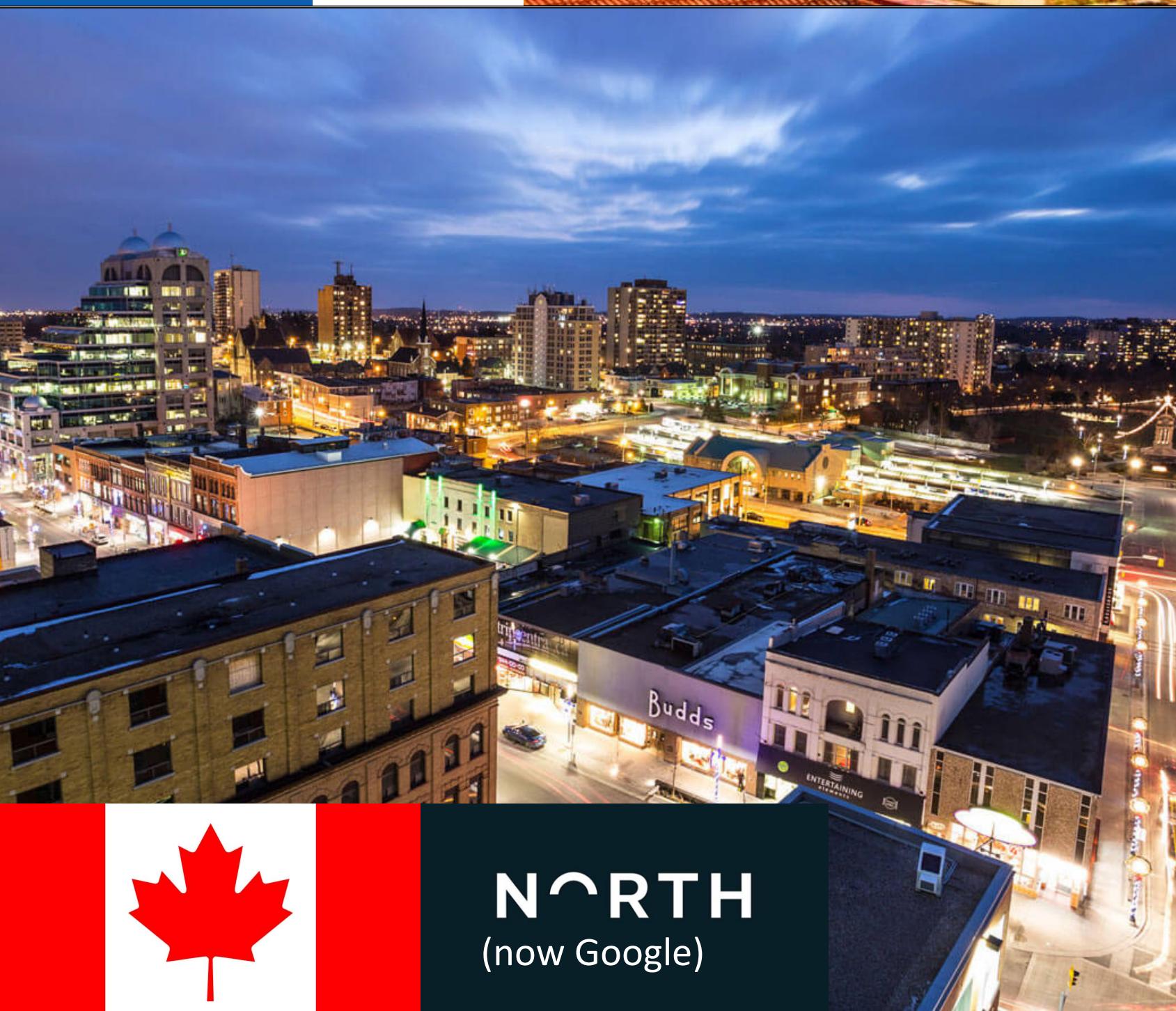
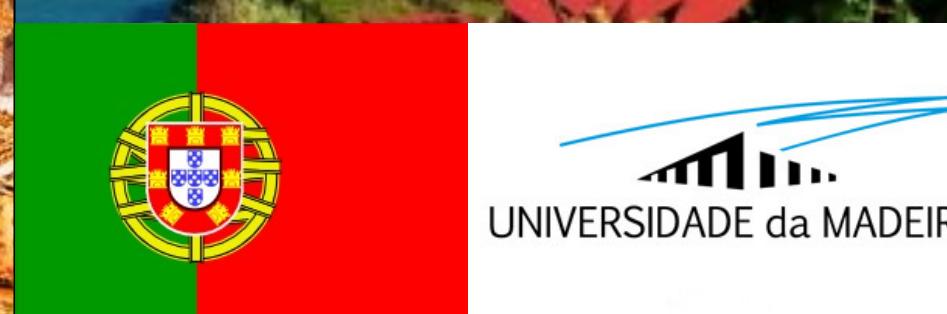
Advanced Machine Learning For Design

Lecture 2 - Machine Learning and Natural
Language Processing / Part 1

Module 1

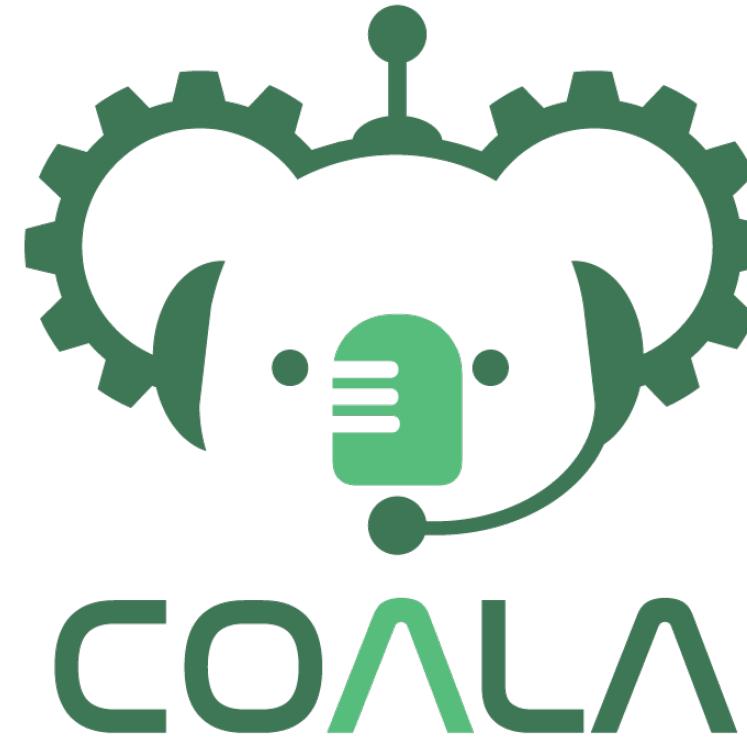
Evangelos Niforatos
28/09/2022

aml4d-ide@tudelft.nl
<https://aml4design.github.io/>



Natural Language Processing

- A sub-field of AI and machine learning in which machines learn to understand natural language as spoken and written by humans
- Goals:
 - Recognize the language, understand it, and respond to it
 - Categorise textual content (e.g. spam vs. Not-spam)
 - Translate between languages
 - Generate new text
- An enabler for technology such as chatbots and digital assistants like Siri or Alexa



COgnitive Assisted
agile manufacturing for a LABor force
supported by trustworthy
Artificial Intelligence

DEMO: COALA Chatbot

A Cognitive Assistant

Dr. Evangelos Niforatos
Assistant Professor | AI-Powered Human Augmentation



Horizon 2020
European Union Funding
for Research & Innovation

*Faculty of Industrial Design Engineering
Delft University of Technology
Delft, the Netherlands*



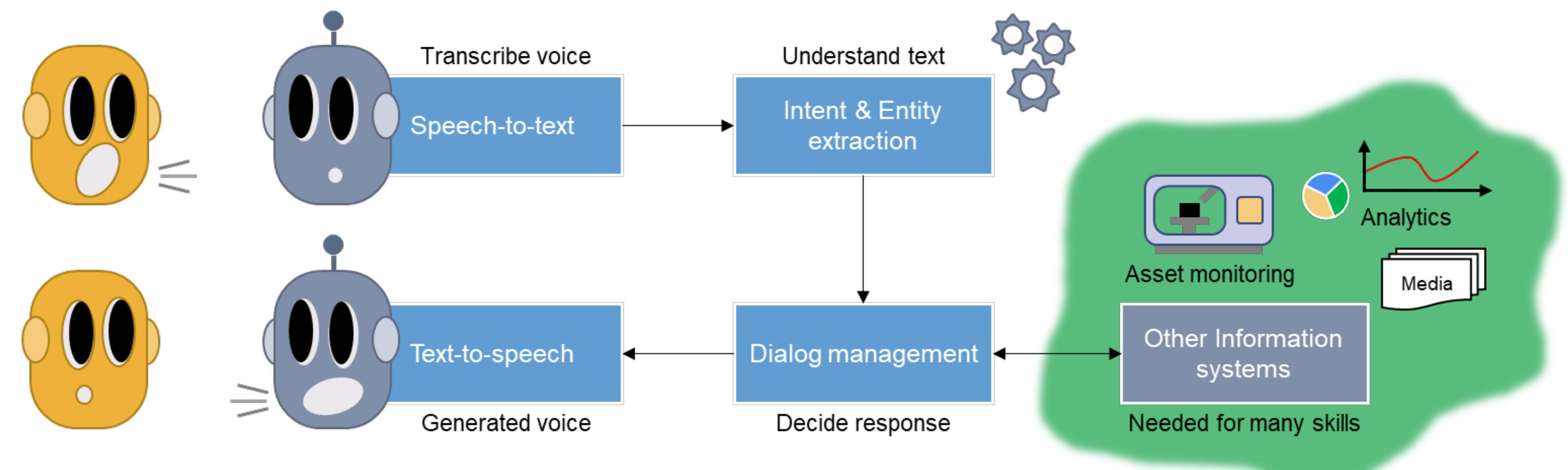
Digital Intelligent Assistants

... interact via natural language.

... support or take over time-consuming, stressful, and other unwanted activities.

... “know” and do things that humans associate with intelligence.

... are amorphous compound technology.



Basic structure of a conversational agent

The amorphous “backend”

The Factory of Today

- Long distances
- Narrow space
- Noisy
- Declining workforce
- Automated tasks
- “Smart” machines
- Sensors, IoT & APIs
- Still: Manual labor



Source: Whirlpool Corporation



Source: MiR / Whirlpool Corporation



Source: Diversey Netherlands Production Bv



Source: FRATELLI PIACENZA S.P.A.



Source: Città Studi.

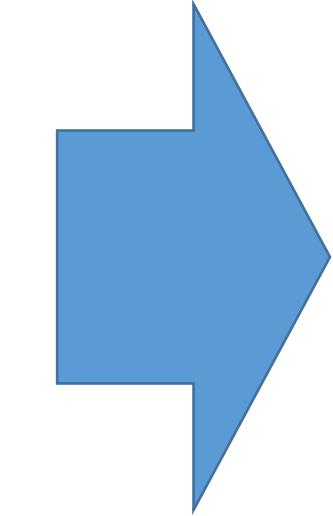
The COALA Chatbot



2x Use Cases:

- Augmented Analytics
 - White goods production
- On-the-Job Training
 - Detergent production
 - Textile production

Cognitive Assistant Benefits
Central access
Customizable
Delegatory
Eyes-free
Instructive & Collaborative
Hands-free
Ubiquitous
Multimodal (text & speech)
Always on
Responsive
Context-aware
Intuitive (no learning curve)



- Reduce training time
- Share knowledge amongst operators
- Improve production performance
- Increase safety

Visit this link to find the related article:

<https://ai4manufacturing.com/anatomy-of-a-digital-assistant/>

COALA Demo and Discussion

1. **Participation is entirely voluntary**
2. You will need both your **Laptop AND** your **Smartphone**
 1. It will take approximately 40 min
3. You will be randomly split between 2x groups and you will receive a post-it with **your unique ID**:
 1. Group **L**: Laptop (**Yellow** post-it)
 2. Group **S**: Smartphone (**Pink** post-it)
4. Watch a 3-minute video about a detergent production line and a simulation we built + Feedback
5. Watch a 5-minute video about the COALA chatbot + Feedback
6. Experience the COALA chatbot by completing a series of tasks and a survey.
7. A round table on what works well, what doesn't and how it can be improved.

Please open the Miro board

<https://tiny.one/miro1> pw: coala2020

REMINDER: include your <ID> in all your post-its!

Part 1

Introduction to a detergent production line and its simulation



Part 1 Feedback

Do you understand the factory context and how the simulation will be used? Any questions or suggestions?

Please use the Miro board to write them down
REMINDER: include your <ID> in all your post-its!

Duration: 3 minutes

Part 2

The COALA chatbot



Part 2 Feedback

Do you understand the capabilities of the chatbot?
Any questions or suggestions?

Please use the Miro board to write them down
REMINDER: include your <ID> in all your post-its!

Duration: 3 minutes

Part 3

Experience the COALA chatbot

Part 3: Experience COALA

- EVERYONE open this link on your laptops: <https://tiny.one/coala-survey>
- **YELLOW** team: On your LAPTOPS, open this link: <https://tiny.one/coala>
- **PINK** team: On your SMARTPHONES, open this link: <https://tiny.one/coala>
- Introduce yourselves to COALA by typing your unique ID:
- e.g., “*Hi, I am S32*” or “*Hi there, I am L26*”
- Complete the tasks (within 10 minutes) and the survey!

Part 3 Feedback

How did you experience using the chatbot? What worked well, what didn't and how can it be improved?

Please use the Miro board to write them down
REMINDER: include your <ID> in all your post-its!

Duration: 10 minutes

Why natural language processing?

And why is it a hard problem?

Fora, social media, blog, products review

Interviews

Design Critiques: Help new and amateur designers improve their designs

Join

r/design_critiques

Posts

Hot New Top All Time ...

Posted by u/erik_messaki 2 years ago

387 Hi, community! How do you like our new illustration? Thanks for your opinion:

VIA AMAZON.COM

My transformation is complete

"It is day 87 and the [horses](#) have accepted me as one of their own. I have grown to understand and respect their gentle ways. Now I question everything I thought I once knew and fear I am no longer capable of following through with my primary objective. I know that those who sent me will not relent. They will send others in my place... But we will be ready." —via Amazon/customer review/[ByronicHero](#).

About Community

Help new and amateur designers improve their designs through reviews and critiques. If you are an experienced designer, please review a submitted post and share your constructive suggestions.

79.8k Members 6 Online

Created Apr 19, 2010

Filter by flair

Website Rebranding

Pinned Tweet

IDE TU Delft @detudelft · Feb 10

"How can we design for societal transitions?" → the central question of @detudelft's new [#research](#) strategy, and we're looking for up to 10 talented design researchers(!) to help us in our mission to redirect [#design](#). Apply here: tinyurl.com/3uz9xvy #wearehiring #delftdesign

1 4

IDE TU Delft @detudelft · 8h

Good afternoon. @tudelft alumni have developed a [#dating](#) app that does not work based on swiping, but with a smart algorithm. www-emerge-nl.translate.goog/interviews/sta...

1 1 4

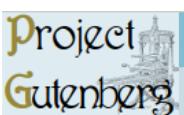
IDE TU Delft @detudelft · 8h

Alumnus Julian Jagtenberg - one of the creators of the world's first sleep robot - about the origin of his startup Somnox:

fd.nl

'Toen ik die Disney-film zag, wist ik het' Julian Jagtenberg van Somnox had een fascinatie voor robots en dan vooral van het pluizige type. O...

Books (digital, or digitised)



About ▾ **Search and Browse** ▾ **Help** ▾

Quick search **Go!** [Donation](#) [PayPal](#)

Frequently Viewed or Downloaded

erviewee: XXX
erviewer: XXX
e of Interview: mm.dd.yy
ation of Interview: XXX
of Acronyms: FP=Frank Peterson, IN=Interviewer

[Transcript 00:00:10]

So what was going on in your life when you joined the Marines?

Well when I joined the navy, actually that was in 1950 at the age of 18. much other than the fact that I wanted to get away from Topeka and see at the rest of world was really all about.

Um-hm.

[00:26]

of course having... gone through the flight training I received my wings
I commission in October of 1952. And the- one of the reasons I opted for
Marines, I knew there had never been a black pilot in the Marine Corps.
I wanted to see if I could achieve that goal, which I was able to do.

