

NLP Topics

Please, record.

Questions?

Modification to the schedule

DOW	Day	Month	Lectures	Literature review	PA
Tue	26	Jan	Introduction		
Thur	28	Jan	NLP Topics		
Tue	2	Feb	Regular Expressions		<i>PA1 assign</i>
Thur	4	Feb	Lecture: Literature search	Topic Chosen	
Tue	9	Feb	Ngrams		
Thur	11	Feb		Citation list (12 papers)	
Tue	16	Feb	Ngrams + Smoothing		<i>PA2 assign; PA1 due</i>
Thur	18	Feb	Association Measures		
Tue	23	Feb	<i>Reading Day</i>		
Thur	25	Feb		Problem Statement	
Tue	2	Mar	Lecture: POS Tagging I		<i>PA3 assign; PA2 due</i>
Thur	4	Mar	Lecture: POS Tagging II		
Tue	9	Mar	Lexical Semantics		
Thur	11	Mar		Annotated Bibliography	
Tue	16	Mar	WSD		<i>PA4 assign; PA3 due</i>
Thur	18	Mar	Vector Representation		
Tue	23	Mar	Machine learning		
Thur	25	Mar		Spreadsheet	
Tue	30	Mar	Sentiment Analysis		<i>PA5 assign; PA4 due</i>
Thur	1	Apr	IE		
Tue	6	Apr	LBD		
Thur	8	Apr		1st Draft due	
Tue	13	Apr	IR		<i>PA6 assign; PA5 due</i>
Thur	15	Apr	Q&A		
Tue	20	Apr	Summarisation		
Thur	22	Apr		2nd Draft due	
Tue	27	Apr	Topic Analysis		<i>PA6 due</i>
Thur	29	Apr			
Tue	4	May	Final exam - take home	Final Draft due	



