

Machine Learning for Design

Lecture 5 - Part *b*
Natural Language Processing

Previously on **ML4D**

Natural Language Processing

- High-level understanding of the language spoken and written by humans
- Also, generation (e.g., ChatGPT)
- An enabler for technology like Siri or Alexa

**why natural
language
processing?**

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

Fora, social media

 **Design Critiques: Help new and amateur designers improve their designs**

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:)



34 Comments Share Save ...

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

79.8k Members 6 Online

Created Apr 19, 2010

Filter by flair
Website Rebranding of an NGO Photo Manipulation

Moderators
Moderator list hidden. [Learn More](#)

Help About
Reddit Coins Careers
Reddit Premium Press

 **IDE TU Delft** 2,782 Tweets Following

Pinned Tweet
IDE TU Delft @idetudelft · Feb 10
"How can we design for societal transitions?" → the central question of @idetudelft'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/3uz9zxy #wearehiring #delftdesign



GIF

1 4

IDE TU Delft @idetudelft · 8h
Good afternoon. @tudelft alumni have developed a #dating app that does not work based on swiping, but with a smart algorithm. www-emerce-nl.translate.goog/interviews/sta...

1 1

IDE TU Delft @idetudelft · 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...

Product review



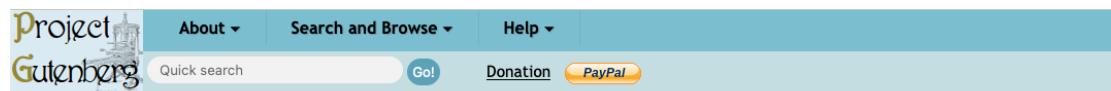
VIA AMAZON.COM

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](#).

Books

Digital, or digitised



Frequently Viewed or Downloaded

These listings are based on the number of times each eBook gets downloaded. Multiple downloads from the same Internet address on the same day count as one download, and addresses that download more than 100 eBooks in a day are considered robots and are not counted.

Downloaded Books

2022-02-27	156396
last 7 days	1167285
last 30 days	4234525

- [Top 100 EBooks yesterday](#)
- [Top 100 Authors yesterday](#)
- [Top 100 EBooks last 7 days](#)
- [Top 100 Authors last 7 days](#)
- [Top 100 EBooks last 30 days](#)
- [Top 100 Authors last 30 days](#)

Top 100 EBooks yesterday

1. [Pride and Prejudice by Jane Austen \(1760\)](#)
2. [Frankenstein; Or, The Modern Prometheus by Mary Wollstonecraft Shelley \(1742\)](#)
3. [Simple Sabotage Field Manual by United States. Office of Strategic Services \(1147\)](#)
4. [Alice's Adventures in Wonderland by Lewis Carroll \(988\)](#)
5. [The Adventures of Sherlock Holmes by Arthur Conan Doyle \(740\)](#)
6. [The Yellow Wallpaper by Charlotte Perkins Gilman \(699\)](#)
7. [The Great Gatsby by F. Scott Fitzgerald \(619\)](#)
8. [The Picture of Dorian Gray by Oscar Wilde \(609\)](#)
9. [A Tale of Two Cities by Charles Dickens \(606\)](#)
10. [The House of the Arrow by A. E. W. Mason \(592\)](#)
11. [Moby Dick; Or, The Whale by Herman Melville \(582\)](#)
12. [Dracula by Bram Stoker \(581\)](#)

Interviews

Interviewee: XXX
Interviewer: XXX
Date of Interview: mm.dd.yy
Location of Interview: XXX
List of Acronyms: FP=Frank Peterson, IN=Interviewer

[Begin Transcript 00:00:10]

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

FP: Well when I joined the navy, actually that was in 1950 at the age of 18. Not much other than the fact that I wanted to get away from Topeka and see what the rest of world was really all about.

IN: Um-hm.

[00:00:26]

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

And 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.

But 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 jumped from the frying pan into the Korean War via El Toro.

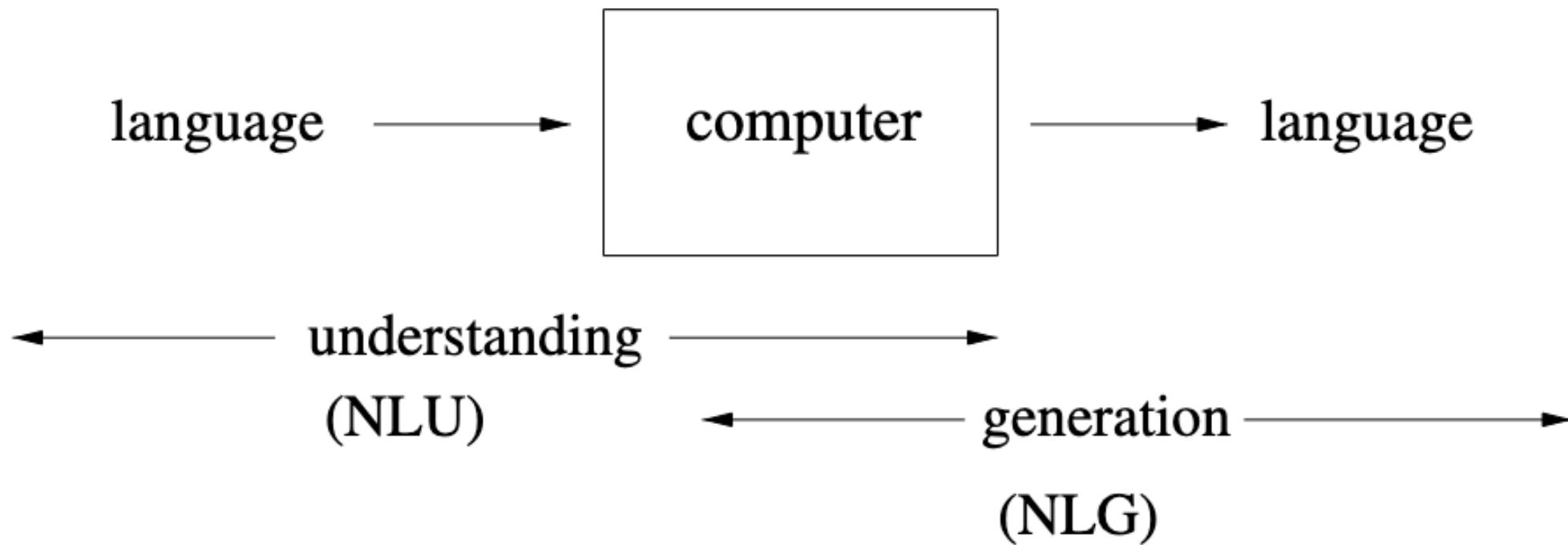
IN: I see.