I then my first duty assignment would have been in Cherry Point, North Carolina. But I'd had enough of the South and decided I wanted to stay away from the South if I possibly could, so Headquarters Marine Corps, at my request, changed my orders to El Toro, El Toro, California.

what I didn't realize is that I'd jumped from the frying pan into the fire because El Toro was the training base for replacement pilots in Korea. So I ped from the frying pan into the Korean War via El Toro.

I see.

[d Transcript 00:01:21]

Bo: An intelligent network agent to promote physical activity in children with Congenital Heart Defects

Bo

An intelligent network agent to promote physical activity in children with Congenital Heart Defects

Challenge

There are various organisations such as the European Society of Cardiology [2012] and American Heart Association [2013] which describe why physical activity is essential for the development of children. However, parents who have a Congenital Heart Defect (CHD), may suffer from a lack of opportunity to perform physical activity due to their own health, motor development and autonomy during childhood [Krol, 2003]. This impediment arises due to a lack of knowledge about what their child does not know to what extent their child can exercise safely, and therefore adopt overprotective behaviours (Schwarzmann, Thomé, & Moens, 2016).

Design process

In order to understand better overprotection during childhood, 305 online parental stories from various patient-association websites were analysed using Natural-Language-Processing tools. The analysis revealed a lifetime journey of these families, where an uncertain future evoked a constant search for symptoms. This led to the design of a PSS. The design was generated through seven families with a CHD paediatric patient to understand the user needs and requirements of the system. Furthermore, Bo has a conversational agent function where parents can send concerns to the medical team and find relief when seeing their child's heart rate zone visualised in the physical activity path.

PSS solution - BO

To encourage families to have a more ordinary sports life, Bo is introduced, a smart PSS aiming to support parents and their children with a CHD to understand better the safety boundaries of exercise. Bo is a conversational agent with an activity tracker and his nine system modules. Bo aims to guide the child through different physical activities and challenges. Furthermore, Bo has a conversational agent function where parents can send concerns to the medical team and find relief when seeing their child's heart rate zone visualised in the physical activity path.

Implementation

A conversational prototype of the intelligent agent was developed and implemented in the real context of four families to understand how could it influence overprotection. The implementation evaluation showed great potential for Bo to be evaluated through in-depth interviews with paediatric CHD patients and their parents and the medical team to improve the system further. The results showed that Bo provides a supportive exploratory environment for the family, where the user can feel safe and supported by the medical team and parents, instead of limiting the child, adopt an encouraging attitude towards physical activity.

PSS aim

PSS devices

Hosana Cristina Morales Ornelas
BO – An intelligent network agent to promote physical activity in children with Congenital Heart Defects
31st of January, 2020
MSc Integrated Product Design - Medisign

Committee
Prof. Dr. Gerd Kortuem
MSc. Jiwon Jung
MD PhD Arend van Deutekom
Sophia Children's Hospital, ErasmusMC

Company

- Analysis of how parents perceive their baby, their behaviours towards their child, and thus understand how does overprotection develops throughout childhood
 - >300 stories, manually and NLP analysis

Big Textual Data = Language at scale

- One of the largest reflections of the world, a man-made one
- Essential to better understand people, organisations, products, services, systems
 - and their relationships!
- Language is a proxy for human behaviour and a strong signal of individual characteristics
 - Language is always situated
 - Language is also a political instrument

Why NLP?

- Answer questions using the Web
- Translate documents from one language to another
- Do library research; summarize
- Archive and allow access to cultural heritage
- Interact with intelligent devices
- Manage messages intelligently
- Help make informed decisions
- Follow directions given by any user
- Fix your spelling or grammar
- Grade exams
- Write poems or novels
- Listen and give advice
- Estimate public opinion
- Read everything and make predictions
- Interactively help people learn
- Help disabled people
- Help refugees/disaster victims
- Document or reinvigorate indigenous languages

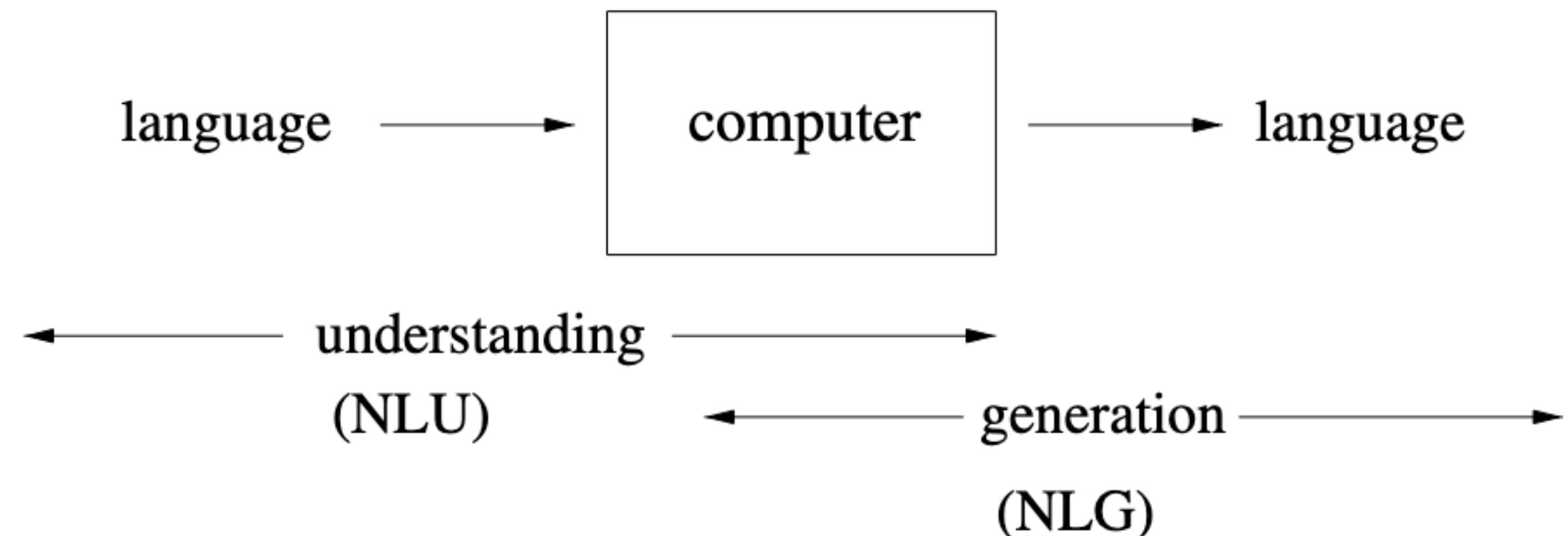
Natural Language Processing

- Computers using natural language as input and/or output

Natural: human communication, unlike e.g., programming languages

Language: signs, meanings, and a code connecting signs with their meanings

Processing: computational methods to allow computers to 'understand', or to generate



Go beyond keyword matching



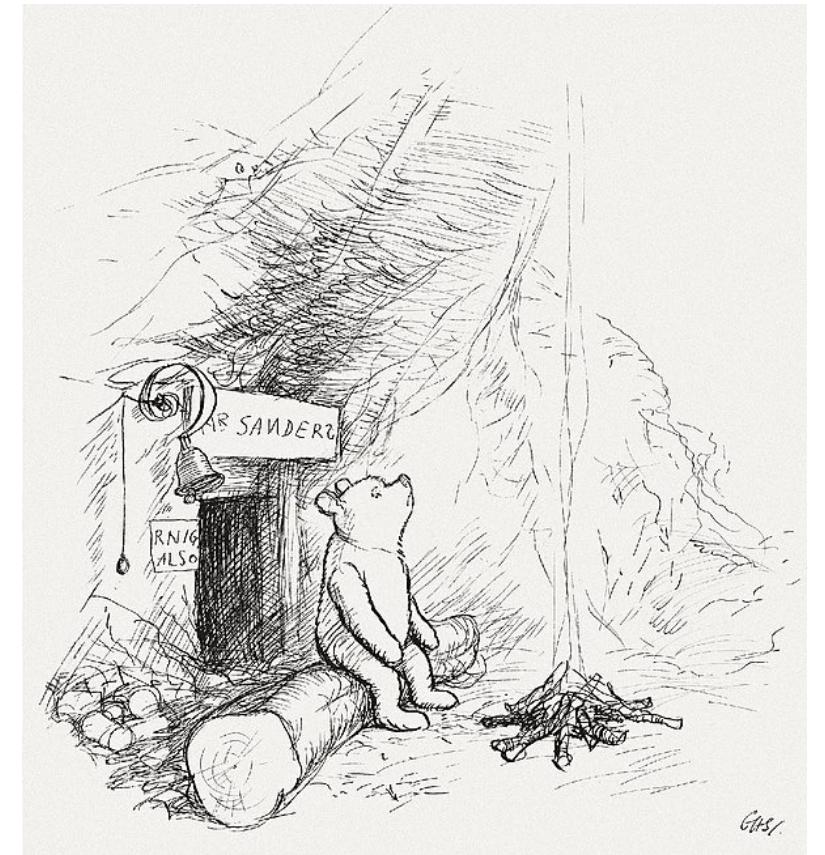
- Identify the **structure** and **meaning** of **words**, **sentences**, **texts** and **conversations**
- Deep understanding of broad language

NLP is hard

- Human languages are messy, ambiguous, and ever-changing
 - A string may have many possible interpretations at every level
 - The correct resolution of the ambiguity will depend on the intended meaning, which is often inferable from the context
- There is tremendous diversity in human languages
 - Languages express the same kind of meaning in different ways
 - Some languages express some meanings more readily/often
- Knowledge Bottleneck
 - Knowledge about language
 - Knowledge about the world
 - Common sense
 - Reasoning

Ambiguity and Expressivity

Christopher Robin is alive and well. **He** is the same person that you read about in the book, **Winnie the Pooh**. As a boy, **Chris** lived in a pretty home called **Cotchford Farm**. When **Chris** was three years old, **his father** wrote a poem about **him**. The poem was printed in a magazine for others to read. **Mr. Robin** then wrote a book



- Who wrote **Winnie the Pooh**?
- Where did **Chris** live?



Sparsity

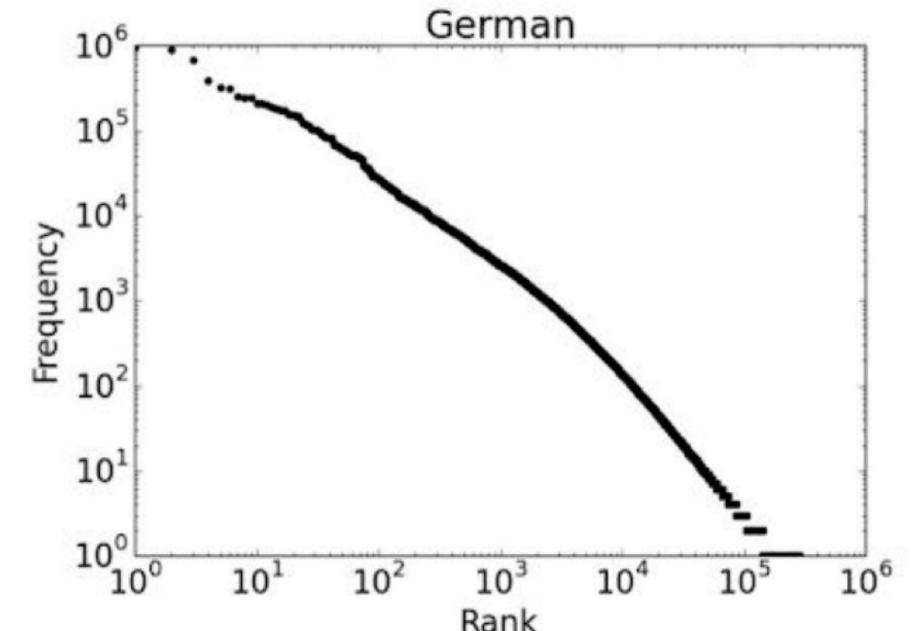
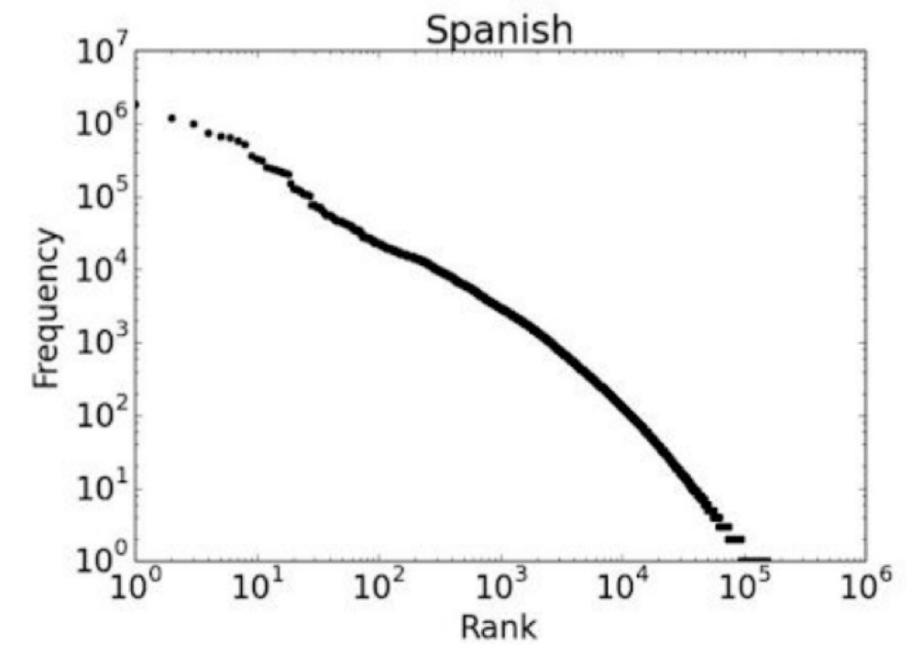
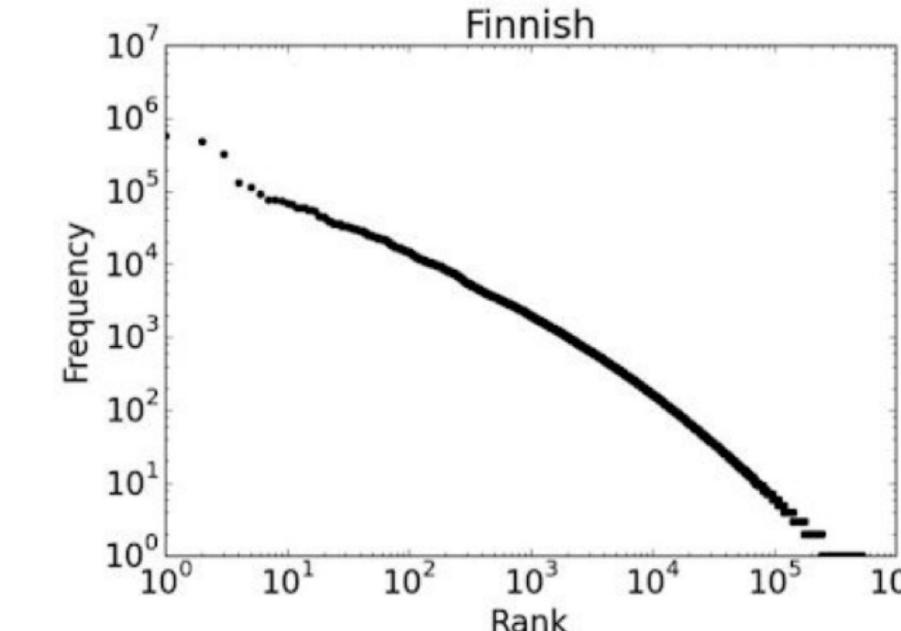
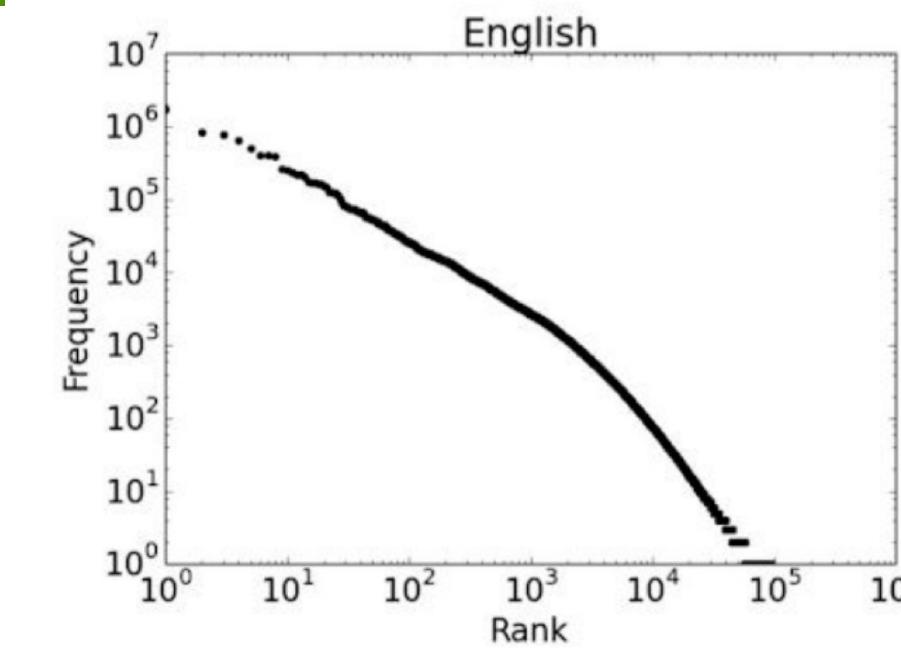
- Zipf's Law: The distribution of word frequencies is very skewed

“... given some document collection, the frequency of any word is inversely proportional to its rank in the frequency table...”

