Introduction to Natural Language Processing

Paul Rad, Ph.D.

Associate Professor
Cyber Analytics and AI
Information Systems and Cyber Security
College of Business School
210.872.7259

Outline

Big Text Data and Processing

- Rule-base approach
- Probabilistic Machine Learning
- Deep Learning approach

What is NLP

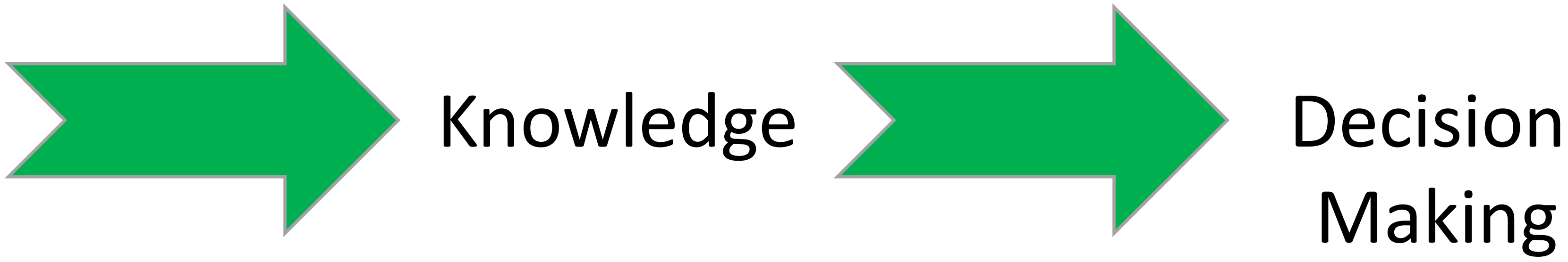
Application of NLP

What are the challenges?

Key NLP components

Big Text Data Analytics

- Internet
- Blogs
- News
- Email
- Literature
- Twitter
- Websites
- Reviews

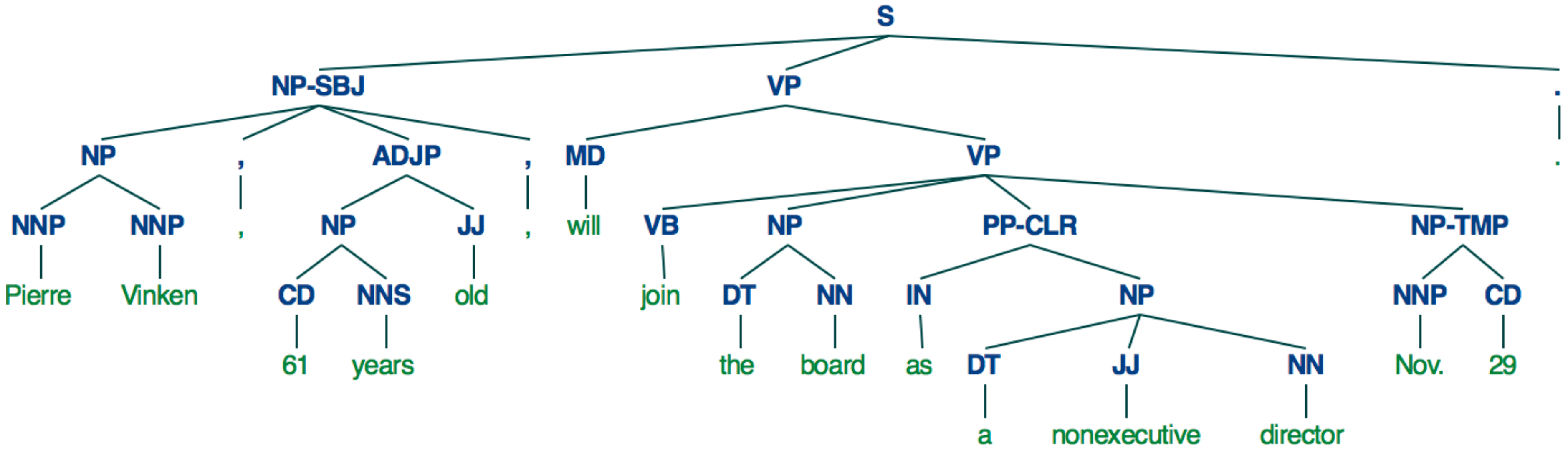


Knowledge

Decision Making

Text Analytics - Syntactic Analysis

Pierre Vinken 61 years old will join the board as a nonexecutive director Nov. 29



Syntactic Analysis

Lexical Analysis
Part of speech tagging

What is NLP

Natural language processing (**NLP**) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (**natural**) languages. – **Wikipidia**

NLP Pyramid

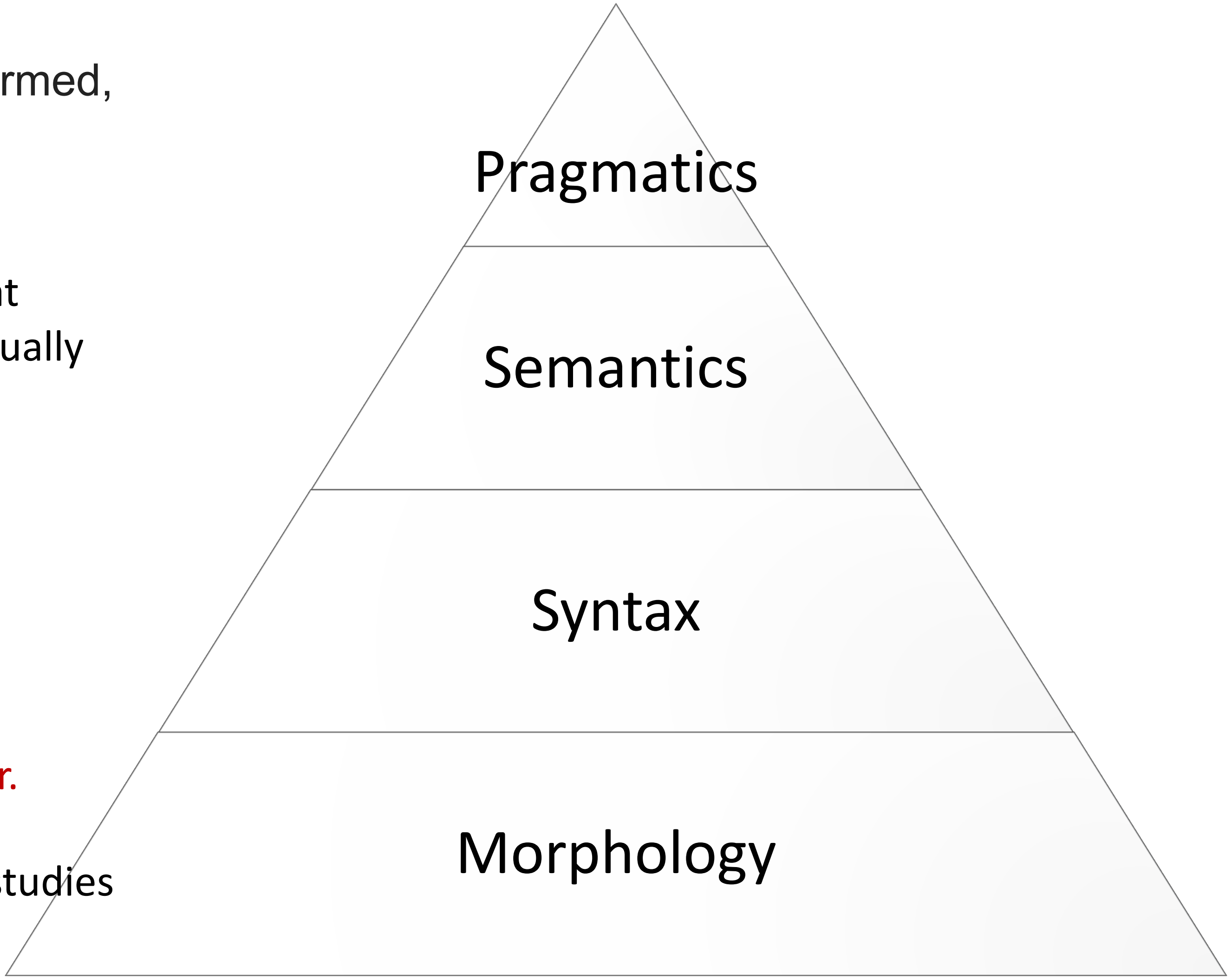
Morphology - is the study of words, how they are formed, and their relationship to other words in the same language.