NLP Topics

Term identification

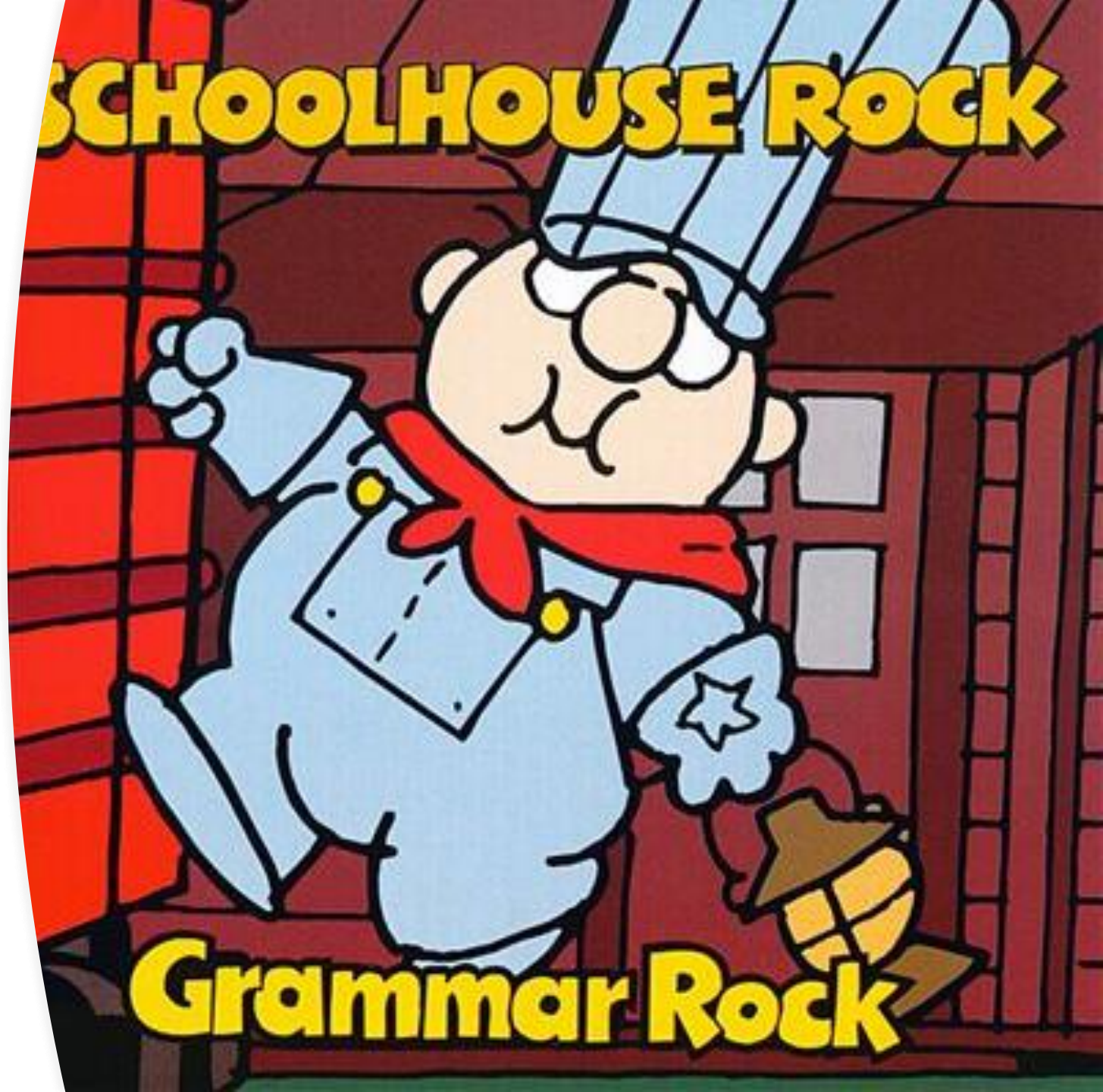


Collocation identification



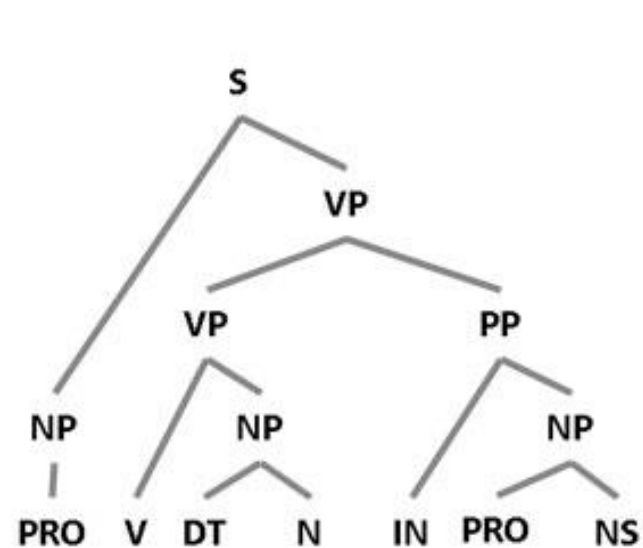
Part of speech tagging

- Noun (e.g. The *dog* barked.)
- Verb (e.g. Tim *jumped* up.)
- Pronoun (e.g. *He* sat down.)
- Preposition (e.g. He hid *under* the covers.)
- Adverb (e.g. She *slowly* stood up.)
- Conjunction (e.g. He talked *and* ate.)
- Participle (e.g. She saw John *eating*)
- Article (e.g. *The* cat ate *a* mouse)

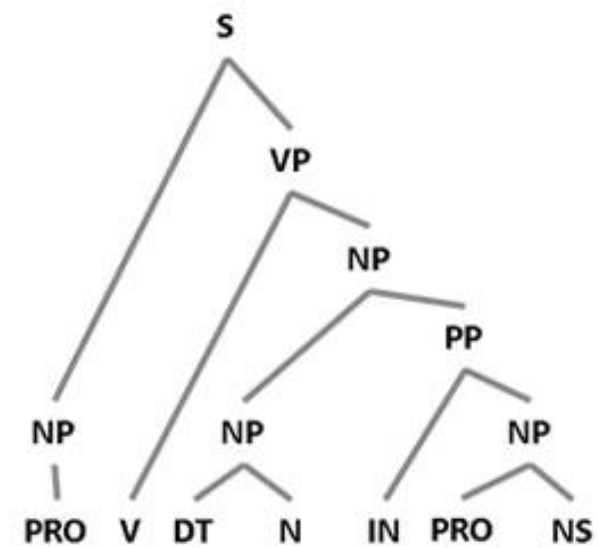


Parsing

How Parse Trees Work




I shot an elephant in my pajamas.



I shot an elephant in my pajamas.