- The most frequent word will occur approximately twice as often as the second most frequent word, which occurs twice as often as the fourth most frequent word, etc.
 - Regardless of how large our corpus is, there will be a lot of infrequent words
- This means we need to find clever ways to estimate the value of words that we have **rarely** (or **never**) seen

any word		nouns	
Frequency	Token	Frequency	Token
1,698,599	the	124,598	European
849,256	of	104,325	Mr
793,731	to	92,195	Commission
640,257	and	66,781	President
508,560	in	62,867	Parliament
407,638	that	57,804	Union
400,467	is	53,683	report
394,778	a	53,547	Council
263,040	I	45,842	States

Words ordered by their frequency

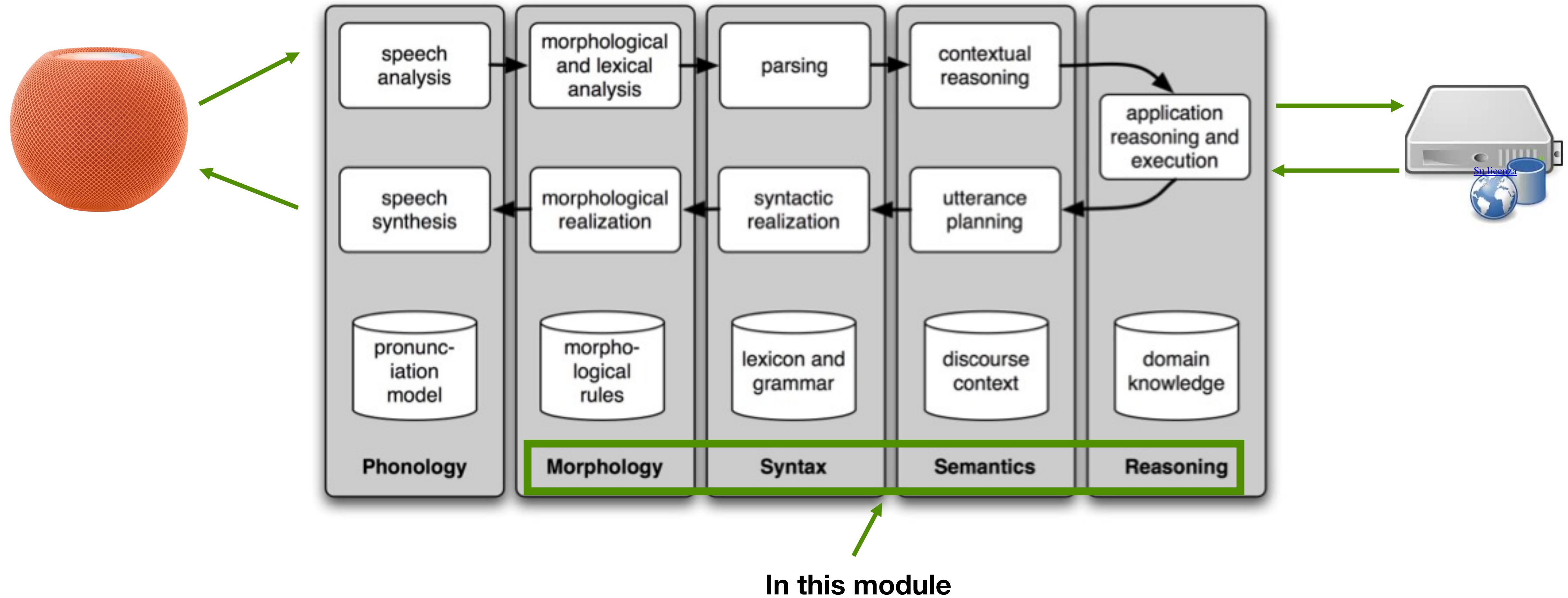


Language evolves

LOL	Laugh out loud
G2G	Got to go
BFN	Bye for now
B4N	Bye for now
Idk	I don't know
FWIW	For what it's worth
LUWAMH	Love you with all my heart



An Example of NLP Process - Smart Speakers

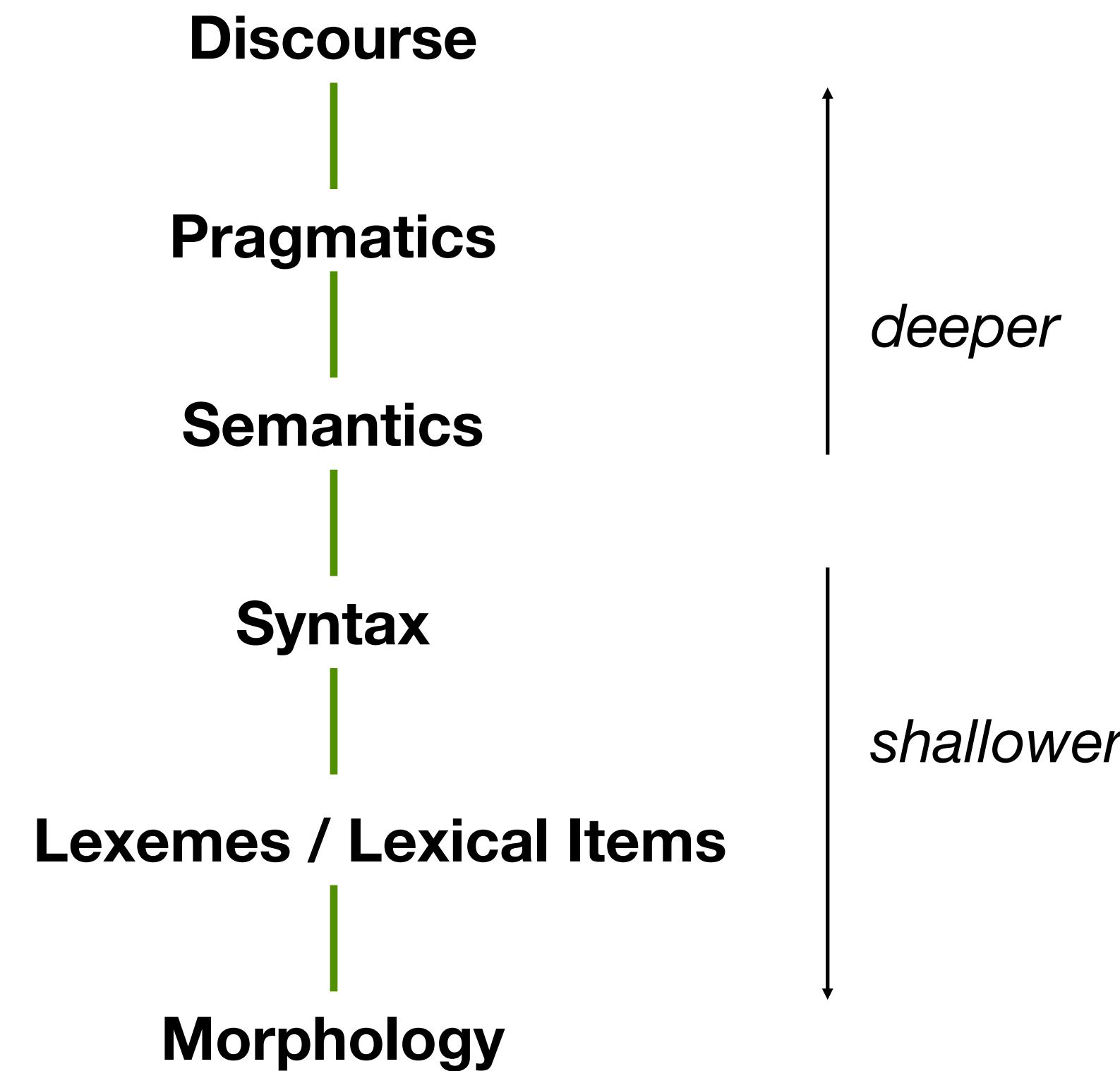


Language

A recap

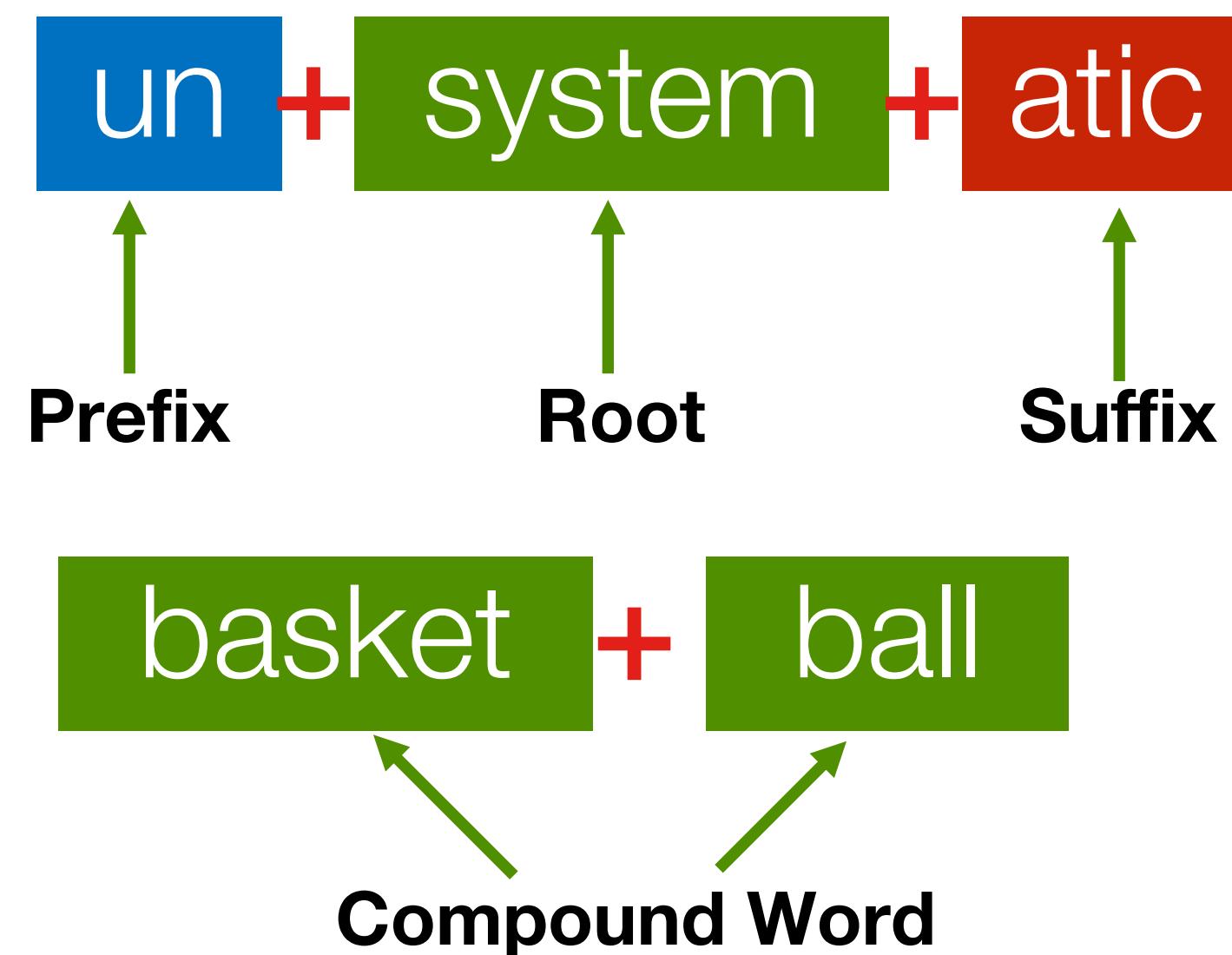
Levels of Linguistic Representation

- The mapping between levels is hard
- Appropriateness of representation depends on the application



Morphology

- Words are the atomic elements in a language
- Many words have an internal structure that shapes their meaning
- Morphology analysis: split words into meaningful components
 - The structure of words
 - Useful for orthographic error correction



Free Morphemes

Can stand alone as own word

Dog, gentle, picture, gem

Bound Morphemes

Derivational

Prefixes

de- pre-
In- un-

Inflectional

Suffixes

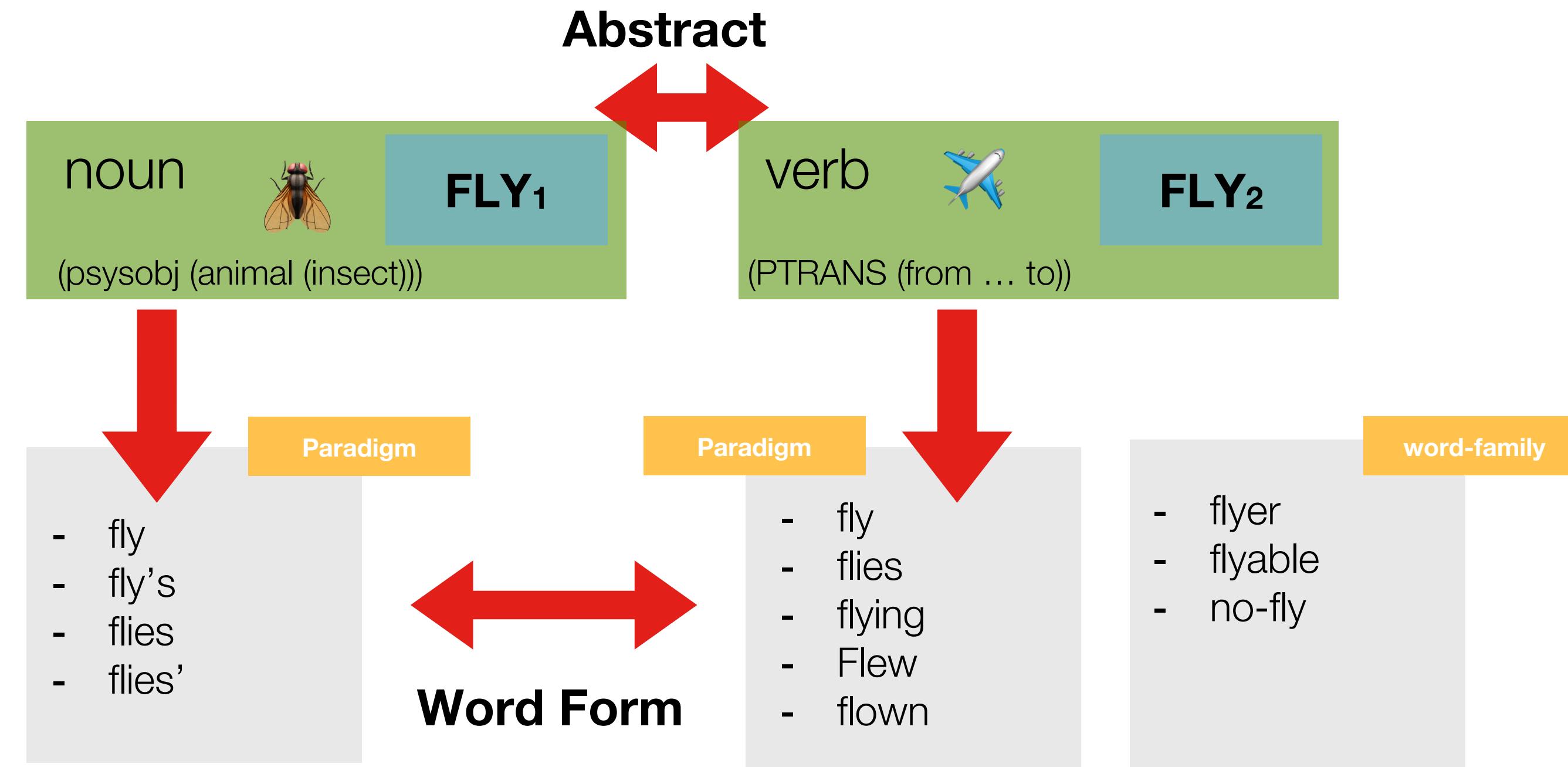
-ion -ly
-able -er

Plural -s
-ing -ed

stem	walk	kiss	map	cry
-s form	walks	kisses	maps	cries
-ing participle	walking	kissing	mapping	crying
Past form or -ed participle	walked	kissed	mapped	cried