[End Transcript 00:01:21]

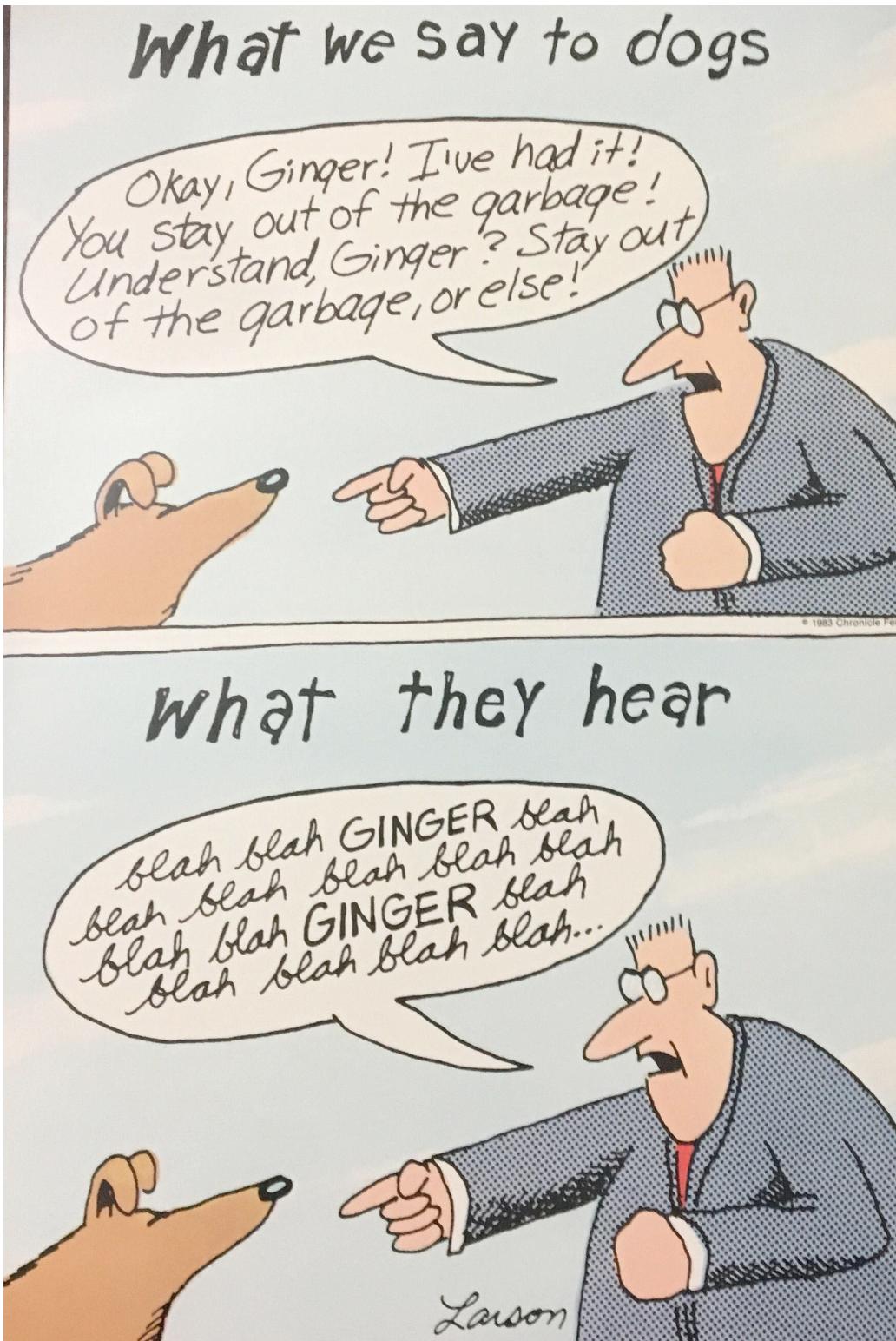
- Answer questions using the Web
- Translate documents from one language to another
- Do library research; summarize
- Archive and allow access to cultural heritage
- Interact with intelligent devices
- Manage messages intelligently
- Help make informed decisions
- Follow directions given by any user
- 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

What is Natural Language Processing?