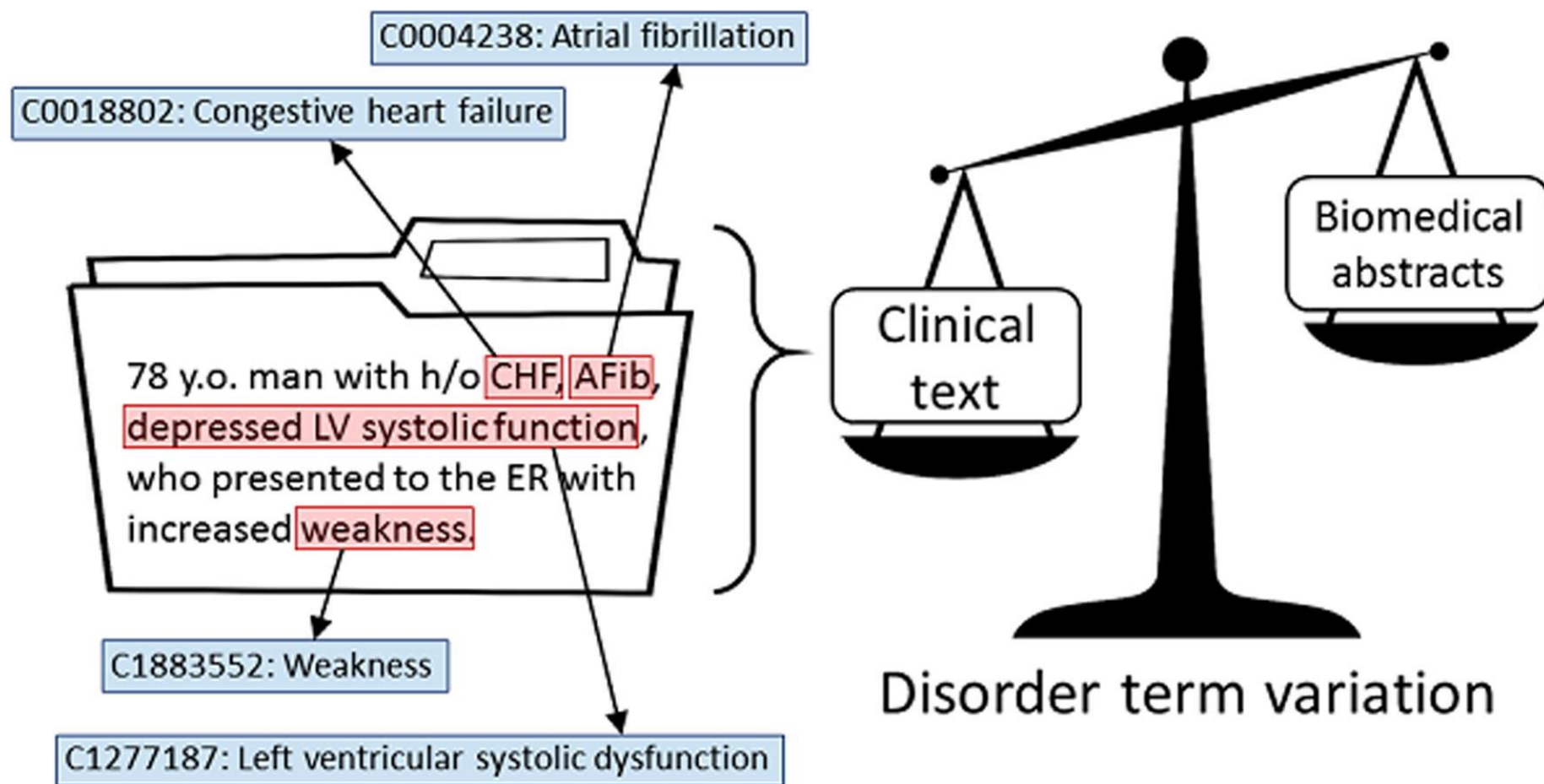
Key: N = Noun | NS = Plural Noun | NP = Noun Phrase | PRO = Pronoun | V = Verb | VP = Verb Phrase |
DT = Determiner | IN = preposition | PP = Prepositional Phrase