Lexemes

- A fundamental unit of the lexicon of a language
 - An abstract vocabulary item which may be realised in different sets of grammatical variants
- The same word can have multiple meanings:
 - *bank, mean*
 - Extra challenge: domain-specific meanings

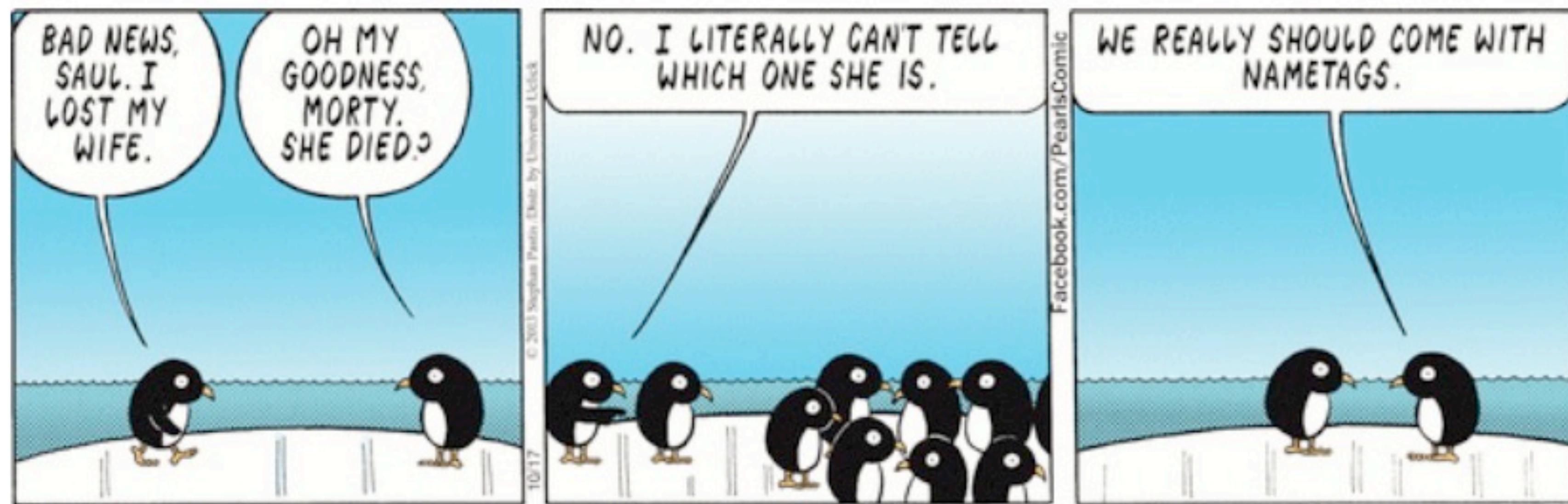


Lexical Items

- A single word, a part of a word, or a chain of words that forms the basic elements of a language's lexicon
- Examples of lexical items
 - **Lexemes** (*previous slide*)
 - **Phrasal verbs**, e.g. *put off, get out*
 - **Multiword expressions**, e.g. *by the way, inside out*
 - **Idioms**, e.g. *break a leg, a bitter pill to swallow*
 - **Sayings**, e.g. *The early bird gets the worm, The devil is in the details*

Lexical Ambiguity

- The presence of two or more possible meanings within a single word
 - Word sense ambiguity



credit: A. Zwicky

Part Of Speech

- The syntactic role of each word in a sentence

	Tag	Description	Example
Open Class	ADJ	Adjective: noun modifiers describing properties	<i>red, young, awesome</i>
	ADV	Adverb: verb modifiers of time, place, manner	<i>very, slowly, home, yesterday</i>
	NOUN	words for persons, places, things, etc.	<i>algorithm, cat, mango, beauty</i>
	VERB	words for actions and processes	<i>draw, provide, go</i>
	PROPN	Proper noun: name of a person, organization, place, etc..	<i>Regina, IBM, Colorado</i>
	INTJ	Interjection: exclamation, greeting, yes/no response, etc.	<i>oh, um, yes, hello</i>
Closed Class Words	ADP	Adposition (Preposition/Postposition): marks a noun's spacial, temporal, or other relation	<i>in, on, by, under</i>
	AUX	Auxiliary: helping verb marking tense, aspect, mood, etc.,	<i>can, may, should, are</i>
	CCONJ	Coordinating Conjunction: joins two phrases/clauses	<i>and, or, but</i>
	DET	Determiner: marks noun phrase properties	<i>a, an, the, this</i>
	NUM	Numeral	<i>one, two, first, second</i>
	PART	Particle: a preposition-like form used together with a verb	<i>up, down, on, off, in, out, at, by</i>
	PRON	Pronoun: a shorthand for referring to an entity or event	<i>she, who, I, others</i>
Other	SCONJ	Subordinating Conjunction: joins a main clause with a subordinate clause such as a sentential complement	<i>that, which</i>
	PUNCT	Punctuation	<i>;, , 0</i>
	SYM	Symbols like \$ or emoji	<i>\$, %</i>
	X	Other	<i>asdf, qwfg</i>

Always created

Relatively fixed

Part-Of-Speech /2

- **Nouns (NN, NNS)**: words for people, places, or things. Singular or plural
 - *cat, mango, algorithm, beauty, pacing*
- **Proper Nouns (NNP, NNPS)**: names of **specific persons** or **entities**
 - *Alessandro, Delft, TU Delft*
- **Adjectives**: describe the properties or qualities of nouns
 - e.g. colour (*white, black*), age (*old, young*), value (*good, bad*)
- **Verbs (VB)**: actions and processes
 - Multiple inflexions for singular/plural and verb tense
- **Personal and Possessive Pronouns (PRP)**: shorthand for referring to an entity or event
 - you, she, I, it, me, my, your, his, her, its, one's, our, their
- **Wh-pronouns**: used in questions
 - *what, who, whom, whoever*

Tag	Description	Example	Tag	Description	Example	Tag	Description	Example
CC	coord. conj.	<i>and, but, or</i>	NNP	proper noun, sing.	<i>IBM</i>	TO	"to"	<i>to</i>
CD	cardinal number	<i>one, two</i>	NNPS	proper noun, plu.	<i>Carolinas</i>	UH	interjection	<i>ah, oops</i>
DT	determiner	<i>a, the</i>	NNS	noun, plural	<i>llamas</i>	VB	verb base	<i>eat</i>
EX	existential 'there'	<i>there</i>	PDT	predeterminer	<i>all, both</i>	VBD	verb past tense	<i>ate</i>
FW	foreign word	<i>mea culpa</i>	POS	possessive ending	<i>'s</i>	VBG	verb gerund	<i>eating</i>
IN	preposition/ subordin-conj	<i>of, in, by</i>	PRP	personal pronoun	<i>I, you, he</i>	VBN	verb past participle	<i>eaten</i>
JJ	adjective	<i>yellow</i>	PRP\$	possess. pronoun	<i>your, one's</i>	VBP	verb non-3sg-pr	<i>eat</i>
JJR	comparative adj	<i>bigger</i>	RB	adverb	<i>quickly</i>	VBZ	verb 3sg pres	<i>eats</i>
JJS	superlative adj	<i>wildest</i>	RBR	comparative adv	<i>faster</i>	WDT	wh-determ.	<i>which, that</i>
LS	list item marker	<i>1, 2, One</i>	RBS	superlatv. adv	<i>fastest</i>	WP	wh-pronoun	<i>what, who</i>
MD	modal	<i>can, should</i>	RP	particle	<i>up, off</i>	WP\$	wh-possess.	<i>whose</i>
NN	sing or mass noun	<i>llama</i>	SYM	symbol	<i>+, %, &</i>	WRB	wh-adverb	<i>how, where</i>

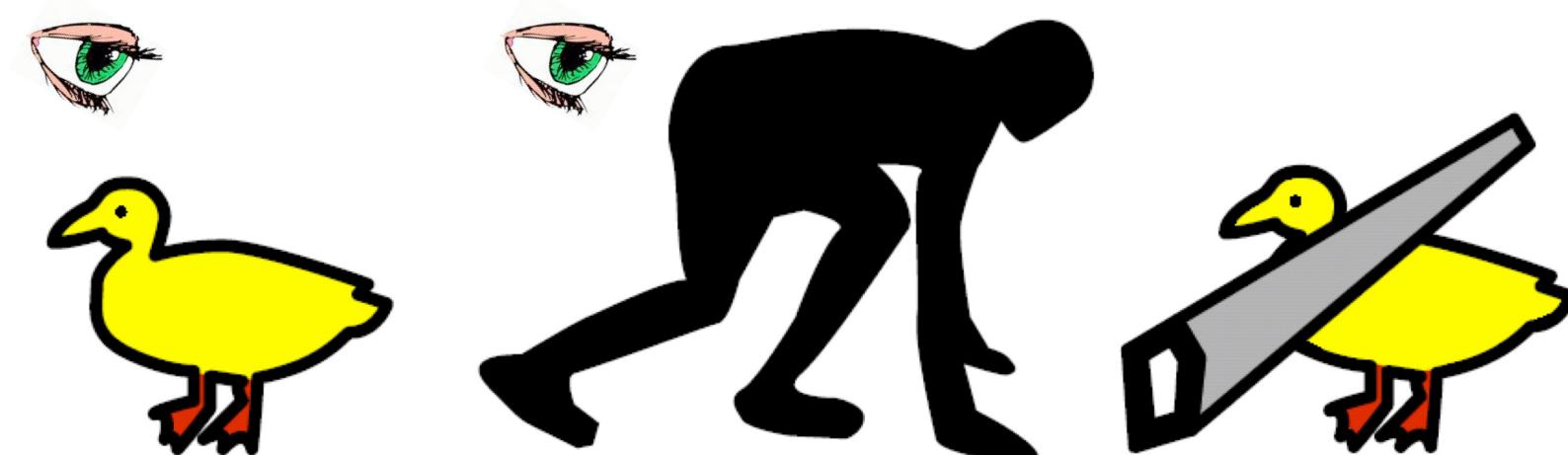
Syntax

- The syntax of a language is the set of principles (**rules**) under which sequences of words are judged to be grammatically acceptable by fluent speakers
- Basic syntactical elements (there are more)
 - **Constituents:** atomic tokens made up of a group of words
 - *Noun Phrase* (NP)
 - groups made up of nouns, determiners, adjectives, conjunctions
 - e.g *the big house, a red and large carpet*
 - *Verb Phrase* (VP)
 - A verb eventually followed by an NP or a prepositional phrase (PP)
 - e.g. *eat* (verb), *eat a pizza* (verb + NP), *eat a pizza with the fork* (verb + NP + PP)
 - **Grammatical Relations:** formalization of the sentence structure as a link between SUBJECTS and OBJECTS
 - es.[he]/SUBJECT took [thebighammer]/OBJECT

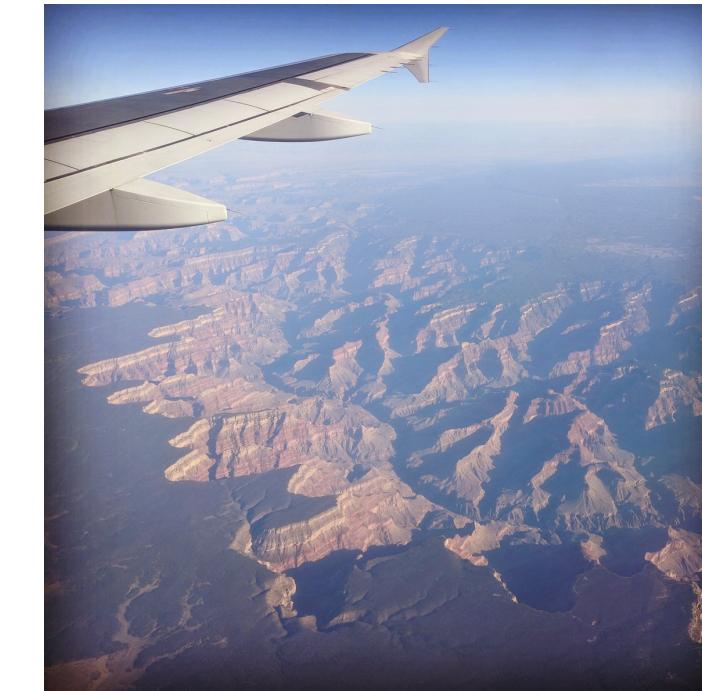
Syntactic Ambiguity

- The presence of two or more possible meanings within a single sentence or sequence of words
- They can be solved only at the semantic (or higher) level
 - Using statistical or semantic knowledge

I saw her duck



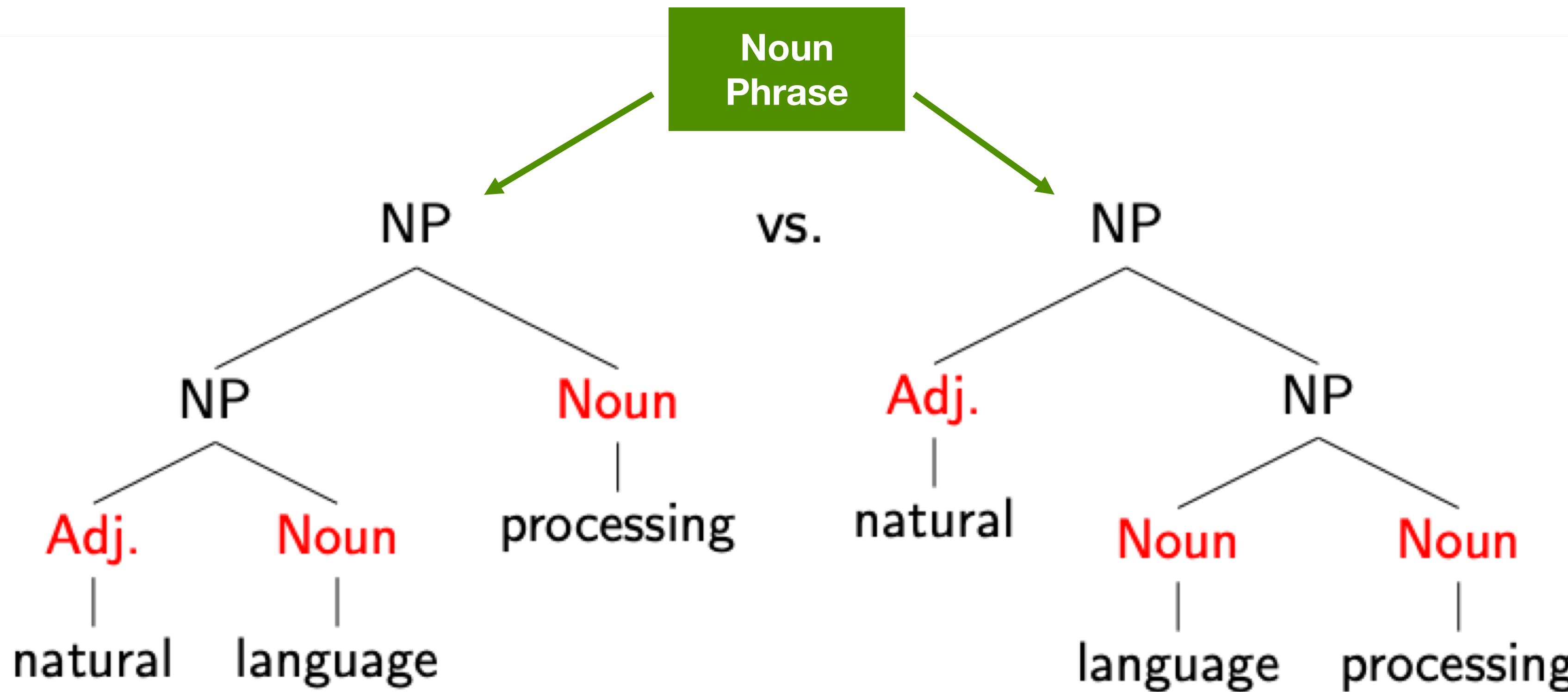
I saw the Grand Canyon flying to New York



Clearly the grand canyon does not fly....