Computer using natural language as input and/or output



- **N**atural: human communication, unlike e.g., programming languages
- **L**anguage: signs, meanings, and a code connecting signs with their meanings
- **P**rocessing: computational methods to allow computers to ‘understand’, or to generate



Beyond keyword matching

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

Why is NLP 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 meaning in different ways
- Some languages express some meanings more readily/often

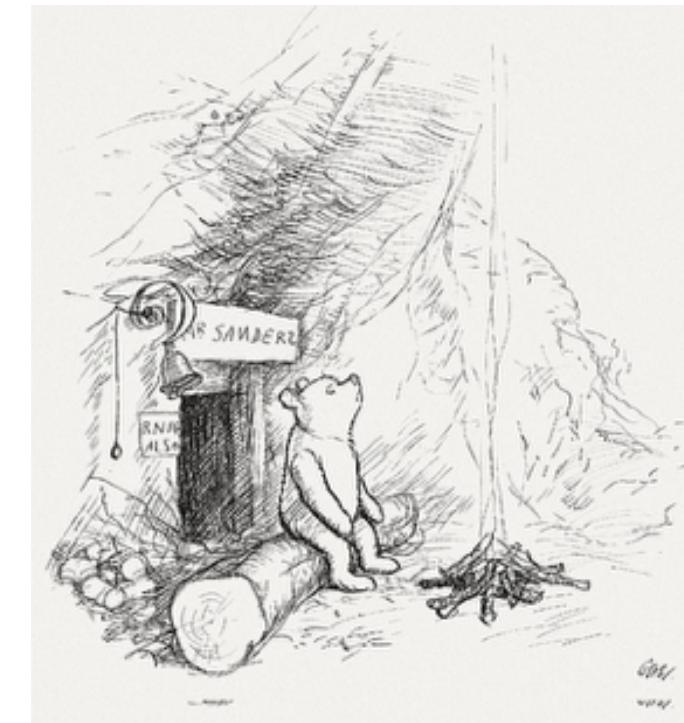
Knowledge Bottleneck

- About language
- About the world: Common sense and Reasoning

Ambiguity and Expressivity

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

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



Lexical ambiguity

The presence of two or more possible meanings within a single word

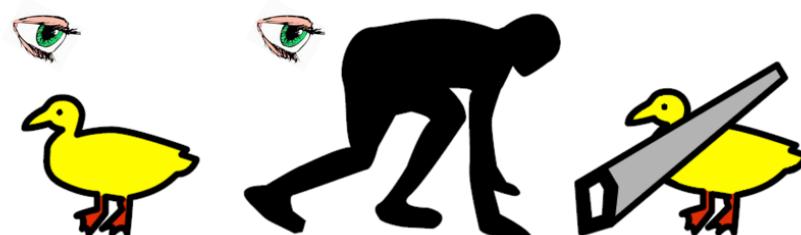


credit: A. Zwicky

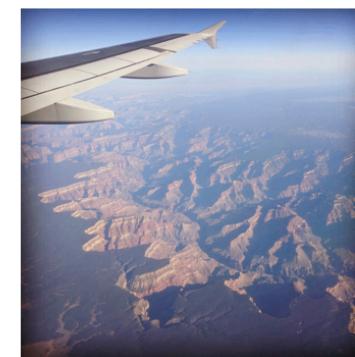
Syntactic ambiguity (Word sense ambiguity)

The presence of two or more possible meanings within a *single sentence or sequence of words*

I saw her duck



I saw the Grand Canyon flying to New York



Clearly the gran
canyon does not
fly....

Attachment ambiguity

*The policeman shot
the thief with the 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!

Semantic Ambiguity



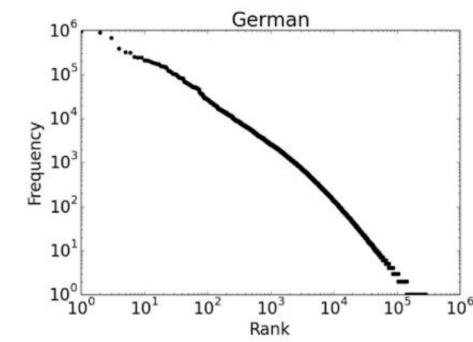
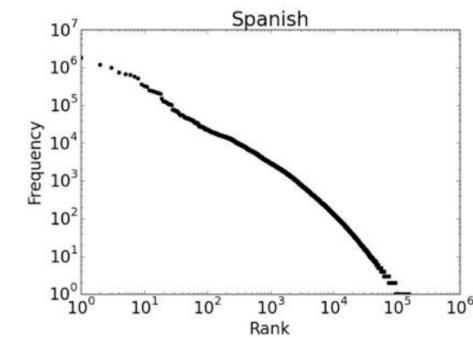
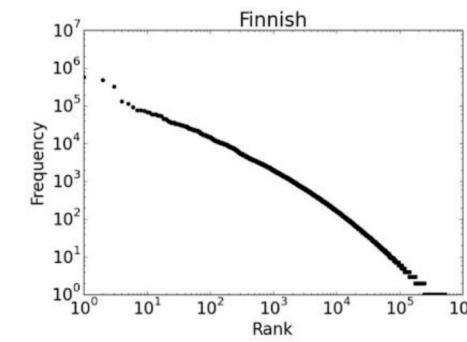
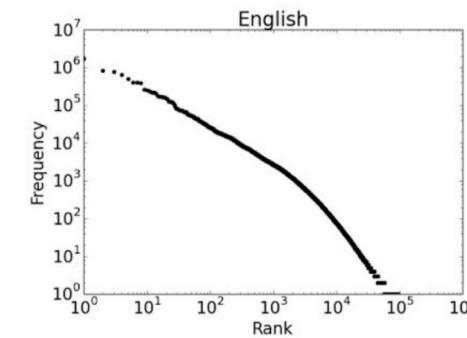
Every fifteen minutes a woman in this country gives birth. Our job is to find this woman, and stop her!