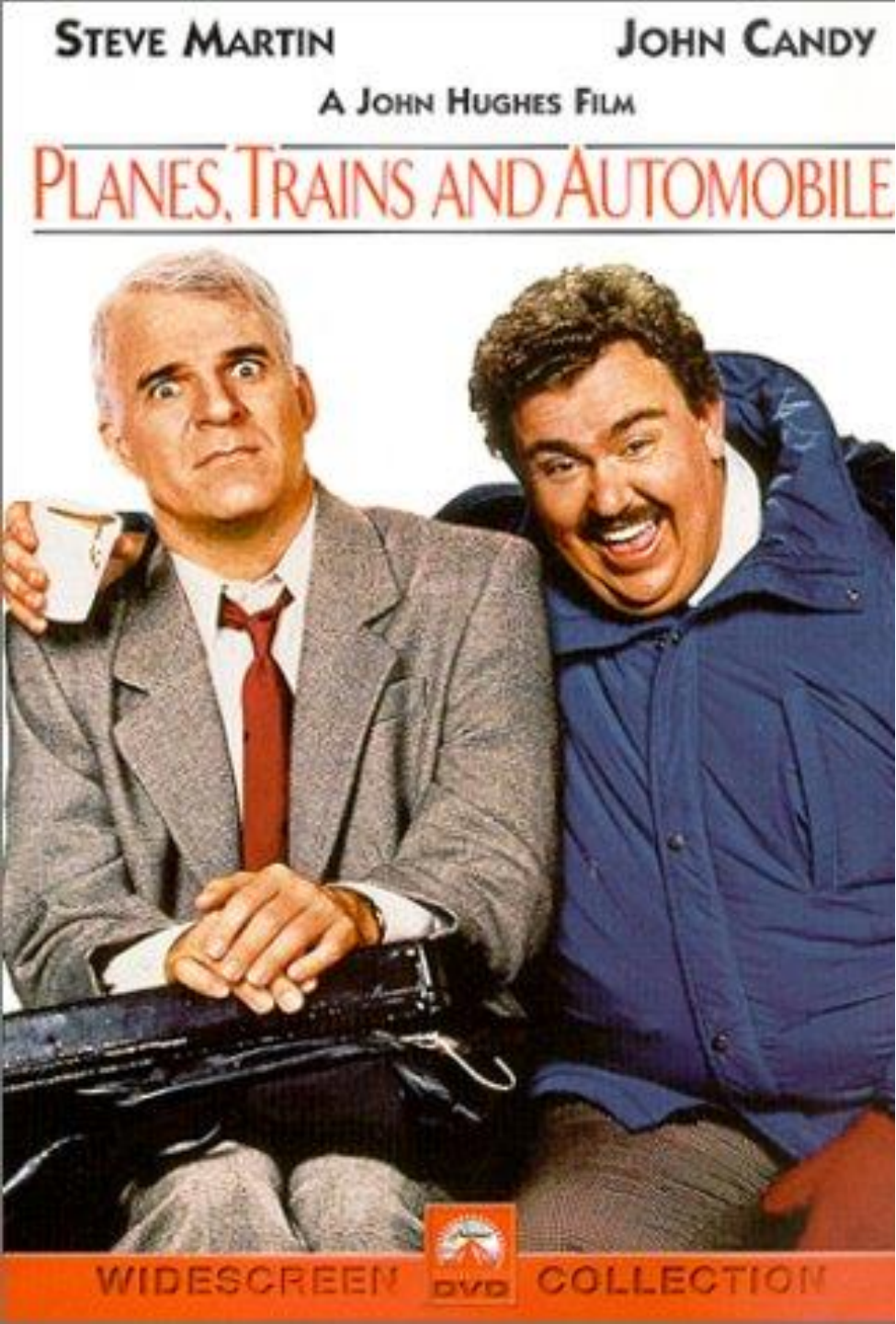


Word sense disambiguation



Normalization



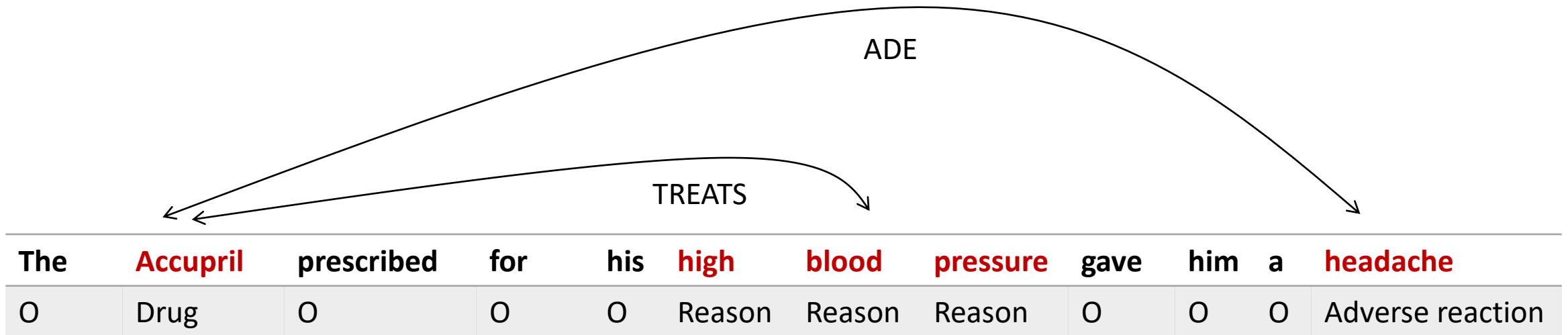
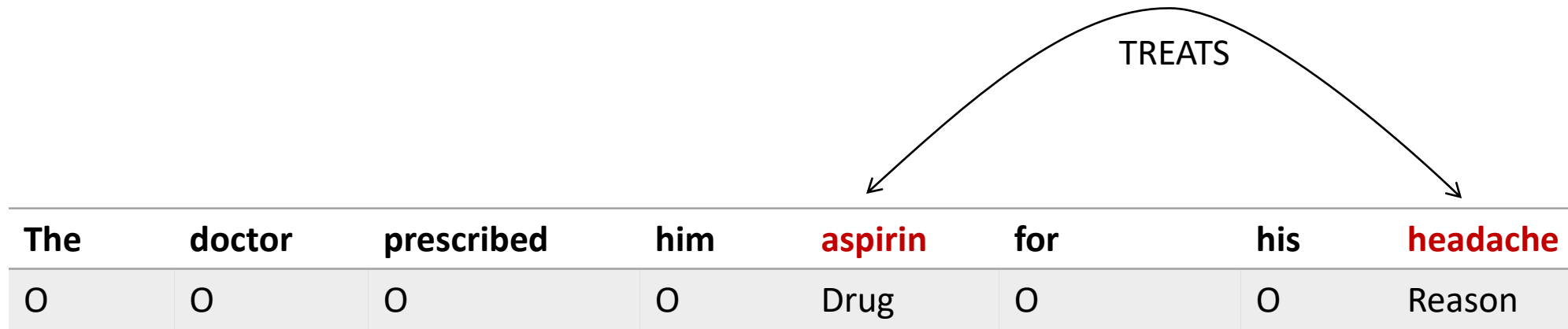