Syntax - is the set of rules, principles, and processes that govern the structure of sentences in a given language, usually including word order

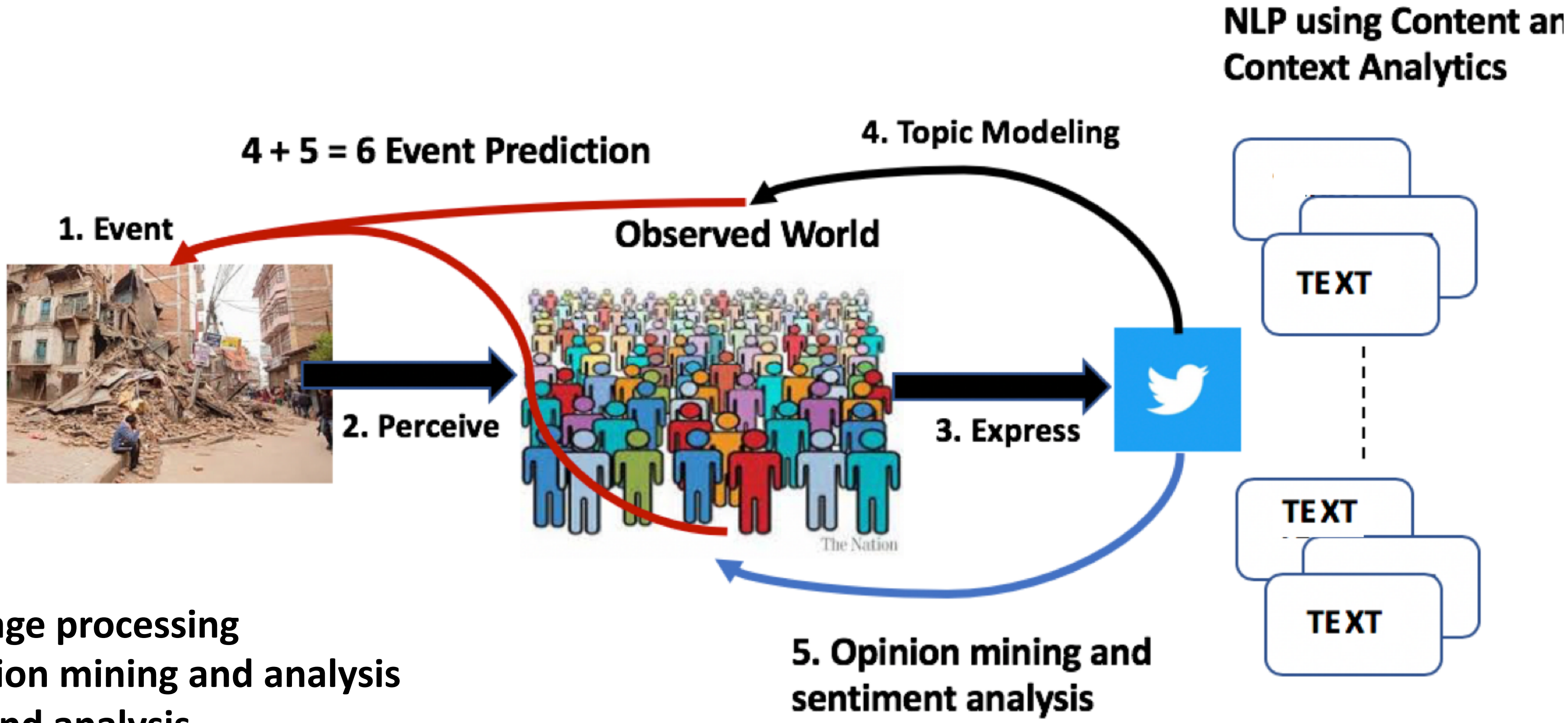
Semantics - is the linguistic and philosophical study of meaning, in language, programming languages, formal logics, and semiotics

Examples of Semantics:
A child could be called a child, kid, boy, girl, son, daughter.

Pragmatics a subfield of linguistics and semiotics that studies the ways in which context contributes to meaning