Groucho Marx

Sparsity

Zip's Law

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



$$f(r) = c/r^z$$

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

Language Evolution

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

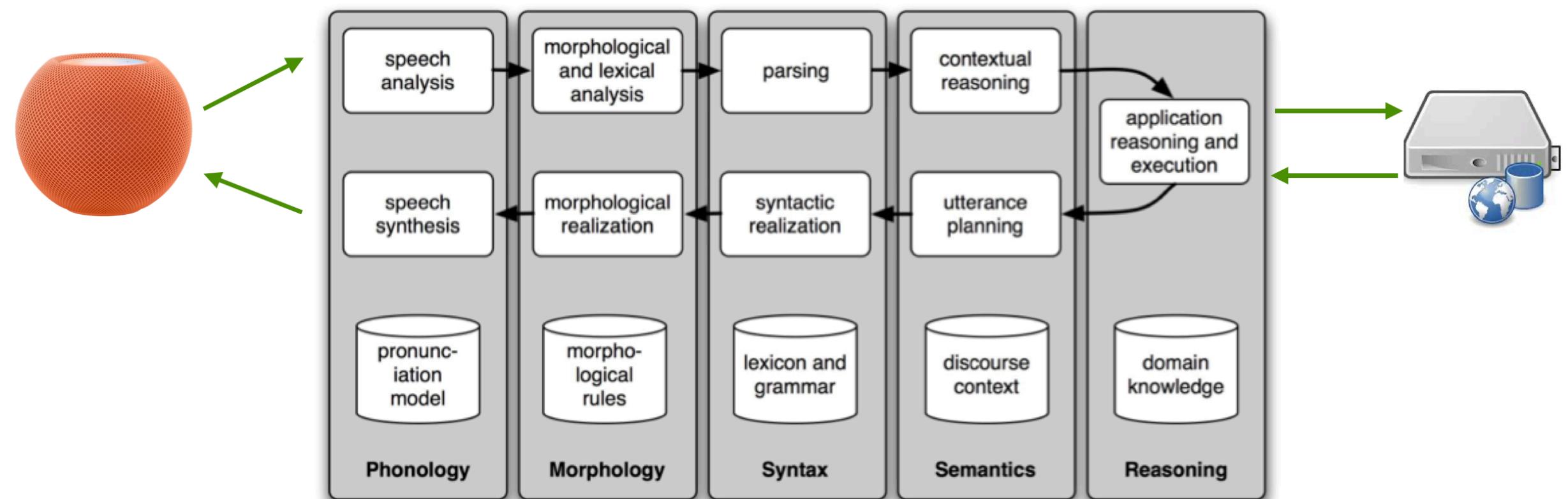
LUWAMH

Love you with all my heart



NLP Tasks

Example NLP Process



Morphology

Tokenisation

„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 .“

朝鲜外务省发言人11月1日在平壤宣布，朝鲜将重返六方会谈，但前提条件是朝鲜与美国在六方会谈框架内讨论解除美国对朝鲜的制裁问题。
针对朝鲜方面的表示欢迎。
Where are the words?
美联社11月1日报道说：“长期以来一直拒绝与平壤进行直接对话的美国总统布什认为，各方达成一致、同意恢复六方会谈应归功于中国的斡旋。

- 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”

Stop-word Removal

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

- 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..)

Lemmatisation

It uses dictionaries and morphological analysis of words to return the base or dictionary form of a word

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 ,
headquarter
Sundar Pichai said in his keynote that users love their new Android phones .
say user phone

Stemming

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 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