Named entity recognition

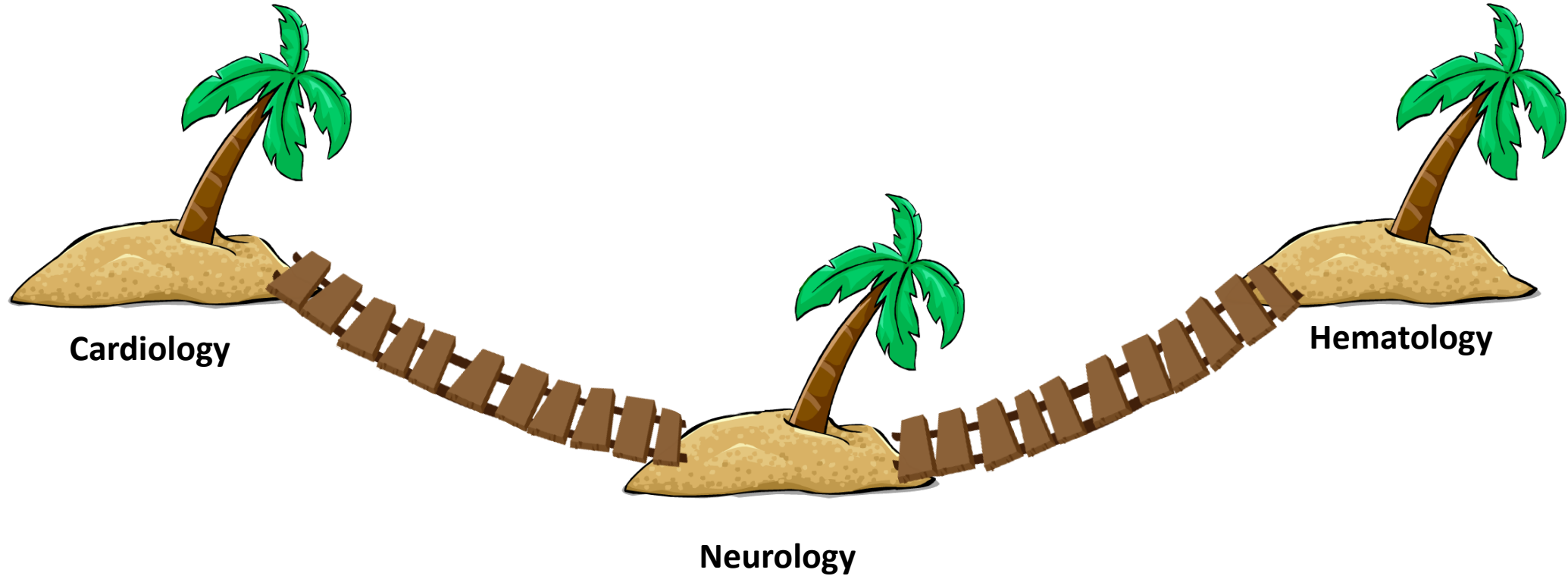
Traditionally anything that can be referred to with a proper name

- People (PER): Individuals, fiction characters
- Organization (ORG): Companies, Agencies
- Location (LOC): Physical extents, mountain ranges, seas
- Geo-political (GP): Countries, states,
- Facility (FAC): Bridges, airports, buildings
- Vehicles (VEH): Planes, trains and automobiles