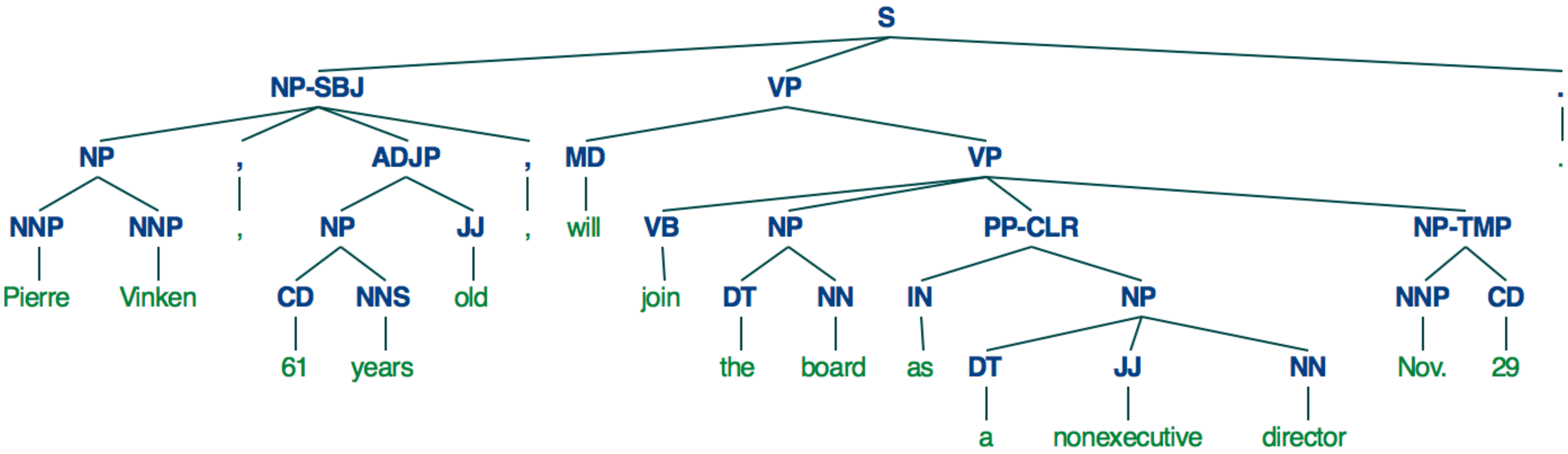
Text Mining



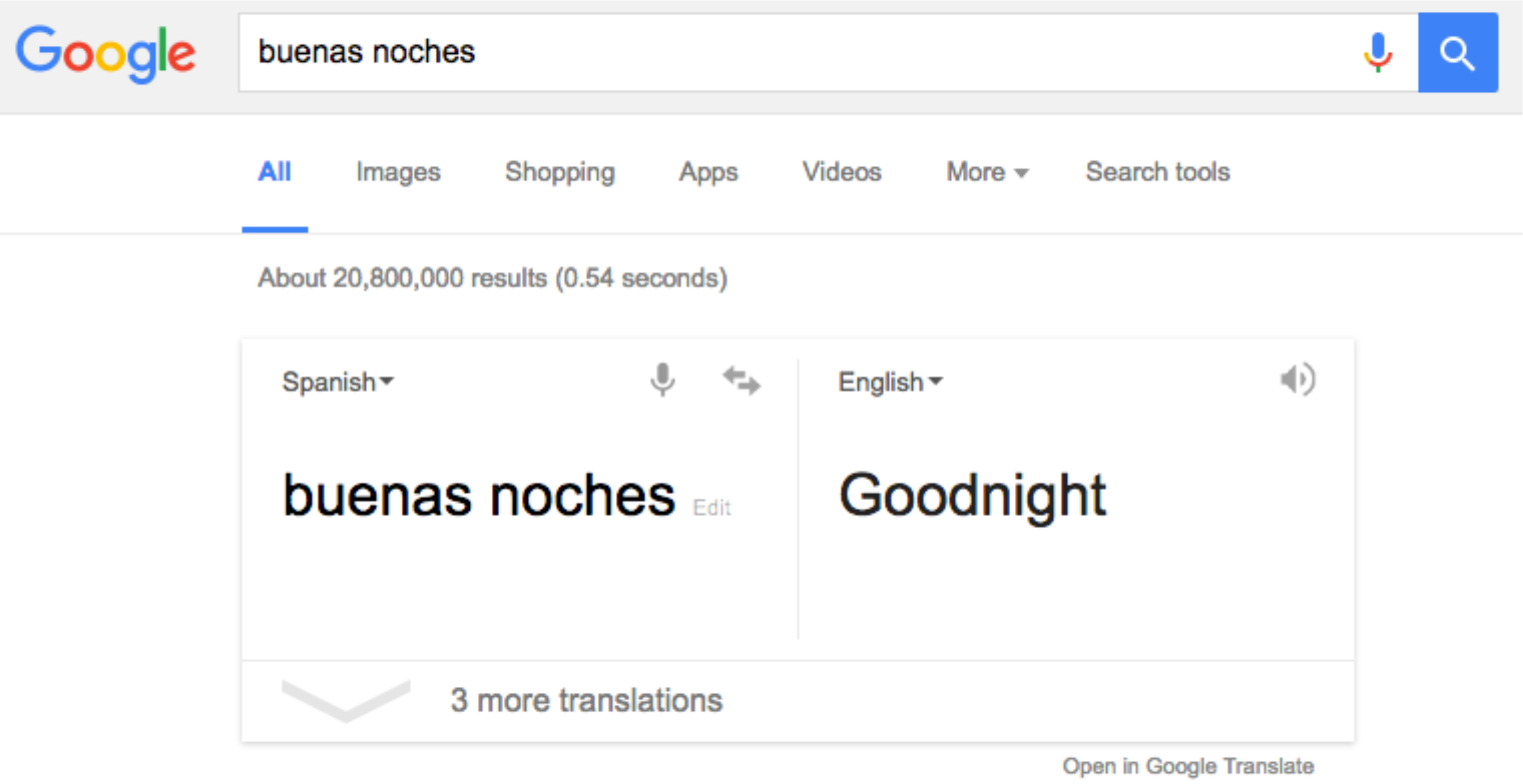
1. Natural language processing
2. Word association mining and analysis
3. Topic mining and analysis
4. Opinion mining and sentiment analysis
5. Event Prediction

Syntactic Analysis

Pierre Vinken 61 years old will join the board as a nonexecutive director Nov. 29



Machine translation



buenas noches

About 20,800,000 results (0.54 seconds)