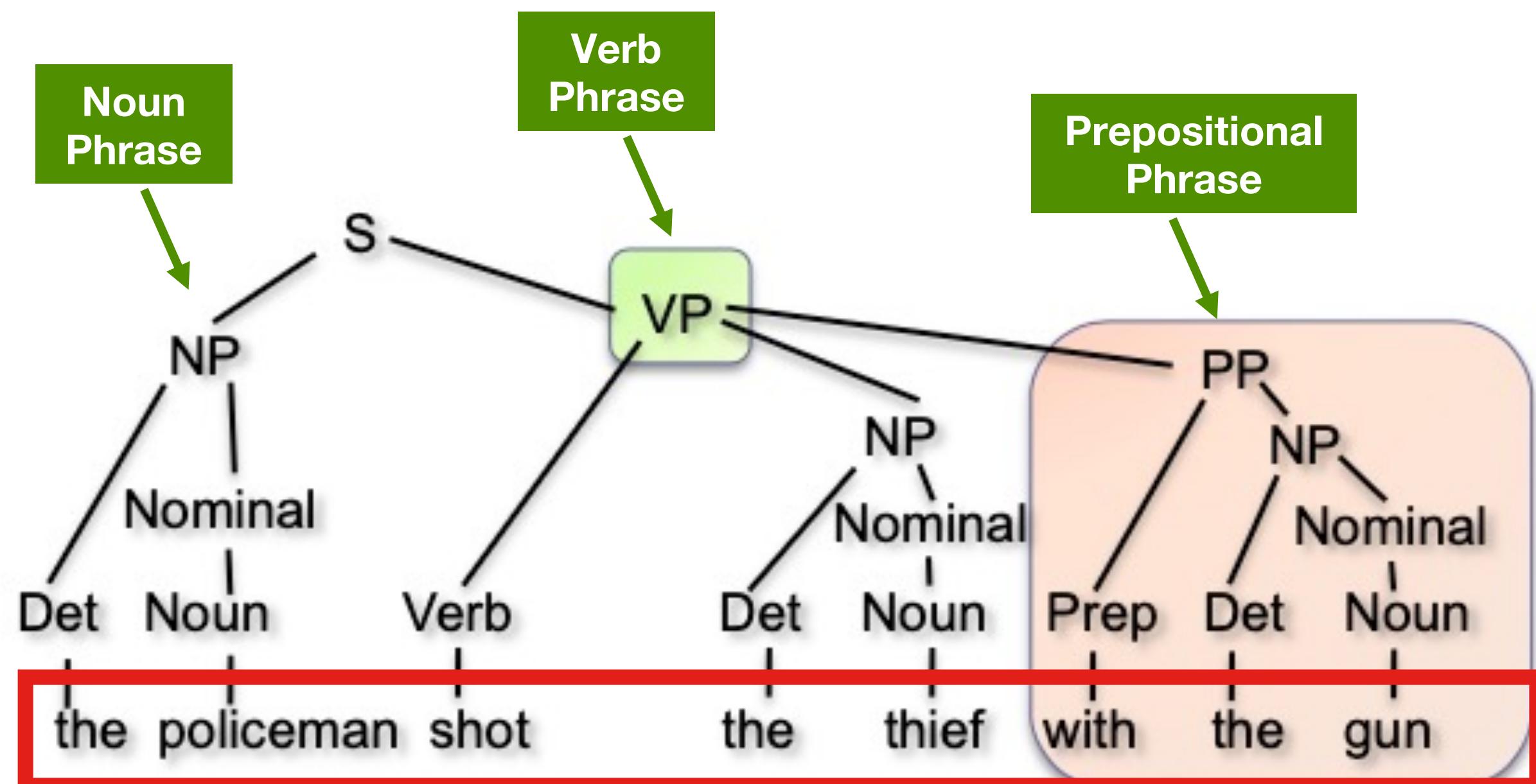
Syntactic Ambiguity

- Different structures lead to different interpretations

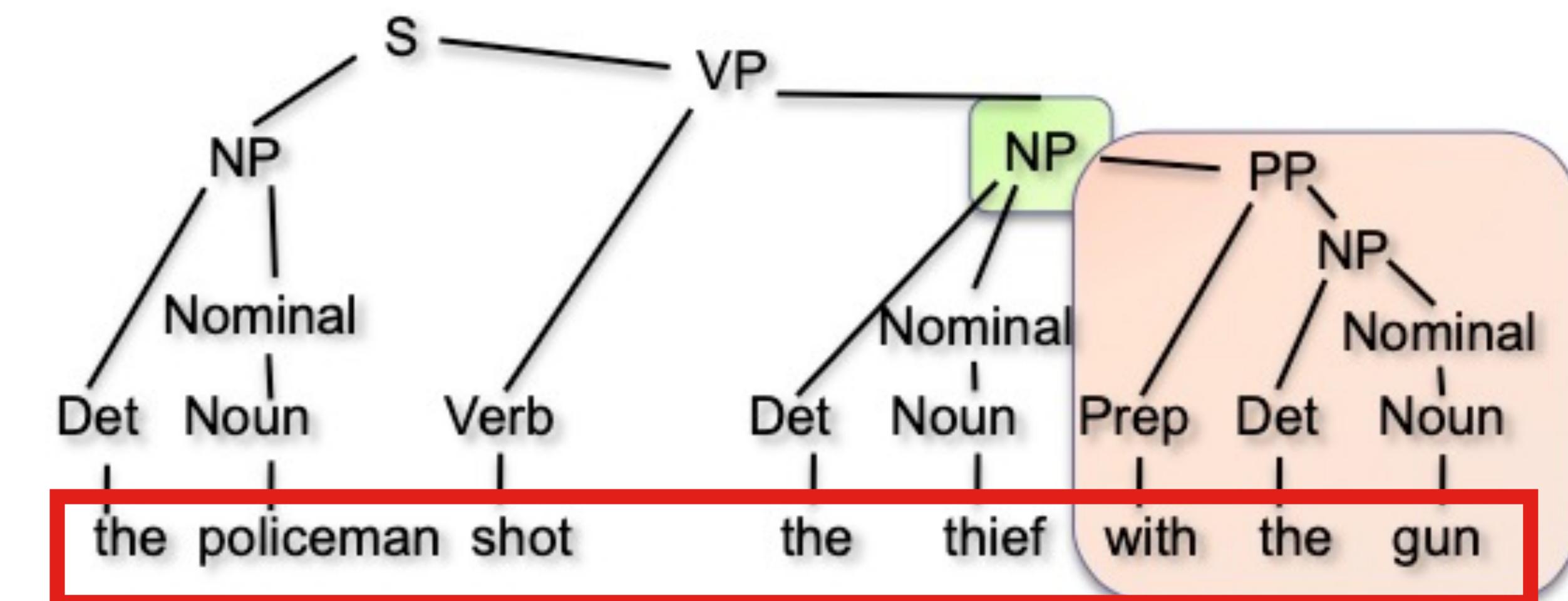


Attachment Ambiguity

The policeman shot the thief with the gun



The policeman used the gun to shoot the thief



The policeman shot a thief that had a gun

Pronoun reference ambiguity



Dr. Macklin often brings his dog Champion to visit with the patients. **He** just loves to give big, wet, sloppy kisses!

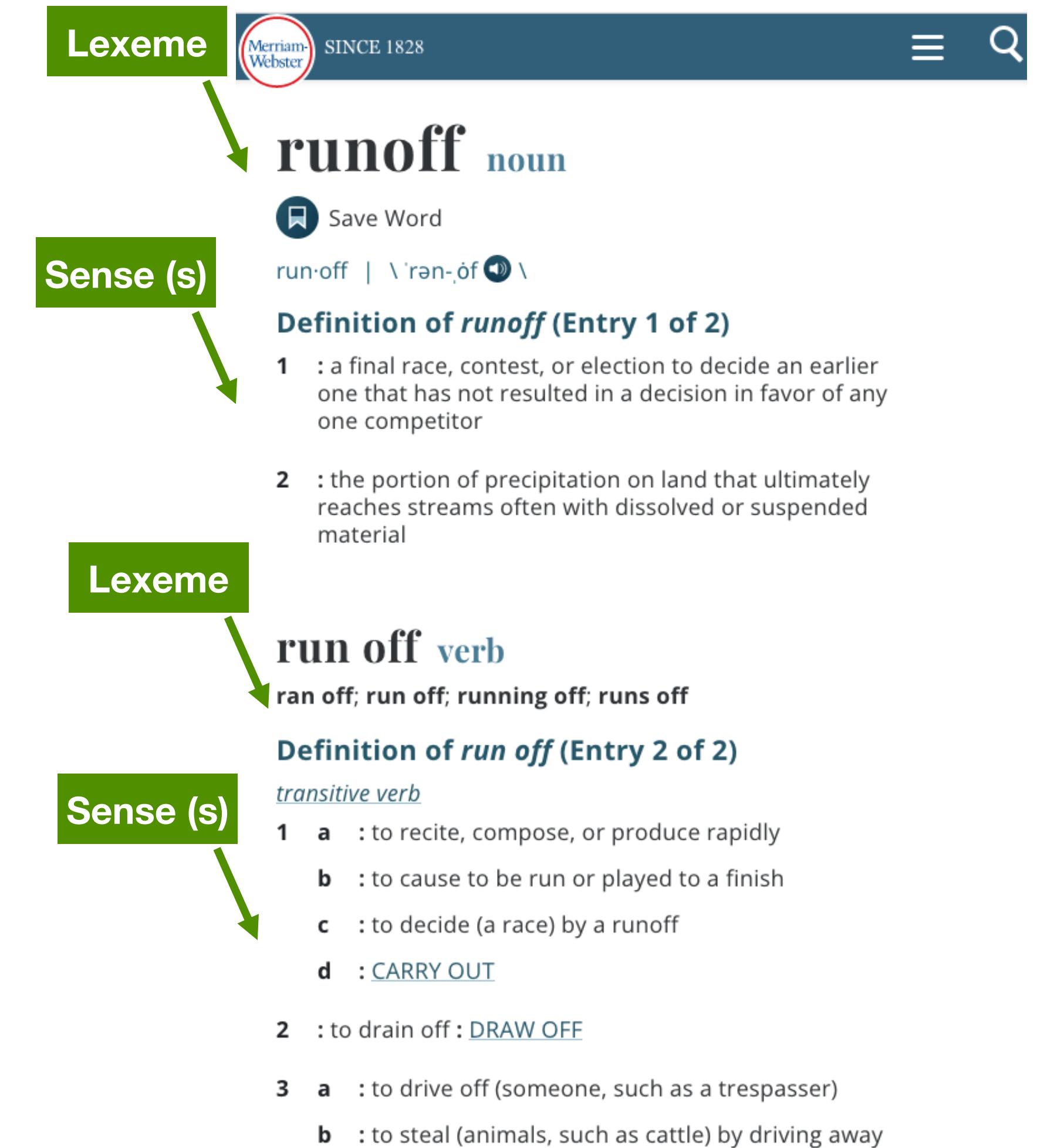
Semantics

- The study of the meaning of words (lexical semantics), and how these combine to form the meanings of sentences (compositional semantics)
- Mapping of natural language sentences into domain representations
 - E.g., a robot command language, a database query, or an expression in a formal logic



Lexical Semantics

- A **lexicon** (the vocabulary of a language) generally has a highly structured form
 - It stores the meanings and uses of each word
 - It encodes the relations between words and meanings
- A **lexeme** is a minimal unit represented in the lexicon. It pairs
 - A **stem**: the orthographic (or phonological) form chosen words (or, sometimes a lexical item)
 - A **sense**: a representation of one aspect of the meaning of a word
- A **dictionary** is a type of lexicon where meanings are expressed through definitions and examples



Lexical and semantic relations among words (senses)

▪ Homonymy

- Lexemes that have the **same form** (and the same PoS) but **unrelated meanings**
- e.g. bank (the financial institution, the river bank)

▪ Polysemy

- It happens when **a lexeme** has **more related meanings**
- It depends on the word etymology - unrelated meaning usually have a different origin)
- e.g. bank (the financial institution), bank (the building hosting the financial institution)

▪ Synonymy

- **distinct lexemes** with the **same meaning**
- e.g. fall, autumn; gift, present

▪ Hyponymy / Hypernymy (is-a relation) {parent: hypernym, child: hyponym}

- A relationship between **two senses** such that one denotes a subclass of the other
- e.g. dog, animal
- The relationship is not symmetric

▪ Holonymy / Meronymy (part-whole relation)

- A relationship between **two senses** such that one is structurally or logically part of the other
- E.g. arm → body (holonomy), bicycle → wheel (meronymy)
- The relationship is not symmetric

▪ Antonymy

- A relationship between two senses exists between words that have opposite meaning
- e.g. tall, short

Wordnet

- A hierarchical database of lexical relations
 - More than 200 languages
- Three Separate sub-databases
 - Nouns
 - Verbs
 - Adjectives and Adverbs
- Each lexeme is associated with a set of senses (synset)
- Synsets are linked by **conceptual**, **semantic** and **lexical** relationships
- Available online or for download
 - <http://wordnetweb.princeton.edu/perl/webwn>

POS	Unique	Synsets	Total
			Word-Sense Pairs
Noun	117798	82115	146312
Verb	11529	13767	25047
Adjective	21479	18156	30002
Adverb	4481	3621	5580
Totals	155287	117659	206941

Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	<i>breakfast</i> ¹ → <i>meal</i> ¹
Hyponym	Subordinate	From concepts to subtypes	<i>meal</i> ¹ → <i>lunch</i> ¹
Instance Hypernym	Instance	From instances to their concepts	<i>Austen</i> ¹ → <i>author</i> ¹
Instance Hyponym	Has-Instance	From concepts to concept instances	<i>composer</i> ¹ → <i>Bach</i> ¹
Member Meronym	Has-Member	From groups to their members	<i>faculty</i> ² → <i>professor</i> ¹
Member Holonym	Member-Of	From members to their groups	<i>copilot</i> ¹ → <i>crew</i> ¹
Part Meronym	Has-Part	From wholes to parts	<i>table</i> ² → <i>leg</i> ³
Part Holonym	Part-Of	From parts to wholes	<i>course</i> ⁷ → <i>meal</i> ¹
Substance Meronym		From substances to their subparts	<i>water</i> ¹ → <i>oxygen</i> ¹
Substance Holonym		From parts of substances to wholes	<i>gin</i> ¹ → <i>martini</i> ¹
Antonym		Semantic opposition between lemmas	<i>leader</i> ¹ ↔ <i>follower</i> ¹
Derivationally		Lemmas w/same morphological root	<i>destruction</i> ¹ ↔ <i>destroy</i> ¹
Related Form			

Noun Relations

Natural language processing tasks

Morphology /1 - Tokenisation

- Separation of words (or of morphemes) in a sentence
- Issues
 - Separators: punctuations
 - Exceptions: „m.p.h“, „Ph.D“
 - Expansions: „we're“ = „we are“
- Multi-words expressions: “New York”, “doghouse”

„Latest figures from the US government show the trade deficit with China reached an **all time** high of **\$ 365.7 bn (£ 250.1 bn)** last **year** . By February this year it had already reached **\$ 57 bn** .“

Morphology /2

■ Normalisation

- Sometimes we need to “normalize” terms
- We want to match U.S.A. and USA

■ Stopword removal

- Removal of high-frequency words, which carry less information
- E.g. determiners, prepositions
- English stop list is about 200-300 terms (e.g., “*been*”, “*a*”, “*about*”, “*otherwise*”, “*the*”, etc..)

any word		nouns	
Frequency	Token	Frequency	Token
1,698,599	the	124,598	European
849,256	of	104,325	Mr
793,731	to	92,195	Commission
640,257	and	66,781	President
508,560	in	62,867	Parliament
407,638	that	57,804	Union
400,467	is	53,683	report
394,778	a	53,547	Council
263,040	I	45,842	States

Morphology /3

■ Stemming

- Heuristic process that *chops* off the ends of words in the hope of achieving the goal correctly most of the time
- Stemming collapses derivationally related words
- Two basic types:
 - Algorithmic: uses programs to determine related words
 - Dictionary-based: uses lists of related words

Example of Stemming with Different Algorithms

Sample text: Such an analysis can reveal features that are not easily visible from the variations in the individual genes and can lead to a picture of expression that is more biologically transparent and accessible to interpretation

Lovins stemmer: such an analys can reve featur that ar not eas vis from th vari in th individu gen and can lead to a pictur of expres that is mor biolog transpar and acces to interpres

Porter stemmer: such an analysi can reveal featur that ar not easili visibl from the variat in the individu gene and can lead to a pictur of express that is more biolog transpar and access to interpret

Paice stemmer: such an analys can rev feat that are not easy vis from the vary in the individ gen and can lead to a pict of express that is mor biolog transp and access to interpret

Morphology /4

■ Lemmatisation

- It uses dictionaries and morphological analysis of words in order to return the base or dictionary form of a word
- Lemmatization collapses the different inflectional forms of a lemma
- Example: Lemmatization of “saw”
—> attempts to return “see” or “saw” depending on whether the use of the token is a verb or a noun

Google, headquartered in Mountain View (1600 Amphitheatre Pkwy, Mountain View, CA 940430), unveiled the new Android phone for \$799 at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

Google , headquartered in Mountain View (1600 Amphitheatre Pkwy , Mountain View ,
headquarter
Sundar Pichai said in his keynote that users love their new Android phones .
say user phone

Syntax: Part-Of-Speech Tagging

■ Why do we care?