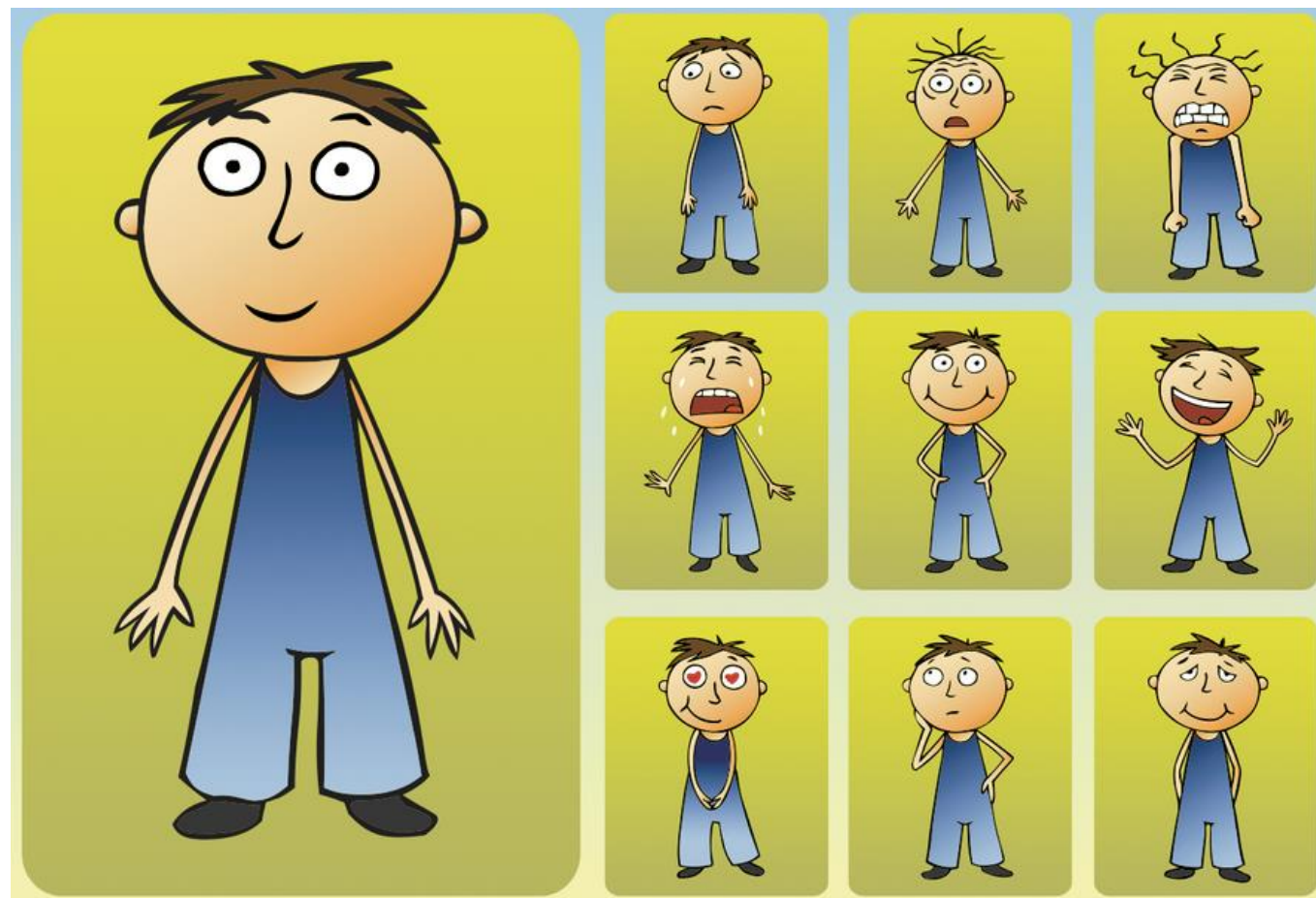
Relationship extraction



Knowledge (literature-based) discovery



Sentiment analysis



Sarcasm detection



Information retrieval



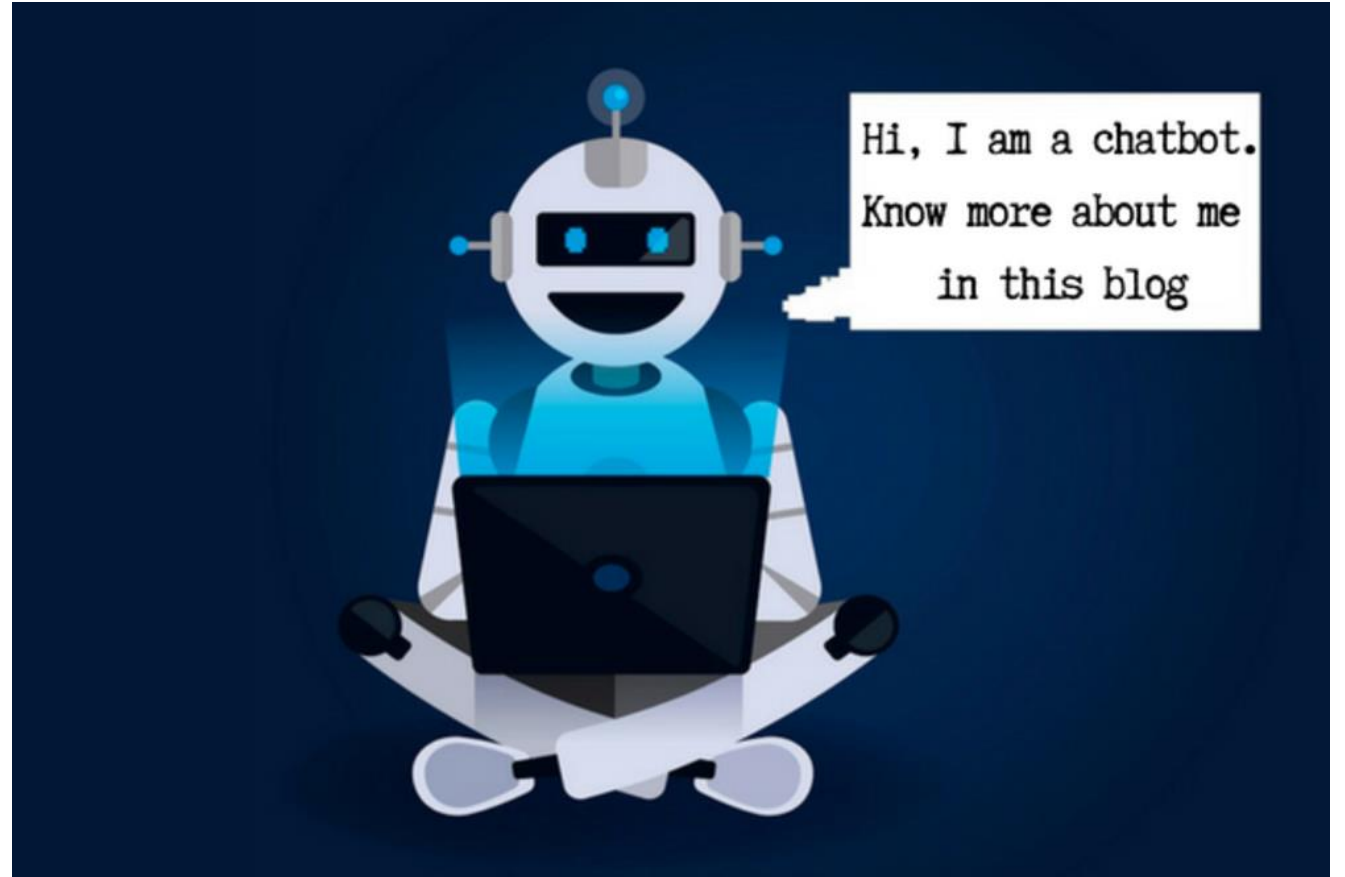
Summarisation



Question and Answering



Chatbots



The Words They Used

The words that the speakers have used during the Democratic convention suggest how the party's themes have changed since the last presidential campaign.

Speakers have hammered home Barack Obama's "change" theme, using the word about eight times as often as they