Spanish

buenas noches Edit

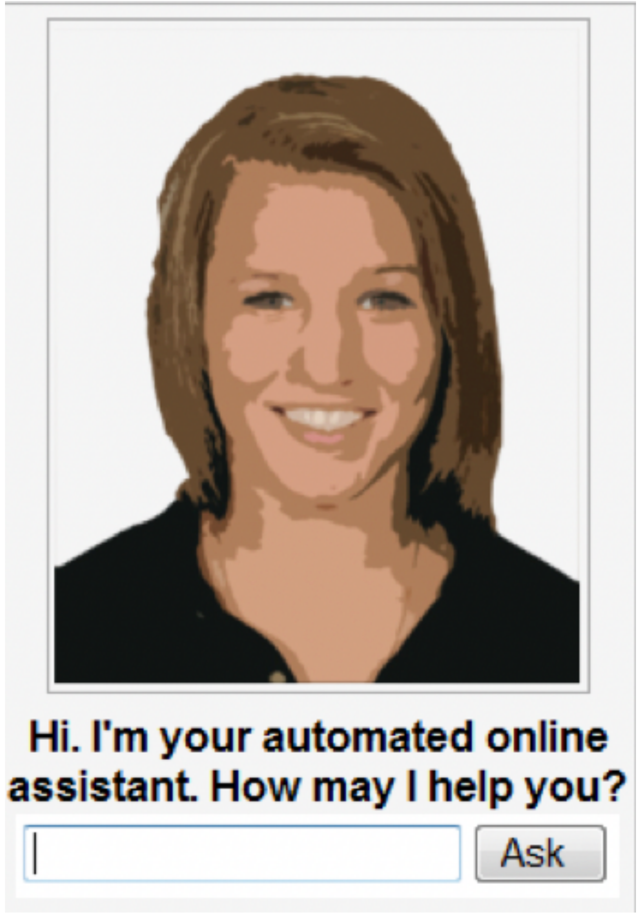
English

Goodnight

3 more translations

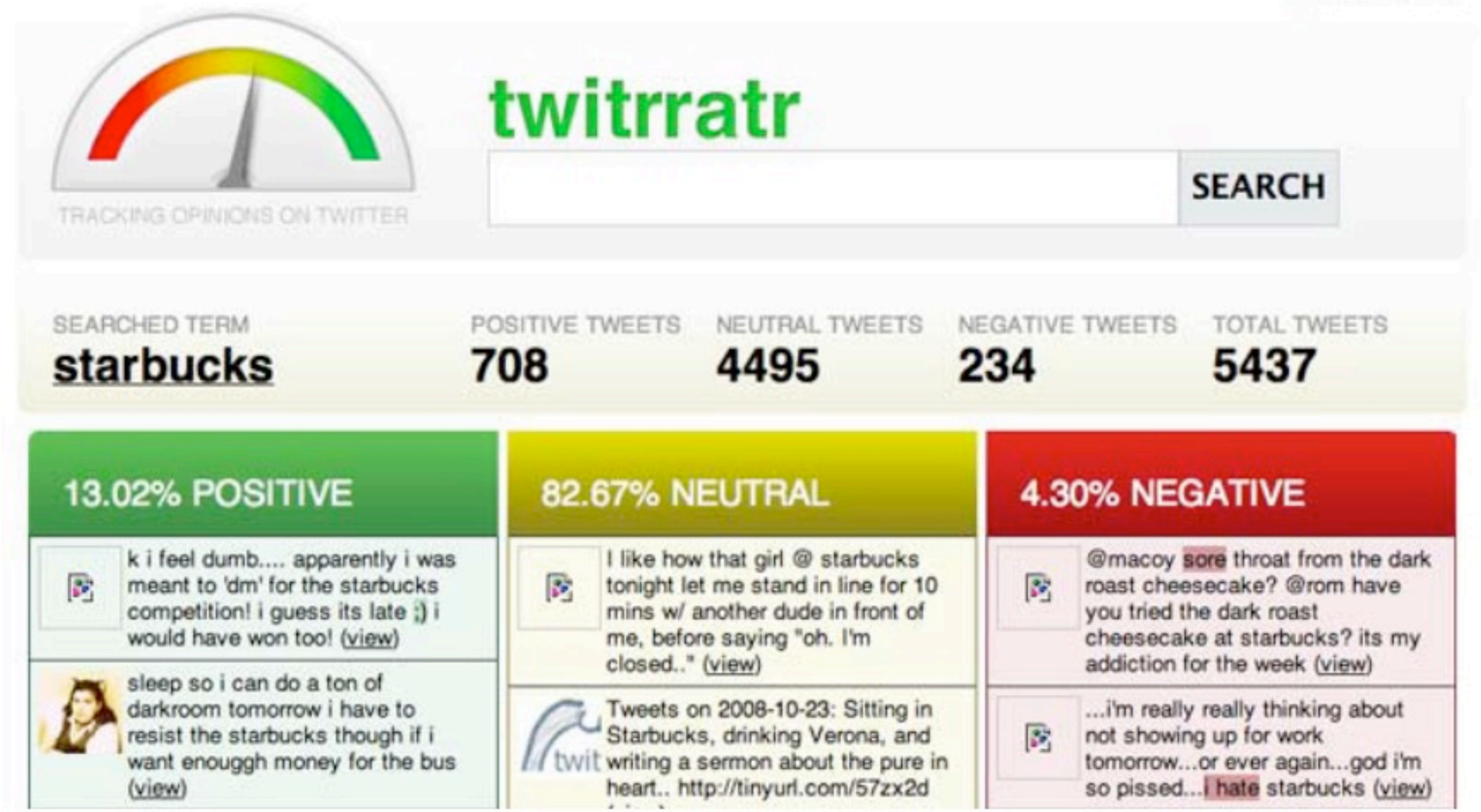
Open in Google Translate

Dialog Systems



Hi. I'm your automated online assistant. How may I help you?