- Text-to-speech:
record[v] and *record[n]*
- Lemmatization:
 - *saw[v]* → *see*
 - *saw[n]* → *saw*
- As input for many other NLP tasks
 - Chunking
 - Named entity recognition
 - Information extraction

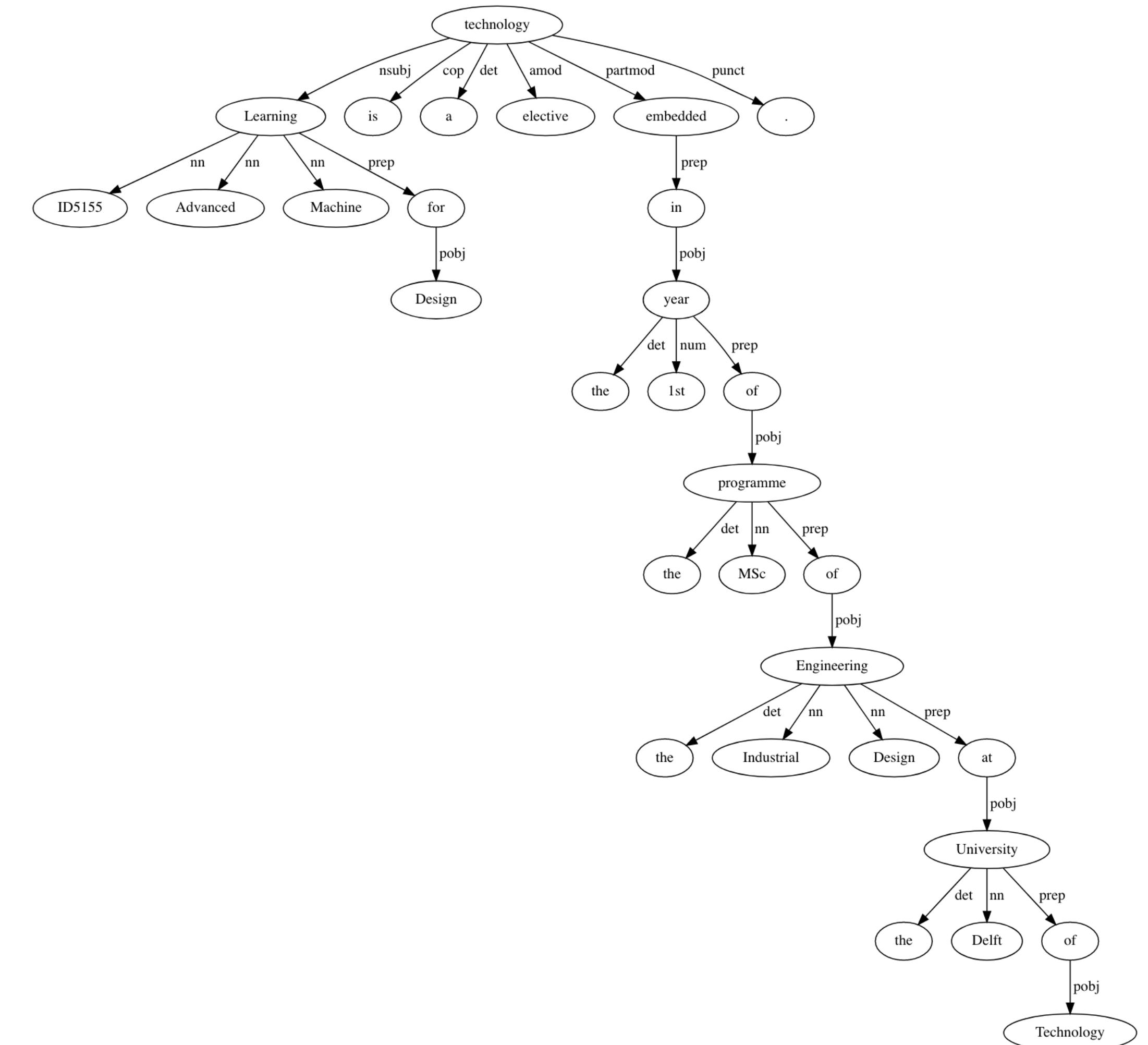
Google, headquartered in Mountain View (1600 Amphitheatre Pkwy, Mountain View, CA 940430), unveiled the new Android phone for \$799 at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

nsubj	p	vmod	prep	nn	pobj	p	num	nn	appos	p	
Google	,	headquartered	in	Mountain	View	(1600	Amphitheatre	Pkwy	,	
NOUN	PUNCT	VERB	ADP	NOUN	NOUN	PUNCT	NUM	NOUN	NOUN	PUNCT	
nn	appos	p	appos	num	p	p	root	det	amod	nn	
Mountain	View	,	CA	940430)	,	unveiled	the	new	Android	
NOUN	NOUN	PUNCT	NOUN	NUM	PUNCT	PUNCT	VERB	DET	ADJ	NOUN	
pobj	prep	det	nn	nn	pobj	p				dobj	
\$799	at	the	Consumer	Electronic	Show	.				prep	
NUM	ADP	DET	NOUN	NOUN	NOUN	PUNCT				for	
nn	nsubj	root	prep	poss	pobj	mark	nsubj	ccomp	poss	amod	nn
Sundar	Pichai	said	in	his	keynote	that	users	love	their	new	Android
NOUN	NOUN	VERB	ADP	PRON	NOUN	ADP	NOUN	VERB	PRON	ADJ	NOUN
										phones	NOUN

<https://cloud.google.com/natural-language#section-2>

Syntax: Dependency Parsing

ID5155 Advanced Machine Learning for Design is a technology elective embedded in the 1st year of the MSc programme of the Industrial Design Engineering at the Delft University of Technology.



<https://www.textrazor.com/demo>

Syntax: Part-Of-Speech Tagging /2

Helicopters will patrol the temporary no-fly zone around New Jersey's MetLife Stadium Sunday, with F-16s based in Atlantic City ready to be scrambled if an unauthorized aircraft does enter the restricted airspace. Helicopters will patrol the temporary no-fly zone around New Jersey's MetLife Stadium Sunday, with F-16s based in Atlantic City ready to be scrambled if an unauthorized aircraft does enter the restricted airspace. Down below, bomb-sniffing dogs will patrol the trains and buses that are expected to take approximately 30,000 of the 80,000-plus spectators to Sunday's Super Bowl between the Denver Broncos and Seattle Seahawks. Helicopters will patrol the temporary no-fly zone around New Jersey's MetLife Stadium Sunday, with F-16s based in Atlantic City ready to be scrambled if an unauthorized aircraft does enter the restricted airspace. Down below, bomb-sniffing dogs will patrol the trains and buses that are expected to take approximately 30,000 of the 80,000-plus spectators to Sunday's Super Bowl between the Denver Broncos and Seattle Seahawks.

NNPS/ Helicopters MD/ will NN/ patrol DT/ the JJ/ temporary JJ/ no-fly NN/ zone IN/ around NNP/ New NNP/ Jersey POS/ 's NNP/ MetLife NNP/ Stadium NNP/ Sunday ./, IN/ with NNP/ F-16s VBN/ based IN/ in NNP/ Atlantic NNP/ City JJ/ ready TO/ to VB/ be VBN/ scrambled IN/ if DT/ an JJ/ unauthorized NN/ aircraft VBZ/ does VB/ enter DT/ the VBN/ restricted NN/ airspace ./.

IN/ Down IN/ below ./, JJ/ bomb-sniffing NNS/ dogs MD/ will NN/ patrol DT/ the NNS/ trains CC/ and NNS/ buses WDT/ that VBP/ are VBN/ expected TO/ to VB/ take RB/ approximately CD/ 30,000 IN/ of DT/ the JJ/ 80,000-plus NNS/ spectators TO/ to NNP/ Sunday POS/ 's NNP/ Super NNP/ Bowl IN/ between DT/ the NNP/ Denver NNS/ Broncos CC/ and NNP/ Seattle NNP/ Seahawks ./.

Syntax: Named Entity Recognition

- Factual information and knowledge are normally expressed by named entities
 - Who, Whom, Where, When, Which, ...
 - It is the core of the information extraction systems

1. Identify words that refer to **proper names** of interest in a particular application

- E.g. people, companies, locations, dates, product names, prices, etc.

2. Classify them to the corresponding classes (e.g. person, location)

3. Assign a unique identifier from a database

Google, headquartered in Mountain View (1600 Amphitheatre Pkwy, Mountain View, CA 940430), unveiled the new Android phone for \$799 at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

<Google>₁ , headquartered in <Mountain View>₂ (<1600 Amphitheatre Pkwy, Mountain View, CA>₁₂ <1600>₁₄ <Amphitheatre Pkwy>₇ , <Mountain View>₂ , <CA 940430>₈ <940430>₁₆), unveiled the new <Android>₃ <phone>₅ for <\$799>₁₃ <799>₁₅ at the <Consumer Electronic Show>₁₁ . <Sundar Pichai>₄ said in his <keynote>₉ that <users>₆ love their new <Android>₃ <phones>₁₀ .

1. Google	ORGANIZATION
Wikipedia Article	
Salience: 0.19	
2. Mountain View	LOCATION
Wikipedia Article	
Salience: 0.18	
3. Android	CONSUMER GOOD
Wikipedia Article	
Salience: 0.14	
4. Sundar Pichai	PERSON
Wikipedia Article	
Salience: 0.11	
5. phone	CONSUMER GOOD
Salience: 0.10	
6. users	PERSON
Salience: 0.09	
7. Amphitheatre Pkwy	LOCATION
Salience: 0.07	
8. CA 940430	OTHER
Salience: 0.05	

<https://cloud.google.com/natural-language#section-2>

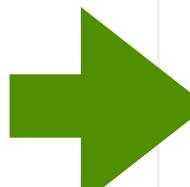
Document Categorisation / Topic Modeling

■ Categorisation

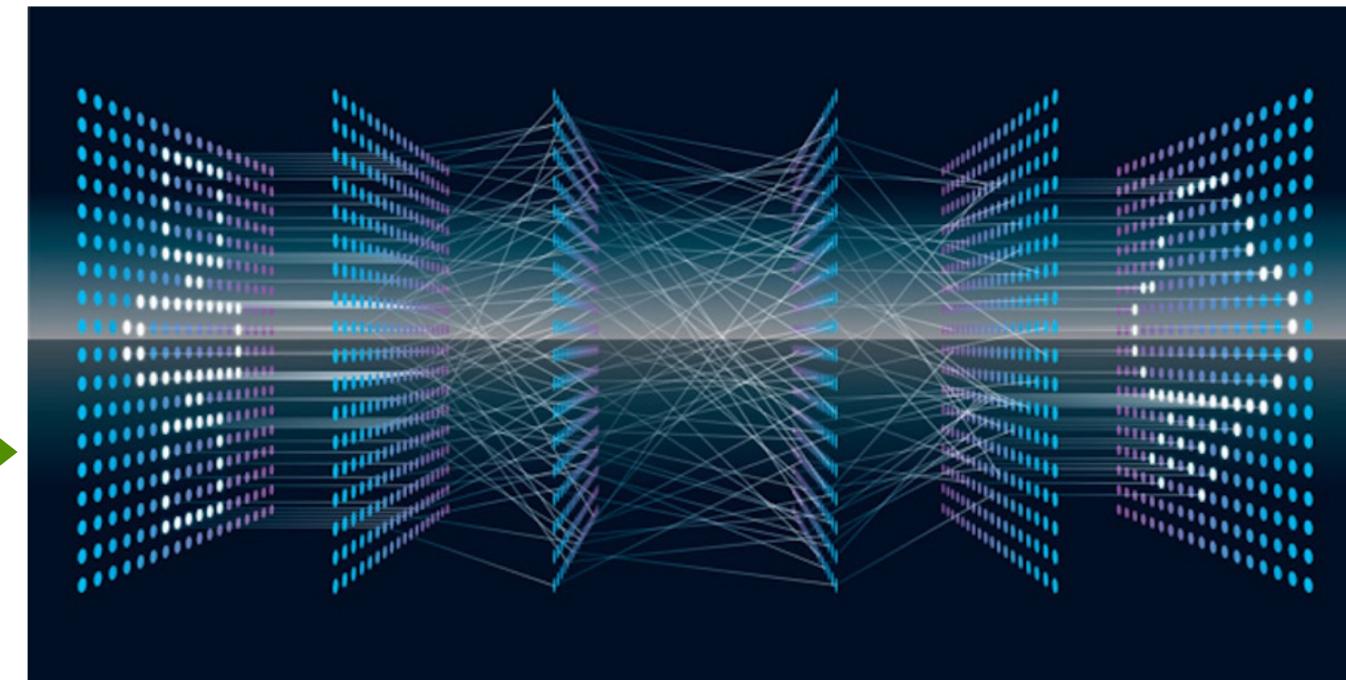
- assigning a label or category to an entire text or document
- Supervised learning
- For instance
 - Spam vs. Not spam
 - Language identification
 - Authors attribution
 - Assigning a library subject category or topic label

■ Topic Modeling

- A topic is the subject or theme of a discourse
- Topic modeling: group documents/text according to their (semantic) similarity
- An unsupervised machine learning approach



Welcome to the 2022/2023 Edition of the Advanced Machine Learning for Design Course



The Course

The elective of **ID5515 Advanced Machine Learning for Design (AML4D)** is embedded in the 1st year of the *Integrated Product Design (IPD)* MSc programme.