Porter stemmer: 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

Paice stemmer: 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

- 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

Syntax

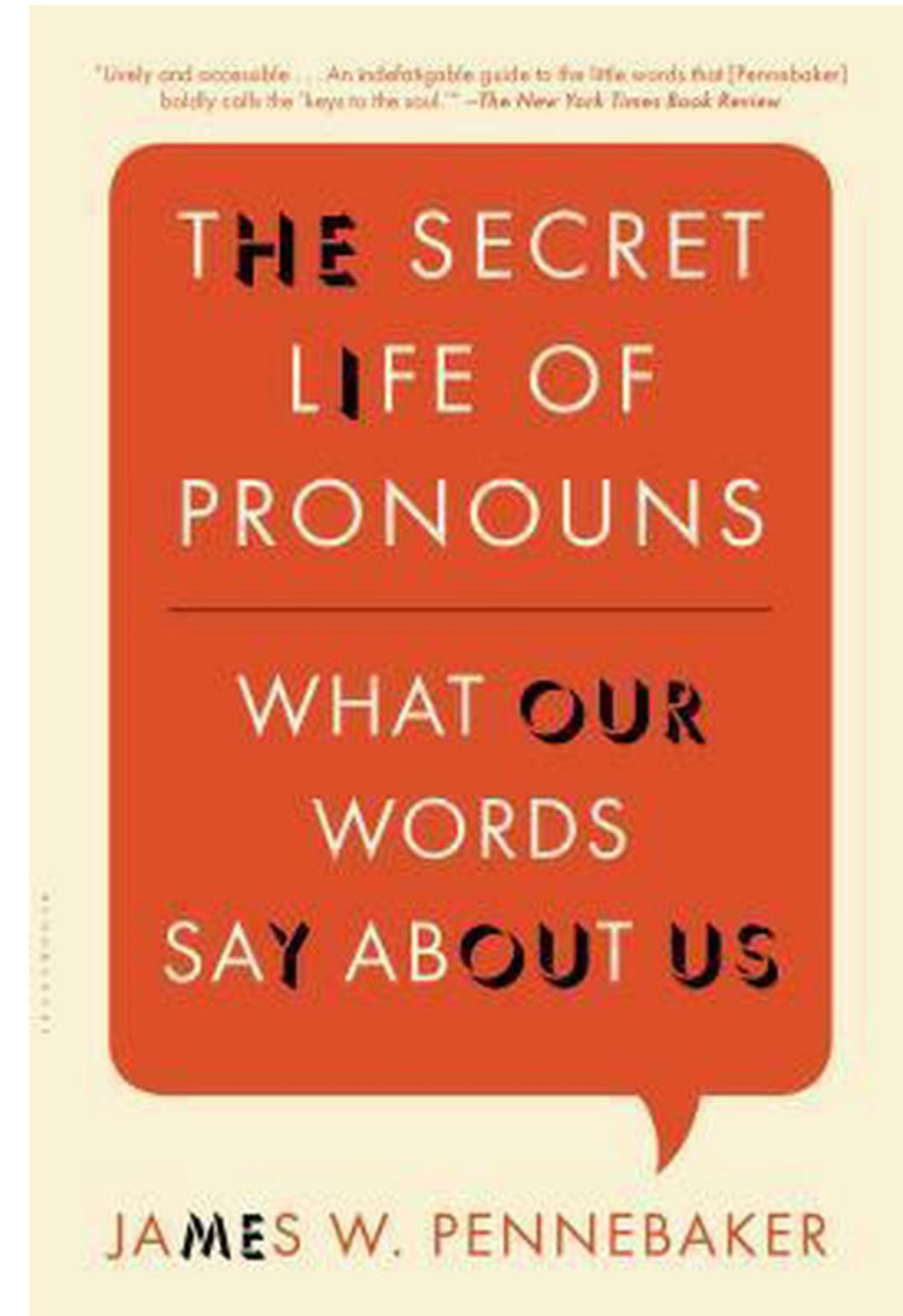
Part-of-speech Tagging

Tagging each word in a sentence with a corresponding *part-of-speech* (e.g. noun, verb, adverbs)

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

Language Analysis

- Idea: people's language can provide *insights into their psychological states* (e.g. emotions, thinking style)
- For instance
 - Frequency of words associated with positive or negative emotions
 - Use of pronouns as a proxy for confidence and character traits



INTRODUCING LIWC-22

A NEW SET OF TEXT ANALYSIS TOOLS AT YOUR FINGERTIPS

People reveal themselves by the words they use. Using LIWC-22 to analyze others' language can help you understand their thoughts, feelings, personality, and the ways they connect with others. It can give you insights you've never had before into the people and world around you.

The Development and Psychometric Properties of LIWC-22

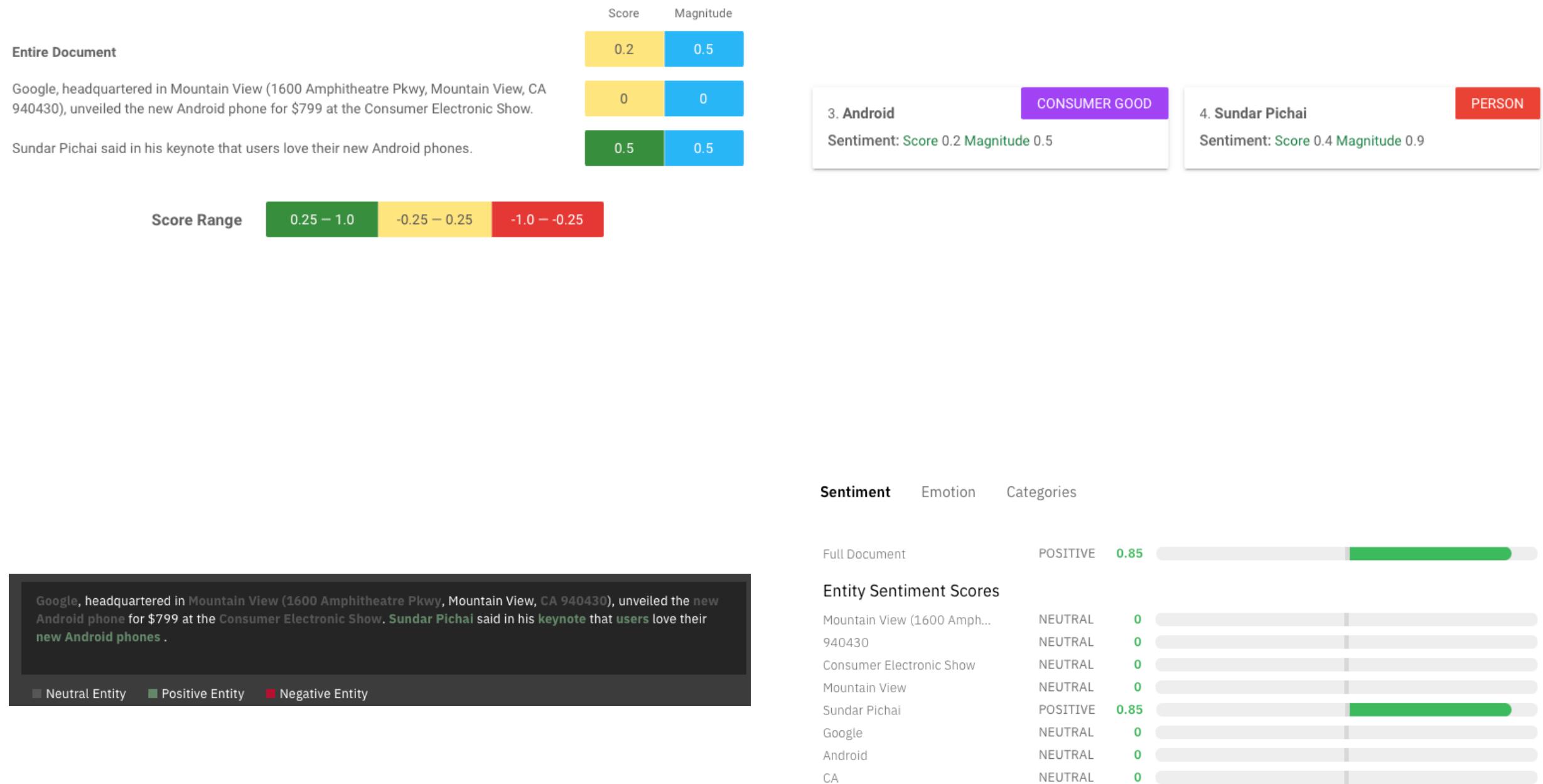
Category	Abbrev.	Description/Most frequently used exemplars
Summary Variables		
Word count	WC	Total word count
Analytical thinking	Analytic	Metric of logical, formal thinking
Clout	Clout	Language of leadership, status
Authentic	Authentic	Perceived honesty, genuineness
Emotional tone	Tone	Degree or positive (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