Sentiment

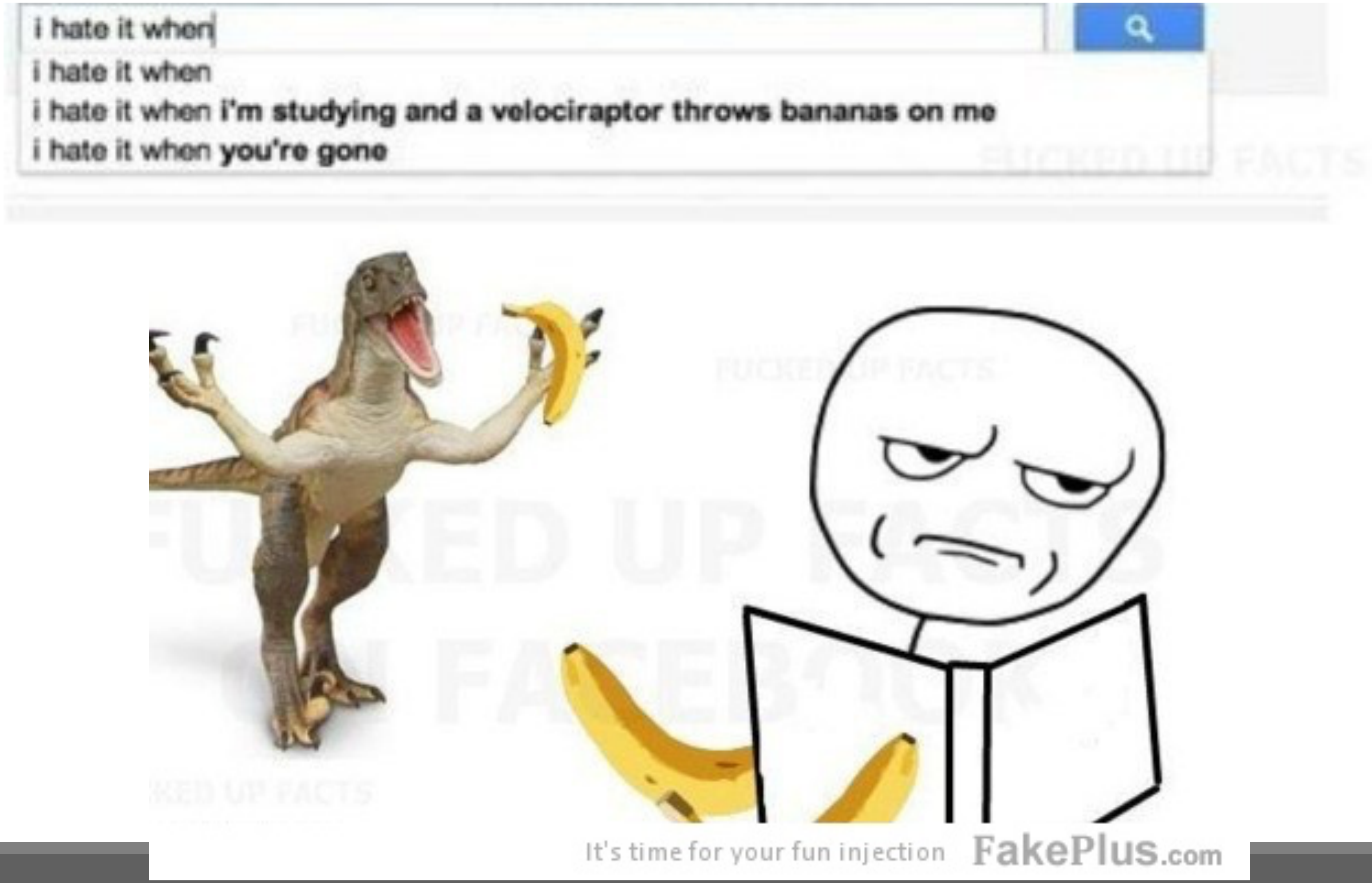


SEARCHED TERM	POSITIVE TWEETS	NEUTRAL TWEETS	NEGATIVE TWEETS	TOTAL TWEETS
starbucks	708	4495	234	5437

13.02% POSITIVE	82.67% NEUTRAL	4.30% NEGATIVE
k i feel dumb.... apparently i was meant to 'dm' for the starbucks competition! i guess its late ;) i would have won too! (view)	I like how that girl @ starbucks tonight let me stand in line for 10 mins w/ another dude in front of me, before saying "oh. I'm closed.." (view)	@macoy sore throat from the dark roast cheesecake? @rom have you tried the dark roast cheesecake at starbucks? its my addiction for the week (view)
sleep so i can do a ton of darkroom tomorrow i have to resist the starbucks though if i want enouggh money for the bus (view)	Tweets on 2008-10-23: Sitting in Starbucks, drinking Verona, and writing a sermon about the pure in heart.. http://tinyurl.com/57zx2d	...i'm really really thinking about not showing up for work tomorrow...or ever again...god i'm so pissed...i hate starbucks (view)

Language model applications

Autocomplete



i hate it when

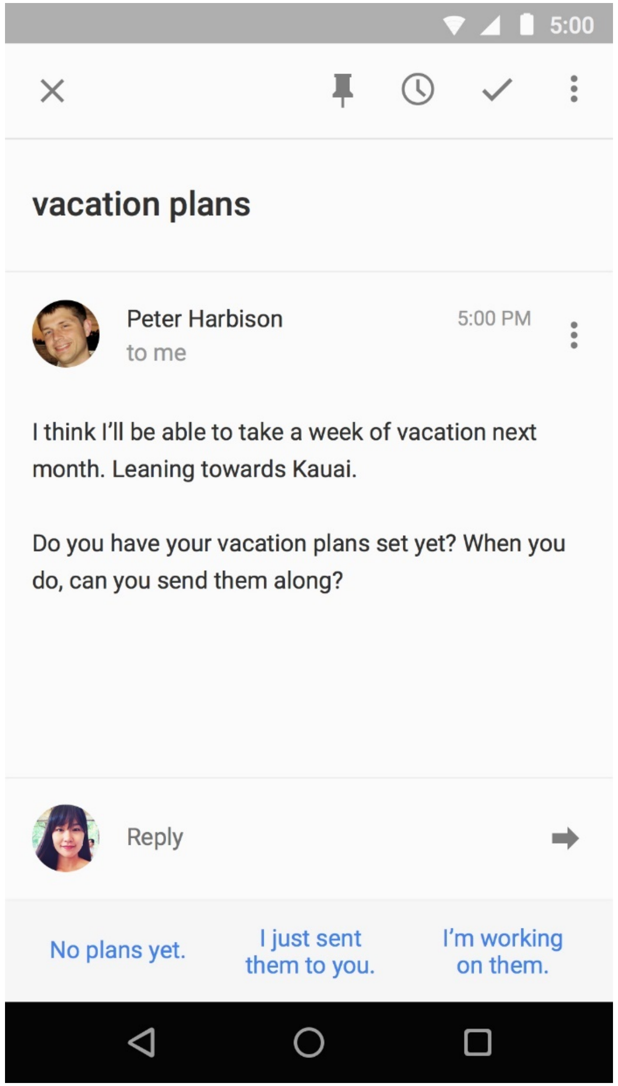
i hate it when

i hate it when **i'm studying and a velociraptor throws bananas on me**

i hate it when **you're gone**

Language model applications

Smart Reply



vacation plans

Peter Harbison
to me
5:00 PM

I think I'll be able to take a week of vacation next month. Leaning towards Kauai.

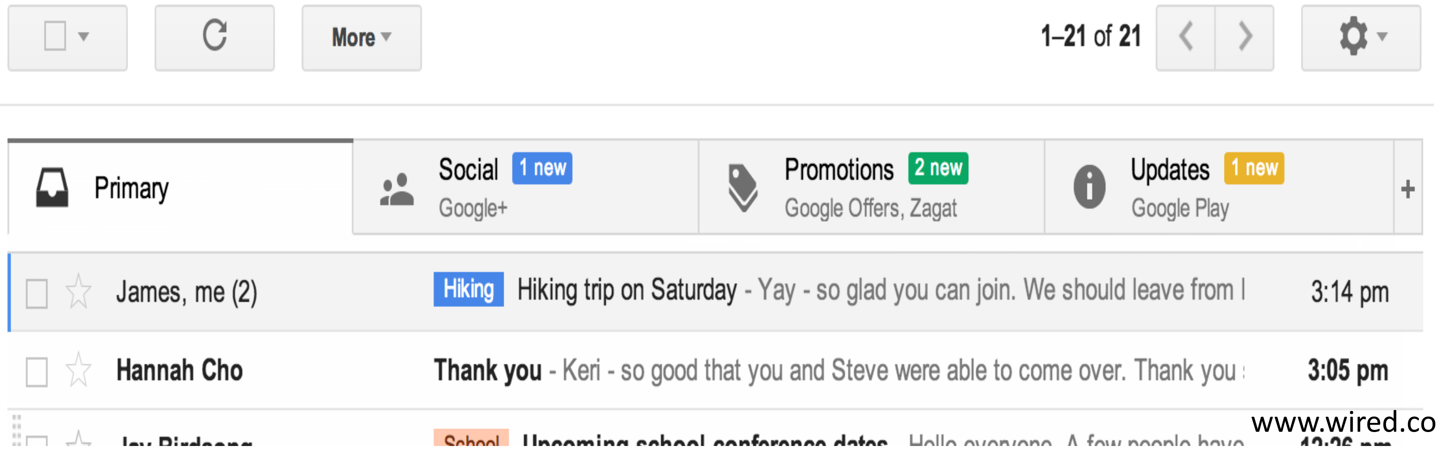
Do you have your vacation plans set yet? When you do, can you send them along?