did in 2004.

Also, unlike 2004, when the Kerry campaign sought to avoid direct attacks on the president at the convention, the speakers have regularly have been mentioning John McCain by name. Speakers in 2004 practiced "the art of the

implicit slam," a veteran Democratic speechwriter said then, indirectly bashing Mr. Bush while barely using his name.

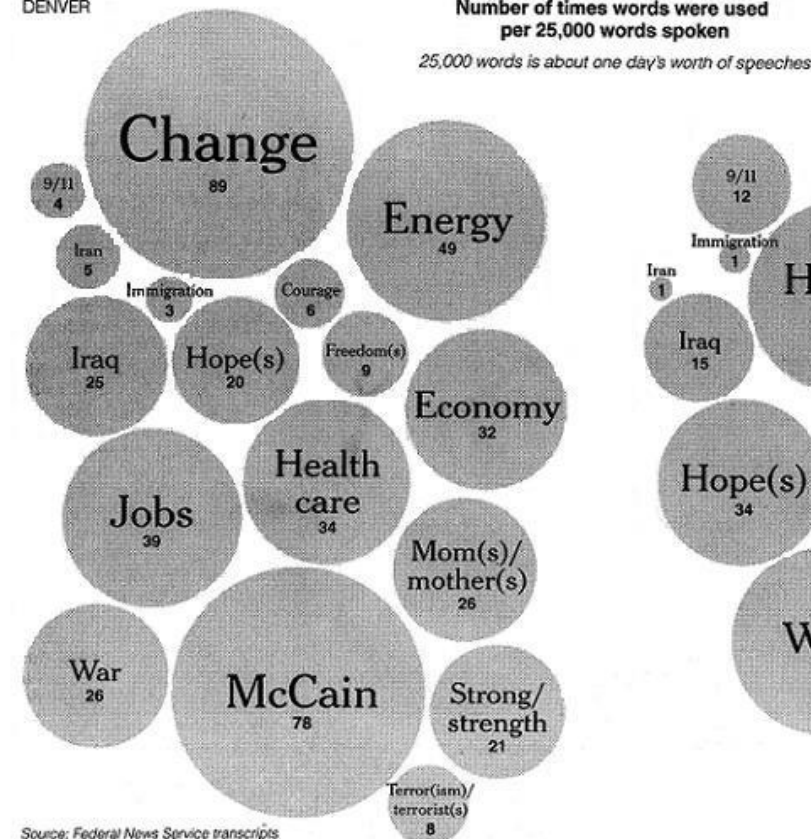
Also on the upswing: more mentions of the economy, energy, Iran and Iraq.

Words less frequently used: freedom, Sept. 11 and terrorism.