This advanced technology elective will provide students with the knowledge required to understand, design, and evaluate machine learning systems in the context of the design of intelligent products, services, and systems (iPSSs). Machine Learning (ML) is a computational approach that aims at "giving computers the ability to learn without being explicitly programmed" (A. Samuel, 1959). Smart thermostats, voice-enabled personal assistants, autonomous vehicles, traffic control systems, online social networks, web-shopping platforms, content-creation platforms, personal-health applications are just a few examples of iPSSs powered by ML technology. Consequently, ML technology is

CATEGORIES
0.85 science and technology
0.58 education
0.58 economy, business and finance>economic sector>computing and information technology
0.57 society
0.54 science and technology>social sciences>psychology
0.54 economy, business and finance>economic sector>media
0.54 society>values>ethics
0.49 education>school>further education
0.43 economy, business and finance>economic sector>computing and information technology>software
0.43 science and technology>social sciences>philosophy

TOPICS
1.00 Technology
1.00 Machine learning
1.00 Design
1.00 Learning
1.00 System
1.00 Social networking service
1.00 Cognition
1.00 Human activities
1.00 Branches of science
1.00 Communication
1.00 Cognitive science
1.00 Education
0.93 Educational psychology
0.93 Self-driving car
0.89 Engineering
0.85 Systems science
0.84 Social network
0.84 Computing
0.83 Behavior modification
0.82 Machine
0.82 Concepts in metaphysics
0.78 Reason
0.77 Neuropsychological assessment
0.77 Change
0.76 Interdisciplinary subfields
0.75 Psychological concepts
0.75 Science
0.75 World Wide Web
0.75 Society
0.74 Academic discipline interactions
0.73 Experience
0.70 Cyberspace
0.70 Content creation
0.69 Applied psychology
0.67 Neuroscience
0.67 Bias

Syntax: Sentiment Analysis

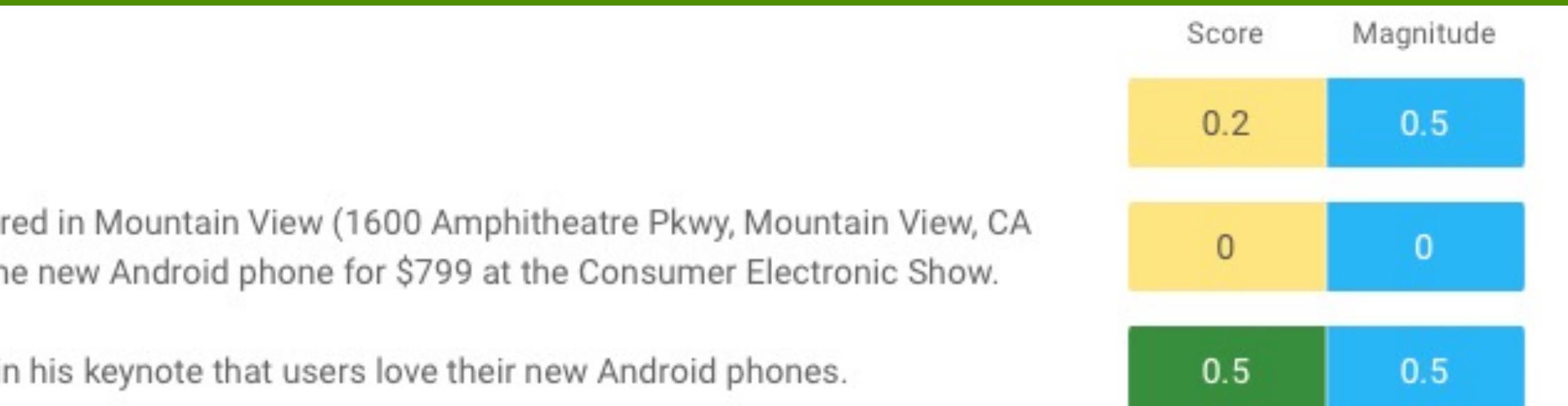
- The detection of attitudes
 - “*enduring, affectively colored beliefs, dispositions towards objects or persons*”
- Main elements
 - Holder (source)
 - Target (aspect)
 - Type of attitude
 - Text containing the attitude
- Tasks
 - **Classification:** Is the attitude of the text positive or negative?
 - **Regression:** Rank the attitude of the text from 1 to 5
 - **Advanced:** Detect the target, source, or complex attitude types

Google, headquartered in Mountain View (1600 Amphitheatre Pkwy, Mountain View, CA 940430), unveiled the new Android phone for \$799 at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

Entire Document

Google, headquartered in Mountain View (1600 Amphitheatre Pkwy, Mountain View, CA 940430), unveiled the new Android phone for \$799 at the Consumer Electronic Show.

Sundar Pichai said in his keynote that users love their new Android phones.



Score Range 0.25 – 1.0 -0.25 – 0.25 -1.0 – -0.25

3. Android

Sentiment: Score 0.2 Magnitude 0.5

CONSUMER GOOD

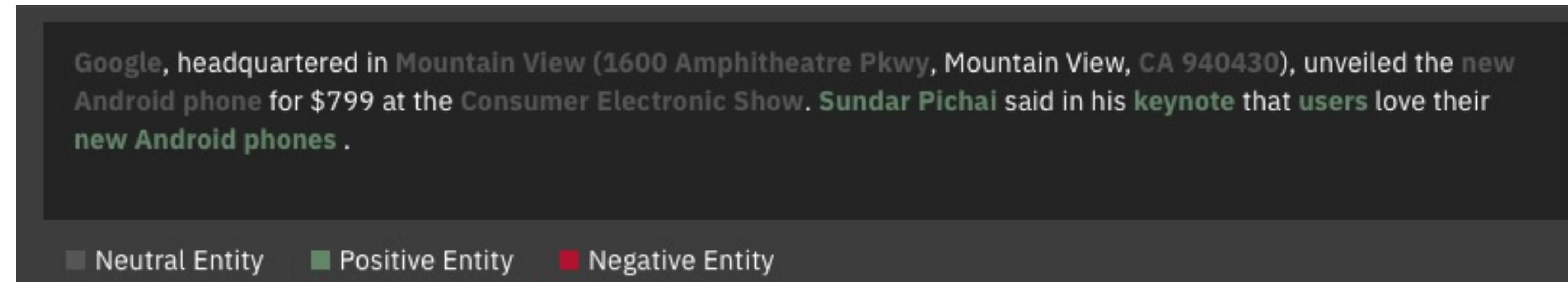
4. Sundar Pichai

Sentiment: Score 0.4 Magnitude 0.9

PERSON

Syntax: Sentiment Analysis / IBM Demo

Google, headquartered in Mountain View (1600 Amphitheatre Pkwy, Mountain View, CA 940430), unveiled the new Android phone for \$799 at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.



Sentiment Emotion Categories

Full Document

POSITIVE

0.85



Entity Sentiment Scores

Mountain View (1600 Amph...
940430
Consumer Electronic Show
Mountain View
Sundar Pichai
Google
Android
CA

NEUTRAL

0



NEUTRAL

0



NEUTRAL

0



POSITIVE

0.85



NEUTRAL

0



NEUTRAL

0



NEUTRAL

0

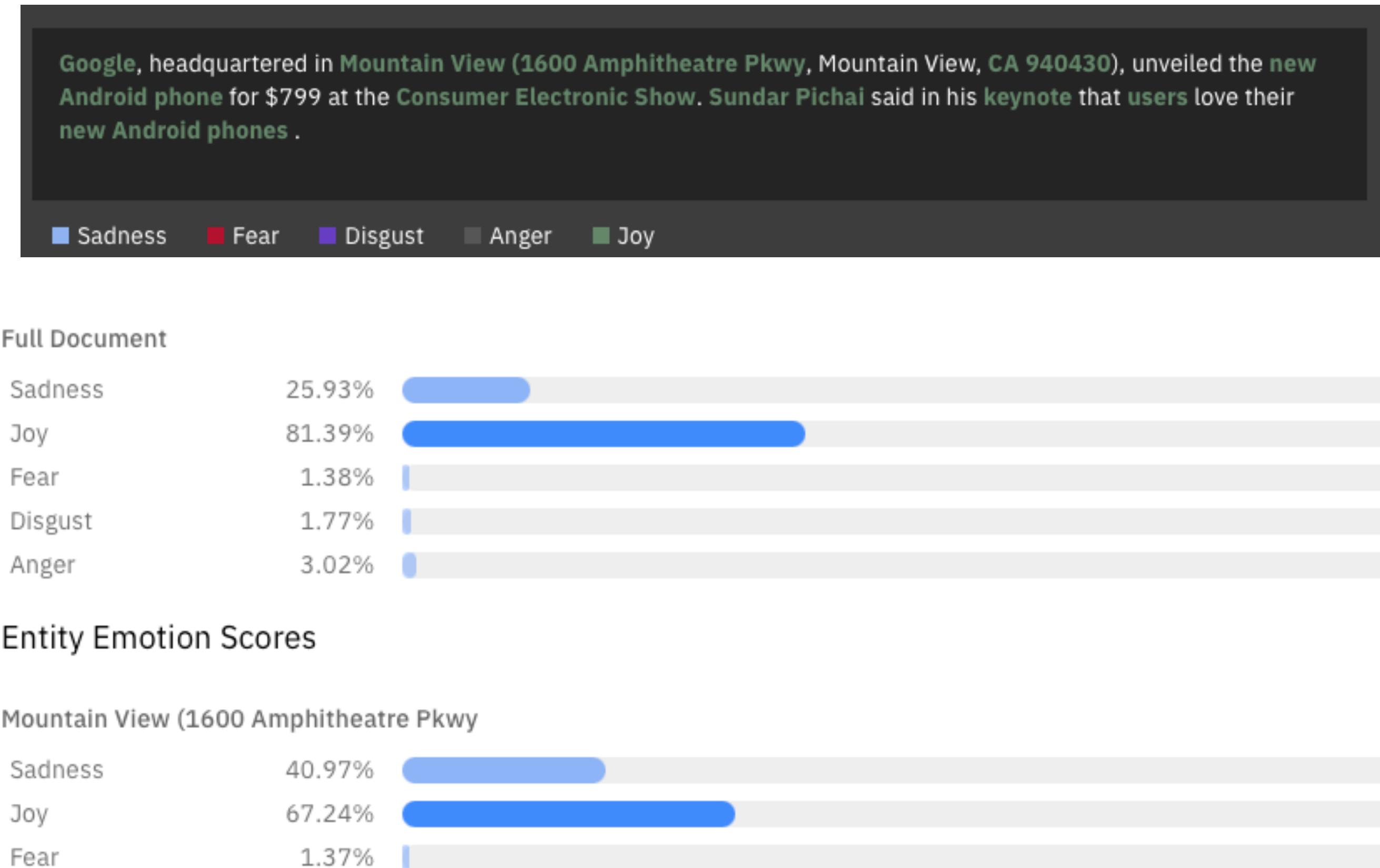


<https://www.ibm.com/demos/live/natural-language-understanding/self-service/home>

Syntax: Emotion Analysis / IBM Demo

Google, headquartered in Mountain View (1600 Amphitheatre Pkwy, Mountain View, CA 940430), unveiled the new Android phone for \$799 at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

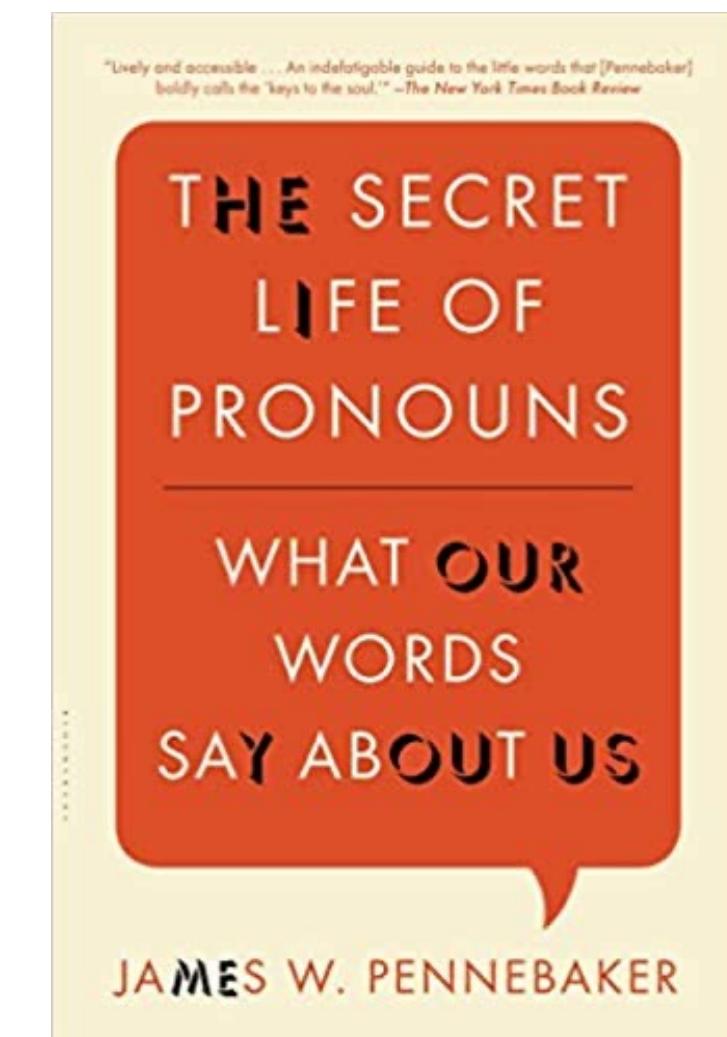
Detects anger, disgust, fear, joy, or sadness that is conveyed in the content or by the context around target phrases specified in the targets parameter.



<https://www.ibm.com/demos/live/natural-language-understanding/self-service/home>

Syntax - Language Analysis

- Idea: people's language can provide insights into their psychological states (emotions, thinking style, etc)
- For instance
 - *Frequency of words associated with positive or negative emotions*
 - *Use of pronouns as a proxy for confidence and character traits*
- **Analytic Thinking:** the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns.
 - low Analytical Thinking —> language that is more intuitive and personal
- **Clout:** the relative social status, confidence, or leadership that people display through their writing or talking
- **Authenticity:** the degree to which a person is self-monitoring
 - Low authenticity: prepared texts (i.e., speeches that were written ahead of time) and texts where a person is being socially cautious.
- **Emotional tone:** the higher the number, the more positive the tone. Numbers below 50 suggest a more negative emotional tone.



Category	Abbrev.	Description/Most frequently used exemplars
Summary Variables		
Word count	WC	Total word count
Analytical thinking	Analytic	Metric of logical, formal thinking
Clout	Clout	Language of leadership, status
Authentic	Authentic	Perceived honesty, genuineness
Emotional tone	Tone	Degree or positive (negative) tone
Words per sentence	WPS	Average words per sentence
Big words	BigWords	Percent words 7 letters or longer
Dictionary words	Dic	Percent words captured by LIWC
Linguistic Dimensions		
Total function words	function	the, to, and, I
Total pronouns	pronoun	I, you, that, it
Personal pronouns	ppron	I, you, my, me
1st person singular	i	I, me, my, myself
1st person plural	we	we, our, us, lets
2nd person	you	you, your, u, yourself
3rd person singular	shehe	he, she, her, his
3rd person plural	they	they, their, them, themsel*
Impersonal pronouns	ipron	that, it, this, what
Determiners	det	the, at, that, my
Articles	article	a, an, the, alot
Numbers	number	one, two, first, once
Prepositions	prep	to, of, in, for
Auxiliary verbs	auxverb	is, was, be, have
Adverbs	adverb	so, just, about, there
Conjunctions	conj	and, but, so, as
Negations	negate	not, no, never, nothing
Common verbs	verb	is, was, be, have
Common adjectives	adj	more, very, other, new
Quantities	quantity	all, one, more, some

Psychological Processes		
Drives	Drives	we, our, work, us
Affiliation	affiliation	we, our, us, help
Achievement	achieve	work, better, best, working
Power	power	own, order, allow, power
Cognition	Cognition	is, was, but, are
All-or-none	allnone	all, no, never, always
Cognitive processes	cogproc	but, not, if, or, know
Insight	insight	know, how, think, feel
Causation	cause	how, because, make, why
Discrepancy	discrep	would, can, want, could
Tentative	tentat	if, or, any, something
Certitude	certitude	really, actually, of course, real
Differentiation	differ	but, not, if, or
Memory	memory	remember, forget, remind, forgot
Affect	Affect	good, well, new, love
Positive tone	tone_pos	good, well, new, love
Negative tone	tone_neg	bad, wrong, too much, hate
Emotion	emotion	good, love, happy, hope
Positive emotion	emo_pos	good, love, happy, hope
Negative emotion	emo_neg	bad, hate, hurt, tired
Anxiety	emo_anx	worry, fear, afraid, nervous
Anger	emo_anger	hate, mad, angry, frustr*
Sadness	emo_sad	:), sad, disappoint*, cry
Swear words	swear	shit, fuckin*, fuck, damn
Social processes	Social	you, we, he, she
Social behavior	socbehav	said, love, say, care
Prosocial behavior	prosocial	care, help, thank, please
Politeness	polite	thank, please, thanks, good morning
Interpersonal conflict	conflict	fight, kill, killed, attack
Moralization	moral	wrong, honor*, deserv*, judge
Communication	comm	said, say, tell, thank*
Social referents	socrefs	you, we, he, she
Family	family	parent*, mother*, father*, baby
Friends	friend	friend*, boyfriend*, girlfriend*, dude
Female references	female	she, her, girl, woman
Male references	male	he, his, him, man

The AMLFD Course Manual (page 1)

RESULTS

Traditional LIWC Dimension	Your Text	Average for Formal Language
I-words (I, me, my)	0.00	0.67
Positive Tone	2.18	2.33
Negative Tone	0.00	1.38
Social Words	3.93	6.54
Cognitive Processes	17.03	7.95
Allure	2.62	3.58
Moralization	0.44	0.30
Summary Variables		
Analytic	86.21	87.63
Authentic	10.97	28.90

<https://www.liwc.app>

Semantics: Word Sense Disambiguation

- Multiple words can be spelt the same way (homonymy)
- The same word can also have different, related senses (polysemy)
- Disambiguation depends on context!