No plans yet.

I just sent them to you.

I'm working on them.

Text Classification



Primary | Social 1 new Google+ | Promotions 2 new Google Offers, Zagat | Updates 1 new Google Play

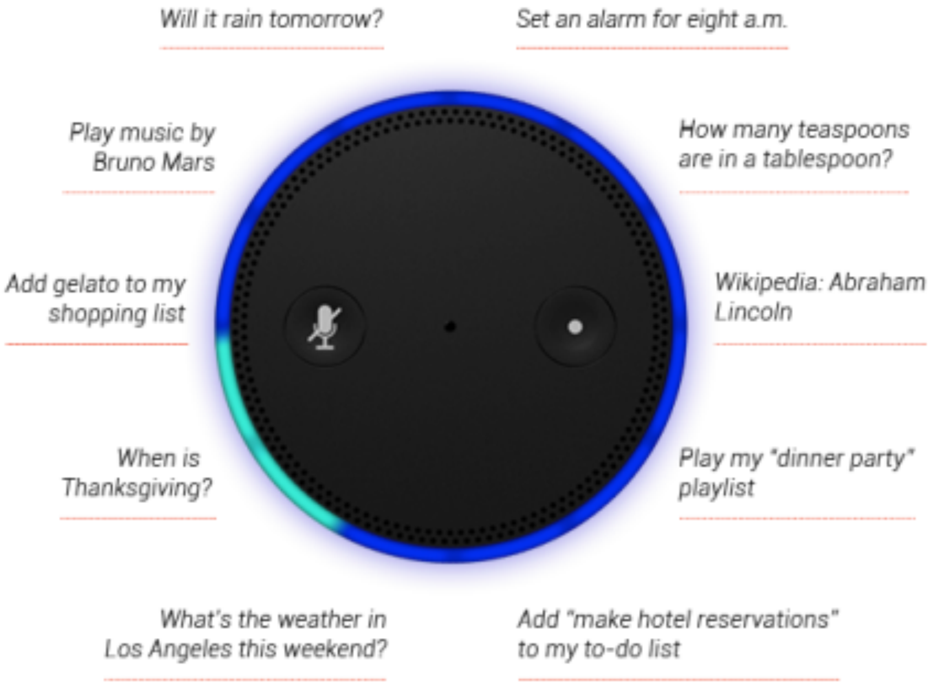
□ ☆	James, me (2)	Hiking Hiking trip on Saturday - Yay - so glad you can join. We should leave from l	3:14 pm
□ ☆	**Hannah Cho**	**Thank you** - Keri - so good that you and Steve were able to come over. Thank you :	**3:05 pm**
□ ☆	**Jay Birdsong**	School **Upcoming school conference dates** - Hello everyone, A few people have	**12:26 pm**

www.wired.com

Question answering

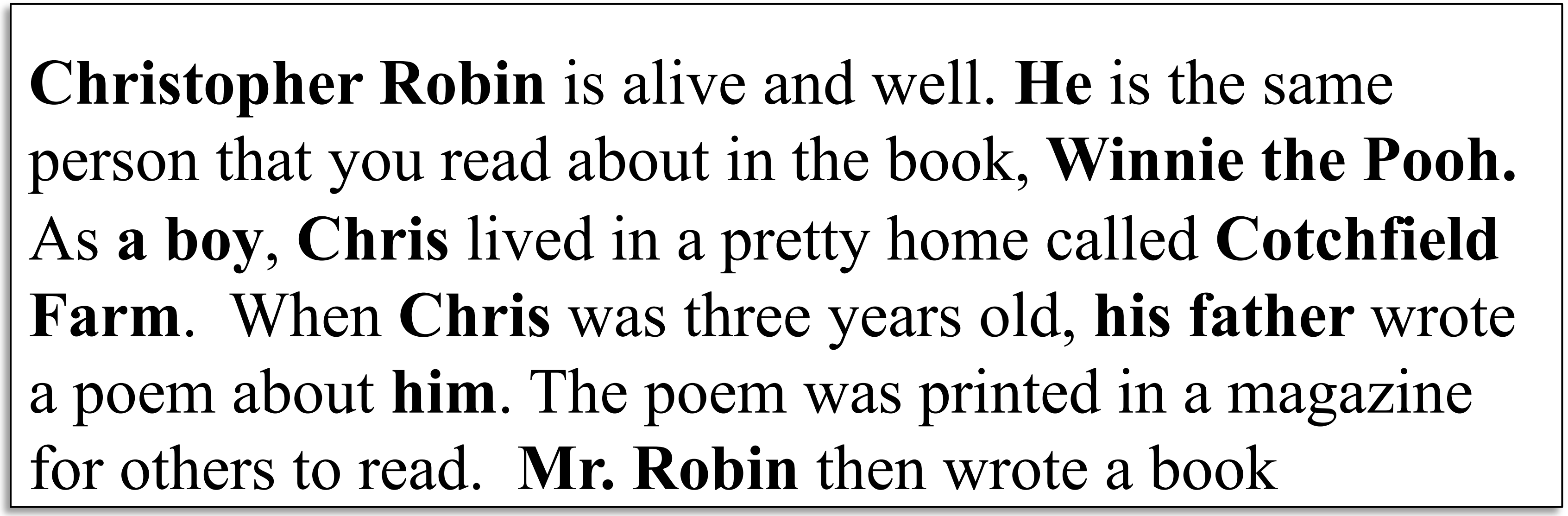
'Watson' computer wins at 'Jeopardy'

Natural language instruction


Will it rain tomorrow?
Set an alarm for eight a.m.
Play music by Bruno Mars
How many teaspoons are in a tablespoon?
Add gelato to my shopping list
Wikipedia: Abraham Lincoln
When is Thanksgiving?
Play my "dinner party" playlist
What's the weather in Los Angeles this weekend?
Add "make hotel reservations" to my to-do list

https://youtu.be/KkOCeAtKHIc?t=1m28s

Language Comprehension



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 **Cotchfield 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

Q: who wrote Winnie the Pooh?