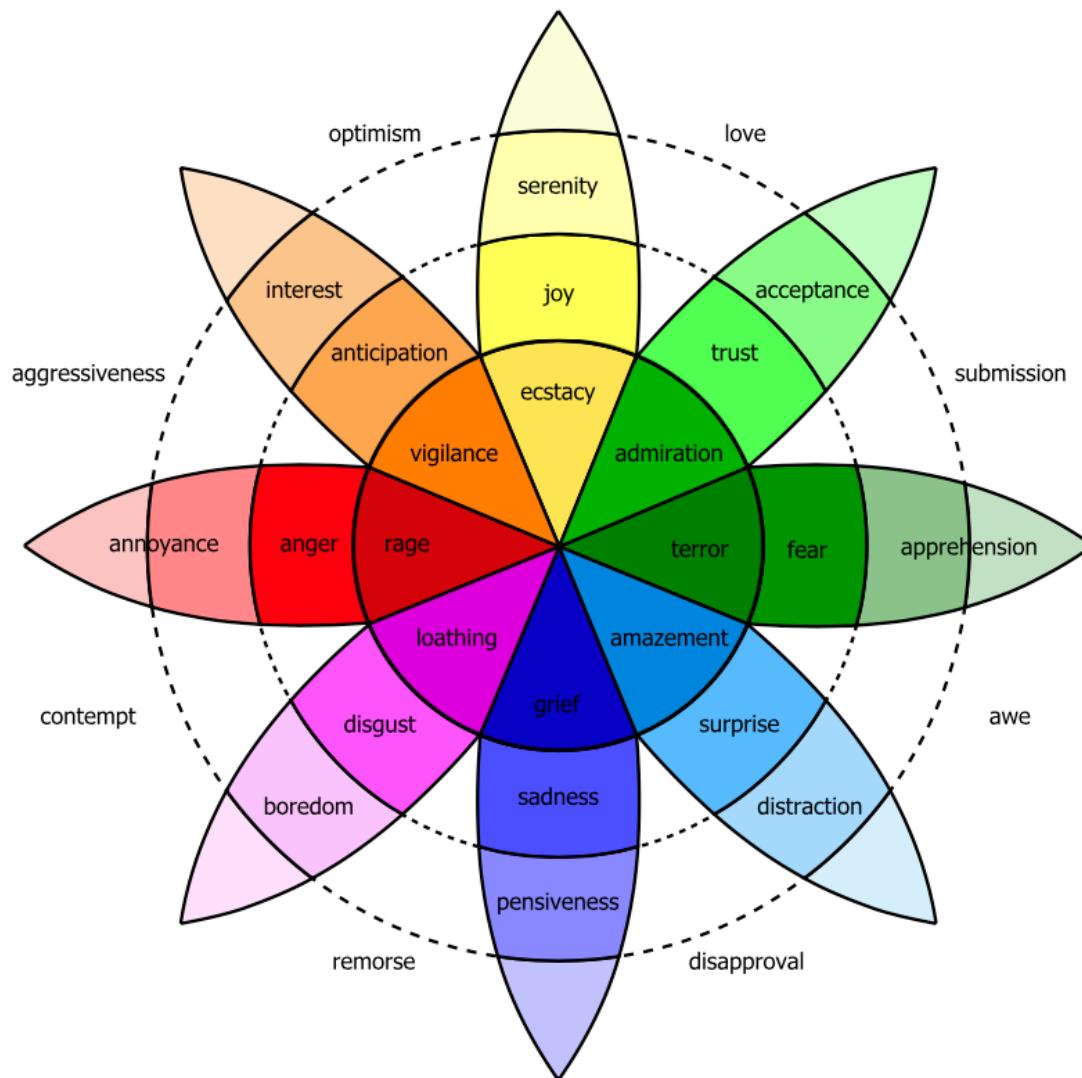
- **Analytic Thinking:** the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns.
 - low Analytical Thinking → language that is more intuitive and personal
- **Clout:** the relative social status, confidence, or leadership that people display through their writing or talking
- **Authenticity:** the degree to which a person is self-monitoring
 - Low authenticity: prepared texts (i.e., speeches 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.