Topic analysis

2008

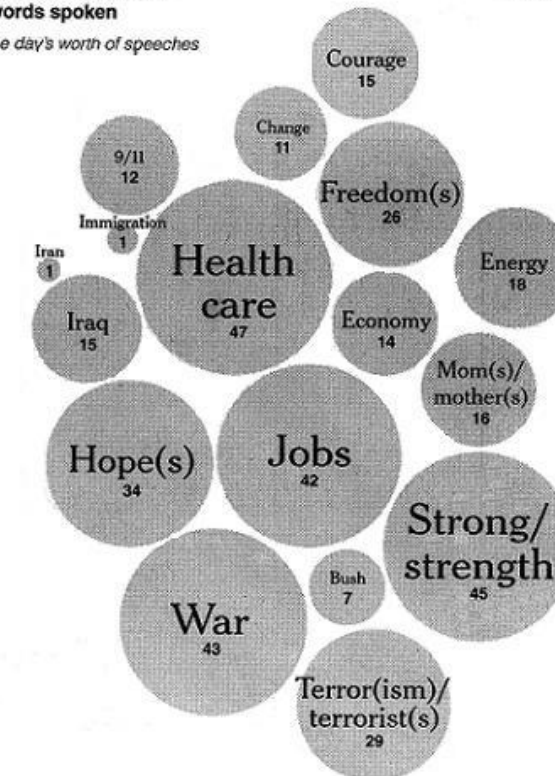
DENVER



Source: Federal News Service transcripts

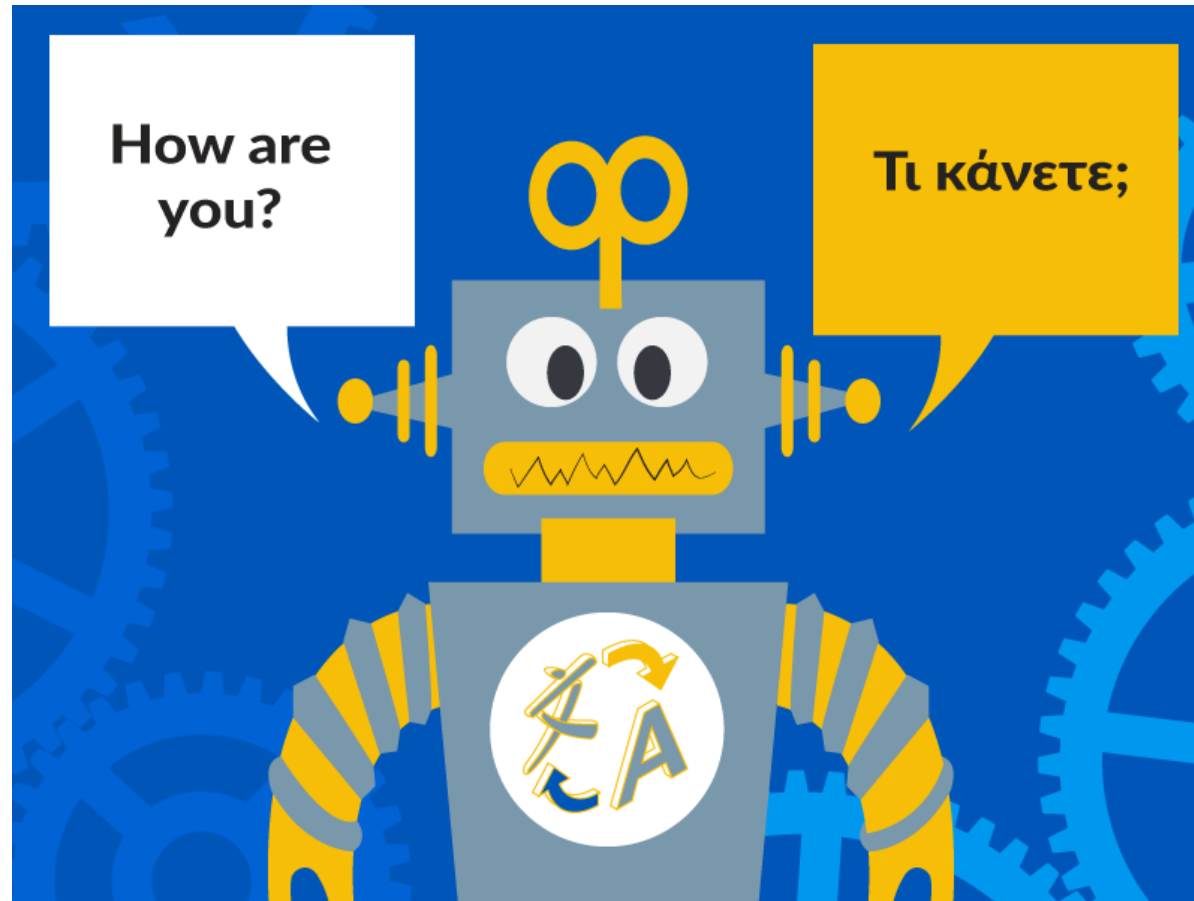
2004

BOSTON



MATTHEW ERICSON/THE NEW YORK TIMES

Machine Translation



Suggested places look at for topics

- SemEval: <https://alt.qcri.org/semeval2020/index.php?id=tasks>
- CLEF: <http://www.clef-initiative.eu/>
- BioCreative: <http://biocreative.sourceforge.net/>
- Wikipedia: https://en.wikipedia.org/wiki/Natural_language_processing

Where to look for academic papers on your topic

Google Scholar

☒ Articles ☐ Case law

Julie Arendt



(804) 828-1539

James Branch Cabell Library,
Room 121-J

arXiv.org



<https://www.library.vcu.edu/about/staff/staff/arendt-julie.html>

Questions?

Next week:

- Tuesday: 2 February 2021: Regular Expressions
- Thursday: Topic choice is due
 - You are required to get the topic approved so meet with me during my office hours or during class time to discuss
 - If those times won't work, send me an email