Q: where is Chris lived?

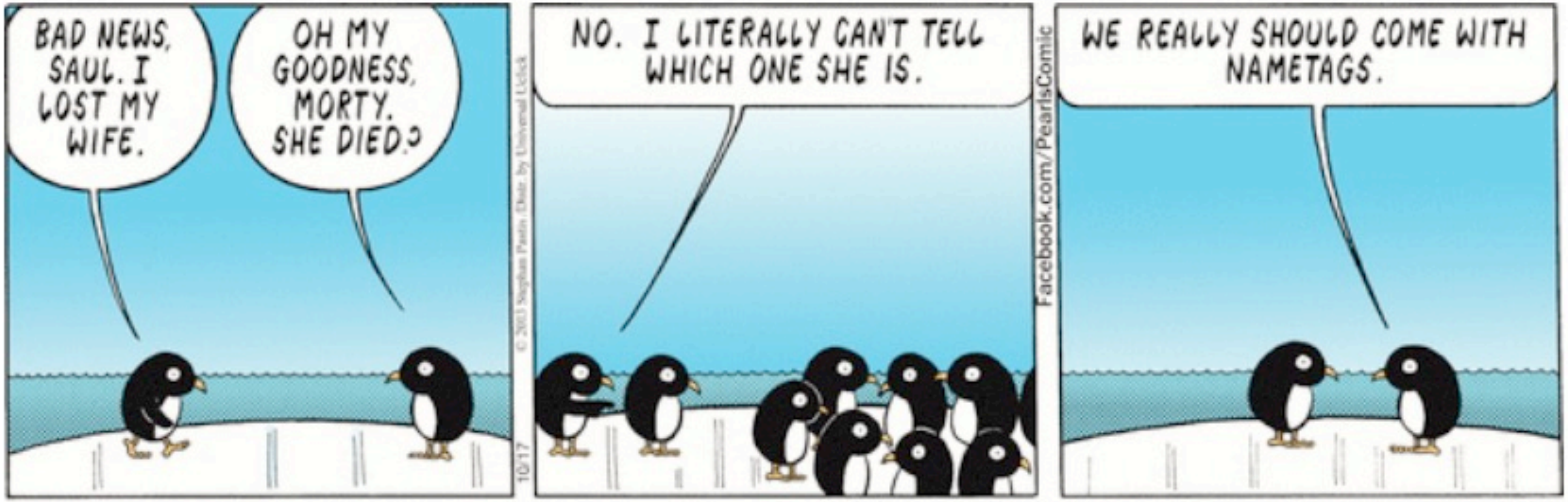
What are the challenges?

Paul Rad, Ph.D.

Associate Professor
Cyber Analytics and AI
Information Systems and Cyber Security
College of Business School
210.872.7259

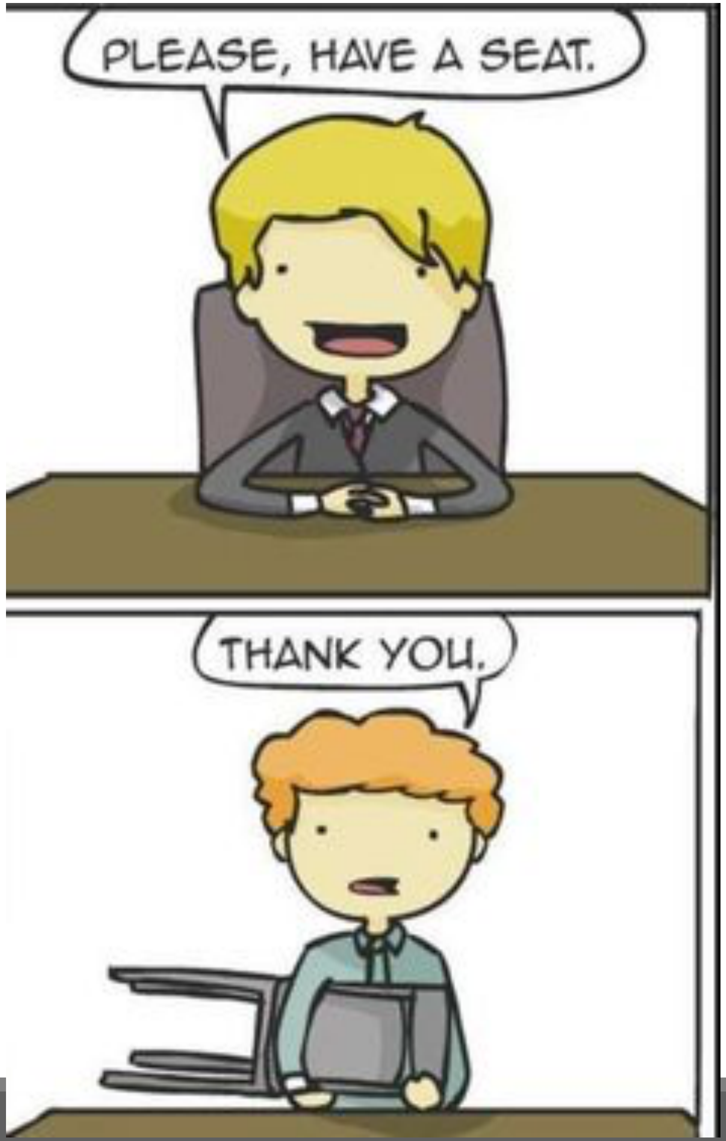
Challenges – ambiguity

Word sense ambiguity


BAD NEWS, SAUL. I LOST MY WIFE.
OH MY GOODNESS, MORTY. SHE DIED?
NO. I LITERALLY CAN'T TELL WHICH ONE SHE IS.
WE REALLY SHOULD COME WITH NAMETAGS.
© 2013 Stephan Pastis/Dist. by Universal Uclick
10/17
Facebook.com/PearlsComic