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 text's attitude positive or negative?
 - *Regression*: Rank the attitude of the text from 1 to 5
 - *Advanced*: Detect the target, source, or complex attitude types



Emotion Analysis

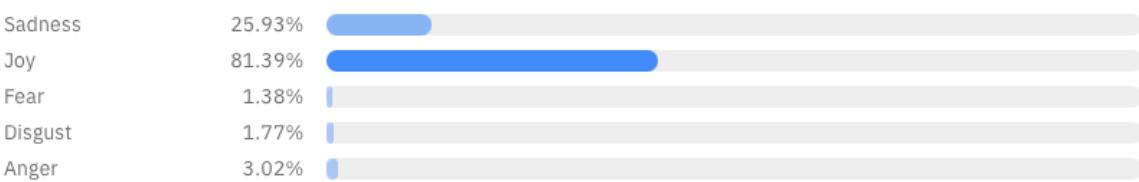


Plutchik wheel of emotion

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 .

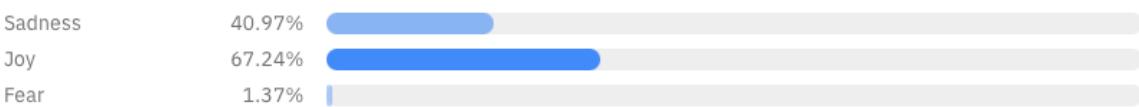
■ Sadness ■ Fear ■ Disgust ■ Anger ■ Joy

Full Document



Entity Emotion Scores

Mountain View (1600 Amphitheatre Pkwy)



Semantics

Named Entity Recognition

- Factual information and knowledge are usually expressed by **named entities**
- Who, Whom, Where, When, Which, ...
- **Identify** words that refer to proper names of interest in a particular application
 - E.g. people, companies, locations, dates, product names, prices, etc.
- **Classify** them to the corresponding classes (e.g. person, location)
- **Assign** a unique identifier from a database

⟨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
Wikipedia Article	
Salience: 0.10	
6. users	PERSON
Wikipedia Article	
Salience: 0.09	
7. Amphitheatre Pkwy	LOCATION
Wikipedia Article	
Salience: 0.07	
8. CA 940430	OTHER
Wikipedia Article	
Salience: 0.05	
9. keynote	OTHER
Wikipedia Article	
Salience: 0.03	
10. phones	CONSUMER GOOD
Wikipedia Article	
Salience: 0.02	
11. Consumer Electro...	EVENT
Wikipedia Article	
Salience: 0.02	
12. 1600 Amphitheatr...	ADDRESS
Wikipedia Article	

Document 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

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

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

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

Word Sense Disambiguation

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

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)

Automated Summarisation

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

Text Summarization Result

Original URL/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.

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).

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.

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://textsummarization.net/>

<https://brevi.app/single-demo>
(not working!)

Machine Translation (popular languages)

The image displays three stacked screenshots of a machine translation application, likely Google Translate, illustrating the bidirectional nature of its popular language support.