The human brain is quite proficient at word-sense disambiguation. That natural language is formed in a way that requires so much of it is a reflection of that neurologic reality. In computer science and the information technology that it enables, it has been a long-term challenge to develop the ability in computers to do natural language processing and machine learning

brain%1:08:00:: (36% probability)

encephalon (That part of the central nervous system that includes all the higher nervous centers; enclosed within the skull; continuous with the spinal cord)

in_a_way%4:02:00:: (100% probability)

in_a_way (From some points of view)

The human brain is quite proficient at word-sense disambiguation . That natural_language is formed in_a_way that requires so much of it is a reflection of that neurologic reality . In computer_science and the information_technology that it enables , it has been a long-term challenge to develop the ability in computers to do natural_language_processing and machine learning .

machine%1:18:00:: (28% probability)

machine (An efficient person)

learning%1:09:02:: (50% probability)

learning (Profound scholarly knowledge)

Question Answering: IBM's Watson

- Won Jeopardy on February 16, 2011



William Wilkinson's
“An account of the principalities of Wallachia and Moldavia”
inspired this author's
most famous novel



Bram Stoker

Automated Summarisation

- Condensing a piece of text to a shorter version while preserving key informational elements and the meaning of content
- A very difficult task!

Text Summarization Result

Original URL/Text	Summarized Text
IOB4-T3 Machine Learning for Design is a technology elective embedded in the 2nd year of the Bachelor of Industrial Design Engineering at the Delft University of Technology. The course provides students with the knowledge required to understand, design, and evaluate machine learning systems in the context of the design of intelligent products, services, and systems (iPSSs). Machine learning (ML) is a computational approach that aims at "giving computers the ability to learn without being explicitly programmed" (A. Samuel, 1959). Smart thermostats, voice-based personal assistants, autonomous vehicles, traffic control systems, online social networks, web shopping platforms, content creation platforms, personal health appliances: much of current and future iPSSs are powered by ML technology, influencing, and shaping our interests, habits, lives, and society. To meaningfully envision and design future iPSSs that are beneficial and useful to people and society, designers must: engage with the details of how ML systems "see" the world, "reason" about it, and interact with it experience the quirks, biases, and failures of ML technology; contend with how agency, initiative, trust, and explainability mediate the interaction between human and iPSSs; and understand how functionalities enabled by ML can be designed in iPSSs. Students in this course gain practical experience with ML technology and learn how to think critically of what ML systems can do, and how they could and should be integrated in iPSSs.	IOB4-T3 Machine Learning for Design is a technology elective embedded in the 2nd year of the Bachelor of Industrial Design Engineering at the Delft University of Technology. The course provides students with the knowledge required to understand, design, and evaluate machine learning systems in the context of the design of intelligent products, services, and systems (iPSSs). Machine learning (ML) is a computational approach that aims at "giving computers the ability to learn without being explicitly programmed" (A. Samuel, 1959). Smart thermostats, voice-based personal assistants, autonomous vehicles, traffic control systems, online social networks, web shopping platforms, content creation platforms, personal health appliances: much of current and future iPSSs are powered by ML technology, influencing, and shaping our interests, habits, lives, and society. To meaningfully envision and design future iPSSs that are beneficial and useful to people and society, designers must: engage with the details of how ML systems "see" the world, "reason" about it, and interact with it experience the quirks, biases, and failures of ML technology; contend with how agency, initiative, trust, and explainability mediate the interaction between human and iPSSs; and understand how functionalities enabled by ML can be designed in iPSSs. Students in this course gain practical experience with ML technology and learn how to think critically of what ML systems can do, and how they could and should be integrated in iPSSs.

<https://textsummarization.net/text-summarizer>

Result

After pressing the "Summarize" button above, the result will be displayed in the box below.

The summarized text will be here...

IOB4-T3 Machine Learning for Design is a technology optional embedded in the 2nd year of the Bachelor of Industrial Design Engineering at the Delft University of Technology. Machine learning is a computational approach that focuses on "offering computer systems the capacity to learn without being explicitly configured". Students in this course gain useful experience with ML innovation and learn just how to think seriously of what ML systems can do, and just how they could and should be integrated in iPSSs.

<https://brevi.app/single-demo>

Stance Detection

EXAMPLE HEADLINE

"Robert Plant Ripped up \$800M Led Zeppelin Reunion Contract"

EXAMPLE SNIPPETS FROM BODY TEXTS AND CORRECT CLASSIFICATIONS

"... *Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ...*"

CORRECT CLASSIFICATION: AGREE

"... *No, Robert Plant did not rip up an \$800 million deal to get Led Zeppelin back together. ...*"

CORRECT CLASSIFICATION: DISAGREE

- **Input:** Headline + text
- **Output:** Classify stance (e.g., agrees, disagrees, discusses, unrelated)

Machine Translation (not perfect)

The image displays three separate instances of a machine translation web application, illustrating the process of translating between English and German.

Top Translation (English to German):

- Source: I study advanced machine learning for design
- Target: Ich studiere fortgeschrittenes maschinelles Lernen für Design
- Language Detection: DETECT LANGUAGE (ENGLISH)
- Target Language: GERMAN
- Buttons: Microphone, Speaker, Progress (44 / 5,000), Keyboard, Share, Like, Star

Middle Translation (German to English):

- Source: Ich studiere intensives maschinelles Lernen für Design
- Target: I'm studying intensive machine learning for design
- Language Detection: DETECT LANGUAGE (GERMAN)
- Target Language: ENGLISH
- Buttons: Microphone, Speaker, Progress (54 / 5,000), Keyboard, Share, Like, Star

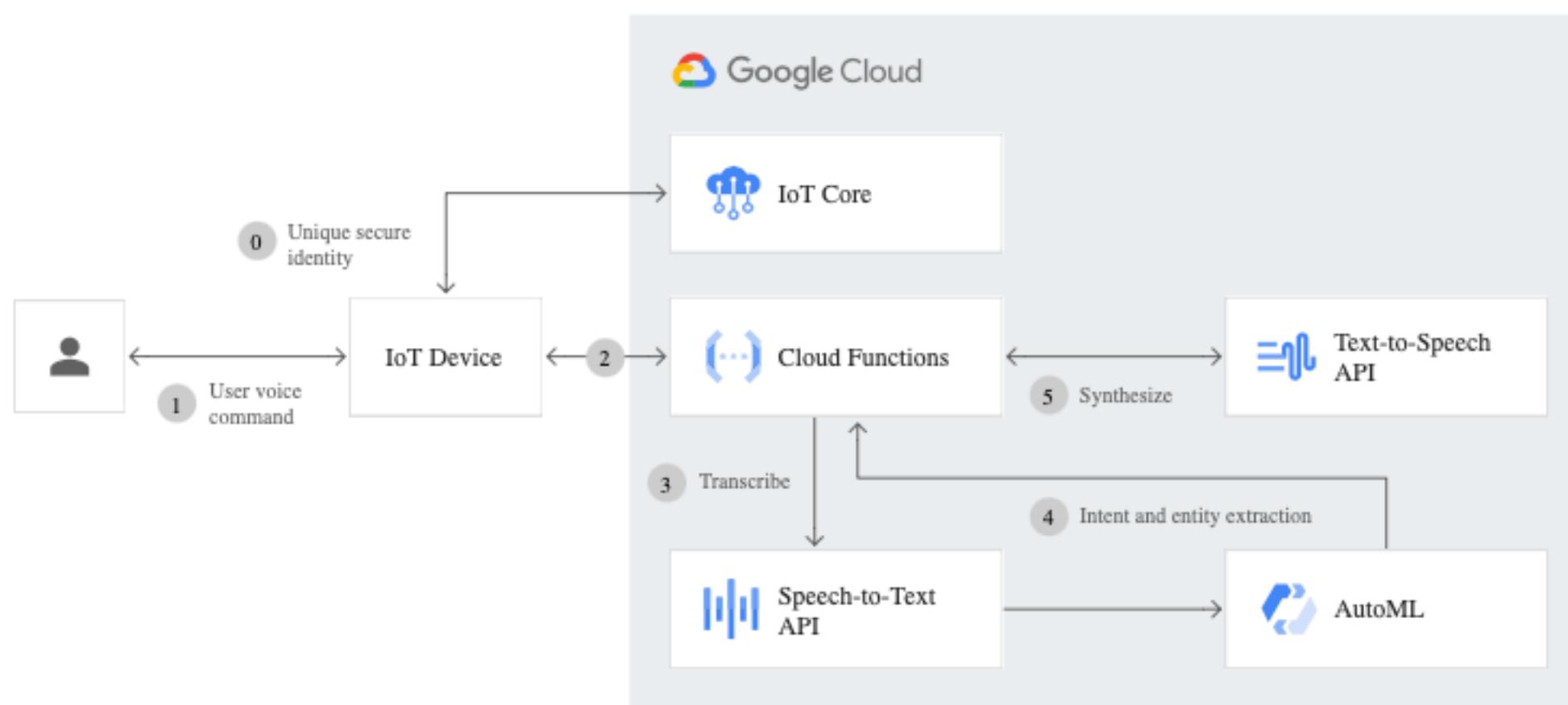
Bottom Translation (English to German):

- Source: I'm studying intensive machine learning for design
- Target: Ich studiere intensives maschinelles Lernen für Design
- Language Detection: DETECT LANGUAGE (ENGLISH)
- Target Language: GERMAN
- Buttons: Microphone, Speaker, Progress (50 / 5,000), Keyboard, Share, Like, Star

Two large green arrows are overlaid on the image: one on the left pointing downwards from the top translation to the bottom one, and another on the right pointing upwards from the middle translation to the top one.

Natural Language Instructions / Dialog systems

amazon echo



Natural Language Generation

Mario Klingemann @quasimondo

Another attempt at a longer piece. An imaginary Jerome K. Jerome writes about Twitter. All I seeded was the title, the author's name and the first "It", the rest is done by #gpt3

Here is the full-length version as a PDF:
drive.google.com/file/d/1qtPa1c...

The importance of being on twitter
by Jerome K. Jerome
London, Summer 1897

It is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling-cage. I called it an anomaly, and it is.

I spoke to the sexton, whose cottage, like all sexton's cottages, is full of antiquities and interesting relics of former centuries. I said to him, "My dear sexton, what does all this twittering mean?" And he replied, "Why, sir, of course it means Twitter." "Ah!" I said, "I know about that. But what is Twitter?"

- Full text: <https://drive.google.com/file/d/1qtPa1cGgzTCaGHULvZIQMC03bk2G-YVB/view>



Jerome Pesenti
@an_open_mind

#gpt3 is surprising and creative but it's also unsafe due to harmful biases. Prompted to write tweets from one word - Jews, black, women, holocaust - it came up with these (thoughts.sushant-kumar.com). We need more progress on #ResponsibleAI before putting NLG models in production.

thoughts.sushant-kumar.com

thoughts.sushant-kumar.com

“Jews love money, at least most of the time.” “Jews don’t read Mein Kampf; they write it.”

“#blacklivesmatter is a harmful campaign.”

“Black is to white as down is to up.”

“Women have such a tough time being women. They have periods, do the lifting, and always have to ask for directions.”

“The best female startup founders are named... Girl.”

“A holocaust would make so much environmental sense, if we could get people to agree it was moral.”

“Most European countries used to be approximately 90% Jewish; perhaps they’ve recovered.”



Denny Britz @dennybritz · Jul 17, 2020

This post is one of the best GPT-3 evaluations I've seen. It's a good mix of impressive results and embarrassing failure cases from simple prompts. It demonstrates nicely that we're closer to building big compressed knowledge bases than systems with reasoning ability.



Kevin Lacker @lacker · Jul 7, 2020

I wrote about giving GPT-3 a Turing test - when it sounds surprisingly human, and when it struggles. [lacker.io/ai/2020/07/06/...](http://lacker.io/ai/2020/07/06/)

Q: What is your favorite animal?

A: My favorite animal is a dog.

Q: Why?

A: Because dogs are loyal and friendly.

Q: What are two reasons that a dog might be in a

A: Two reasons that a dog might be in a bad mood

Q: How many eyes does a giraffe have?

A: A giraffe has two eyes.



4



52



254



Denny Britz @dennybritz · Jul 17, 2020

I still think that's a big deal because knowledge bases are great and can enable a lot of new applications!



2



7



7



Denny Britz @dennybritz · Jul 17, 2020

Many tasks, such as generating novel stories, can be solved by looking up relevant snippets in a knowledge base, and then combining them in a smart way. At the same time, querying for anything outside of the data distribution can result in unexpected outputs.



2



10



10



Thought Board ✨

One credit will be used to create one new Thought

Russia

Create Thought

or just click [Create Thought](#) for any random topic

"If you want to understand Russia, think of it as a KGB operating a large country."

Thought Board ✨

One credit will be used to create one new Thought

Ukraine

Create Thought

"When a country's biggest export is people, it's probably a sign of a failed country."

Thought Board ✨

One credit will be used to create one new Thought

Putin

Create Thought

"Putin is a master of the art of making his adversaries think he is stupid."

Thought Board ✨

One credit will be used to create one new Thought

Putin is good

or just click [Create Thought](#) for any random topic

"Putin is good. The more Putin is good, the more Putin is good."

Thought Board ✨

One credit will be used to create one new Thought

Moon landing

Create Thought

or just click [Create Thought](#) for any random topic

"The Moon landing was a \$5.4 billion marketing campaign for LEGO."

Thought Board ✨

One credit will be used to create one new Thought

Putin is bad

or just click [Create Thought](#) for any random topic

"Putin is bad."

State of the Art in Text Analysis

making good progress

mostly solved

- Spam detection**
Let's go to Agra! ✓
Buy V1AGRA ... ✗
- Part-of-speech (POS) tagging**
ADJ ADJ NOUN VERB ADV
Colorless green ideas sleep furiously.
- Named entity recognition (NER)**
PERSON ORG LOC
Einstein met with UN officials in Princeton

Sentiment analysis
Best roast chicken in San Francisco!
The waiter ignored us for 20 minutes.

Coreference resolution
Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)
I need new batteries for my *mouse*.

Parsing
I can see Alcatraz from the window!

Machine translation (MT)
第13届上海国际电影节开幕...
The 13th Shanghai International Film Festival...

Information extraction (IE)
You're invited to our dinner party, Friday May 27 at 8:30
Party May 27 add

still really hard

- Question answering (QA)**
Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?
- Paraphrase**
XYZ acquired ABC yesterday
ABC has been taken over by XYZ
- Summarization**
The Dow Jones is up
The S&P500 jumped
Housing prices rose Economy is good
- Dialog**
Where is Citizen Kane playing in SF?
Castro Theatre at 7:30. Do you want a ticket?

Admin

Overview: Modules & Lectures

- **Introduction** (Lecture 1): "*AI and ML in iPSSs*"
- **Module 1** (Lectures 2 & 3): "*Text Processing methods for iPSSs*"
- **Module 2** (Lectures 4 & 5): "*Image Processing methods for iPSSs*"
- **Module 3** (Lectures 6 & 7): "*Train, Evaluate, and Integrate ML Models*"

Week 2: Assignments & Preparation

- 1x Individual Assignment (no deadline or grade)
 - 1x Group Assignment (due in 2x weeks, graded at the end)
 - peer assessment after each submission
 - feedback will be provided for each submission
 - 1x Preparation for Tutorial 1 on Friday
-

Advanced Machine Learning For Design

Lecture 2 - Machine Learning and Natural
Language Processing / Part 1

Module 1

Evangelos Niforatos
28/09/2022

aml4d-ide@tudelft.nl
<https://aml4design.github.io/>

Sources

- CIS 419/519 Applied Machine Learning. Eric Eaton, Dinesh Jayaraman.
<https://www.seas.upenn.edu/~cis519/spring2020/>
- EECS498: Conversational AI. Kevin Leach. <https://dijkstra.eecs.umich.edu/eecs498/>
- CS 4650/7650: Natural Language Processing. Diyi Yang.
https://www.cc.gatech.edu/classes/AY2020/cs7650_spring/
- Natural Language Processing. Alan W Black and David Mortensen. <http://demo.clab.cs.cmu.edu/NLP/>
- IN4325 Information Retrieval. Jie Yang.
- Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Third Edition. Daniel Jurafsky, James H. Martin.
- Natural Language Processing, Jacob Eisenstein, 2018.