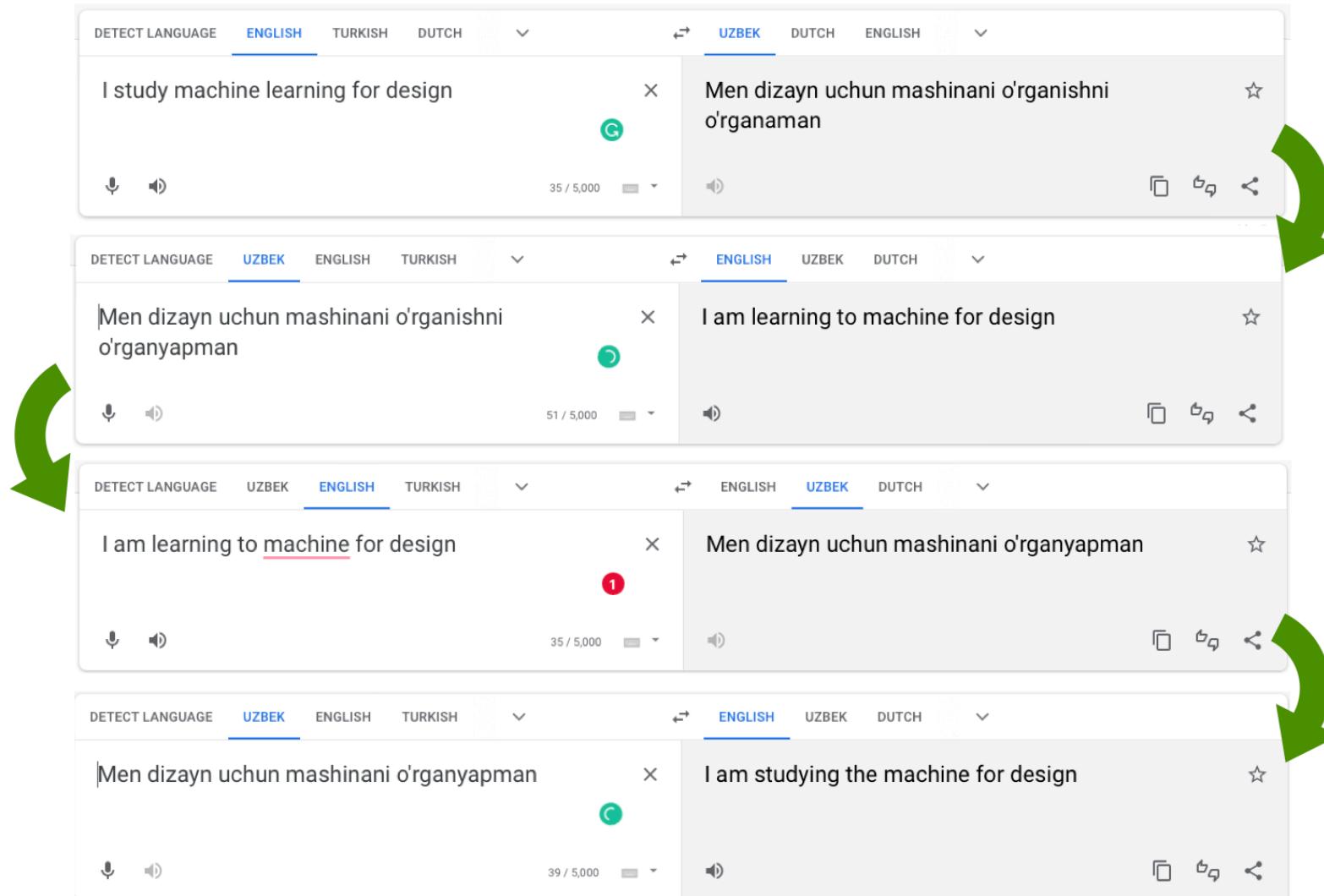
Screenshot 1 (Top): Shows the interface with "DETECT LANGUAGE" set to "GERMAN", "ENGLISH" selected as the target language, and "DUTCH" as the source language. The input text is "I study machine learning for design". The output is "Ich studiere maschinelles Lernen für Design".

Screenshot 2 (Middle): Shows the interface with "DETECT LANGUAGE" set to "GERMAN", "GERMAN" selected as the target language, and "ENGLISH" as the source language. The input text is "Ich studiere maschinelles Lernen für Design". The output is "I'm studying machine learning for design".

Screenshot 3 (Bottom): Shows the interface with "DETECT LANGUAGE" set to "GERMAN", "ENGLISH" selected as the target language, and "DUTCH" as the source language. The input text is "I'm studying machine learning for design". The output is "Ich studiere maschinelles Lernen für Design".

Green arrows on the left and right sides of the screenshots indicate the bidirectional nature of the translation process between English and German.

Machine Translation (languages with fewer resources)



Natural Language Instructions / Dialog systems



Natural Language Generation

 **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.



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

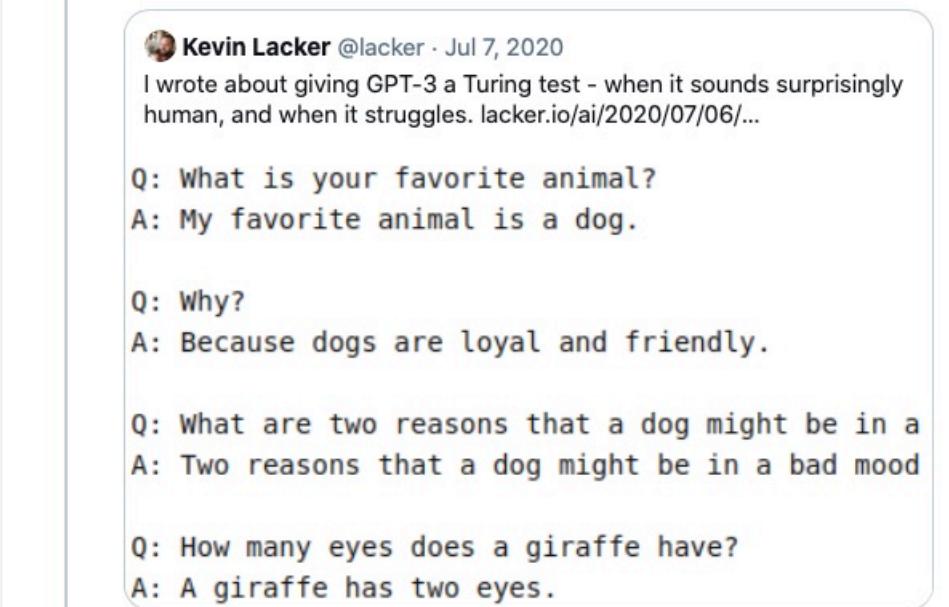
"#blacklivesmatter is a harmful campaign." "Black is to white as down is to up."

"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/...](https://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 bad mood
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 ↑

 **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 ↑

State of the Art in NLP

As of 2022

making good progress

mostly solved

Spam detection

Let's go to Agral
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?



Credits: Nava Tintarev

**What is the state of the art
in 2024?**

Try with Aelixr!

Machine Learning for Design

Lecture 5 - Part *b*
Natural Language Processing

Credits

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.

Text as Data. Grimmer, Justin; Roberts, Margaret E.; Stewart, Brandon M. Princeton University Press.