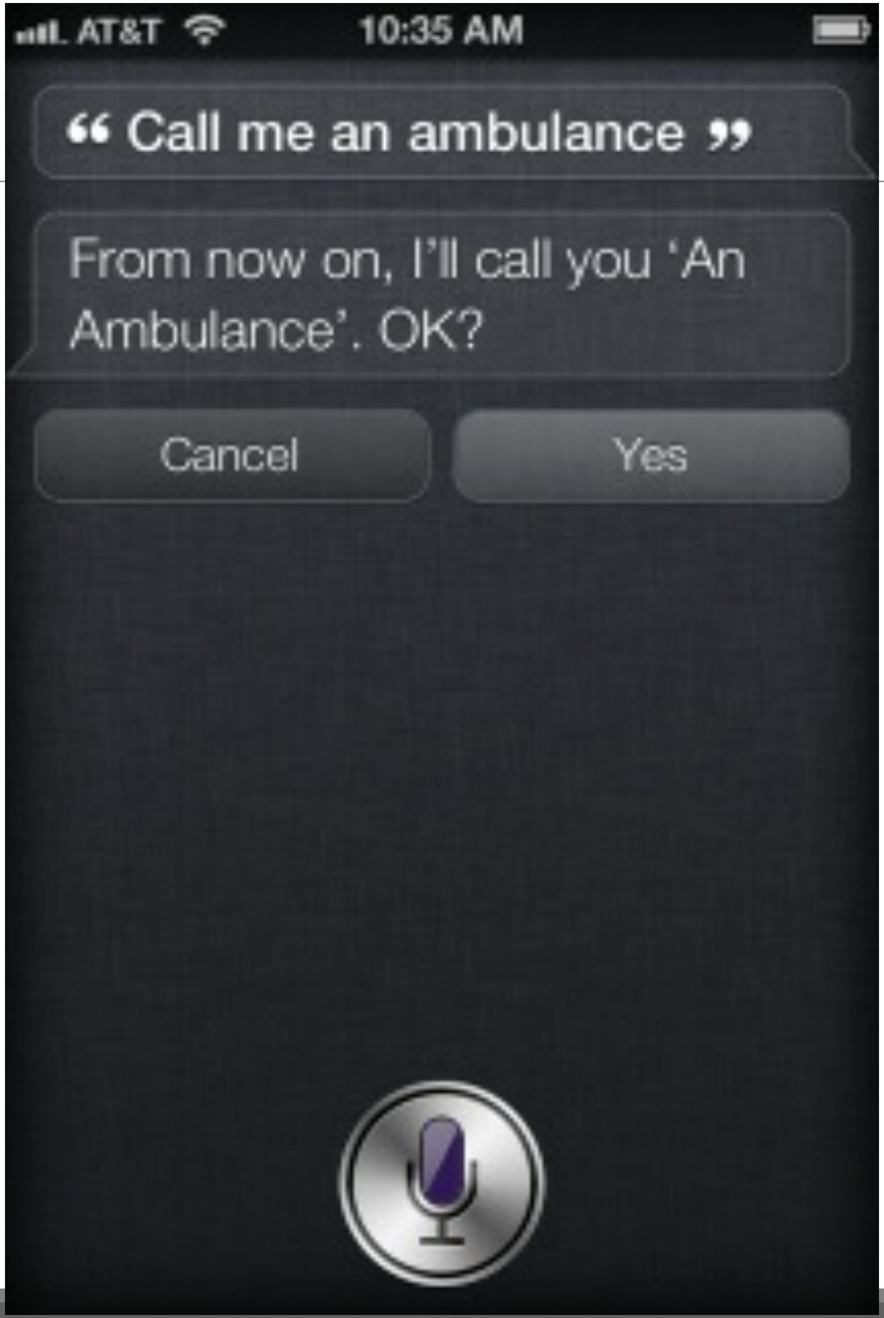
credit: A. Zwicky

Word sense ambiguity



Challenges – ambiguity

Word sense / meaning ambiguity



“Call me an ambulance”

From now on, I’ll call you ‘An Ambulance’. OK?

Credit: http://stuffsirisaid.com

Challenges -- ambiguity

Ambiguous headlines:

- Include your children when baking cookies
- Local High School Dropouts Cut in Half
- Hospitals are Sued by 7 Foot Doctors
- Iraqi Head Seeks Arms

- Safety Experts Say School Bus Passengers Should Be Belted
- Teacher Strikes Idle Kids

Challenges – ambiguity

Pronoun reference ambiguity

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

Credit: http://www.printwand.com/blog/8-catastrophic-examples-of-word-choice-mistakes

Challenges – language is not static

Language grows and changes

- e.g., cyber lingo

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

Challenges – scale

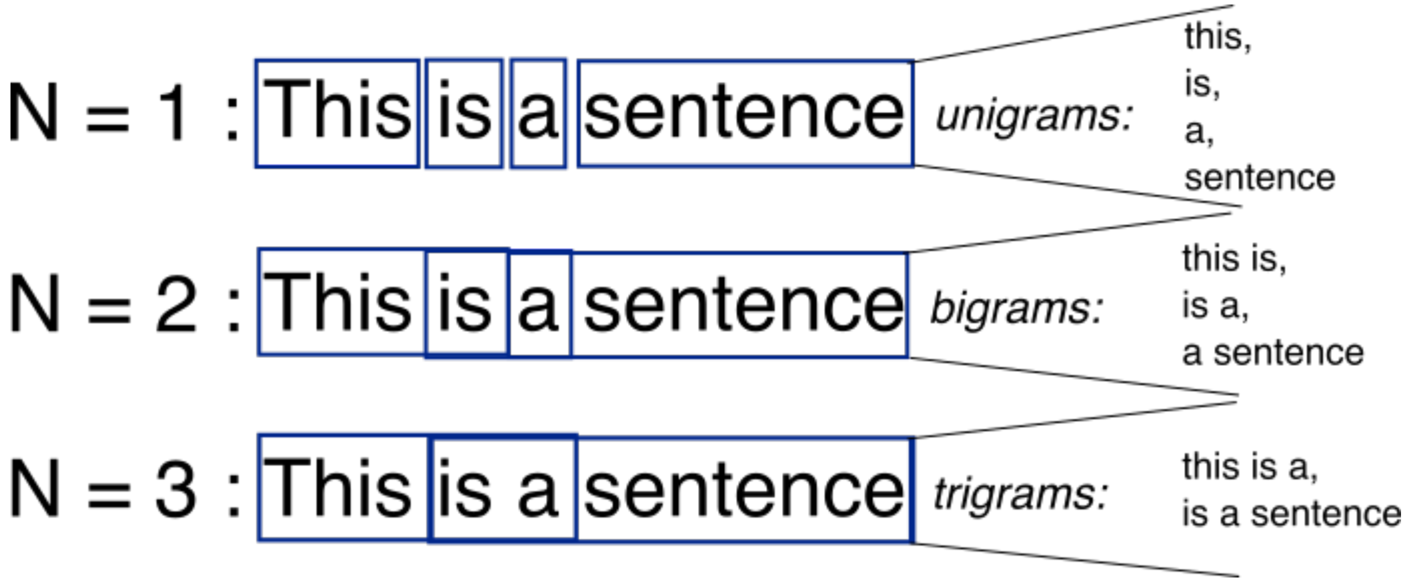
Examples:

- Bible (King James version): ~700K
- Penn Tree bank ~1M from Wall street journal
- Newswire collection: 500M+
- Wikipedia: 2.9 billion word (English)
- Web: several billions of words

Part of Speech Tagging

Bag-of-Words with N-grams

N-grams: a contiguous sequence of n tokens from a given piece of text

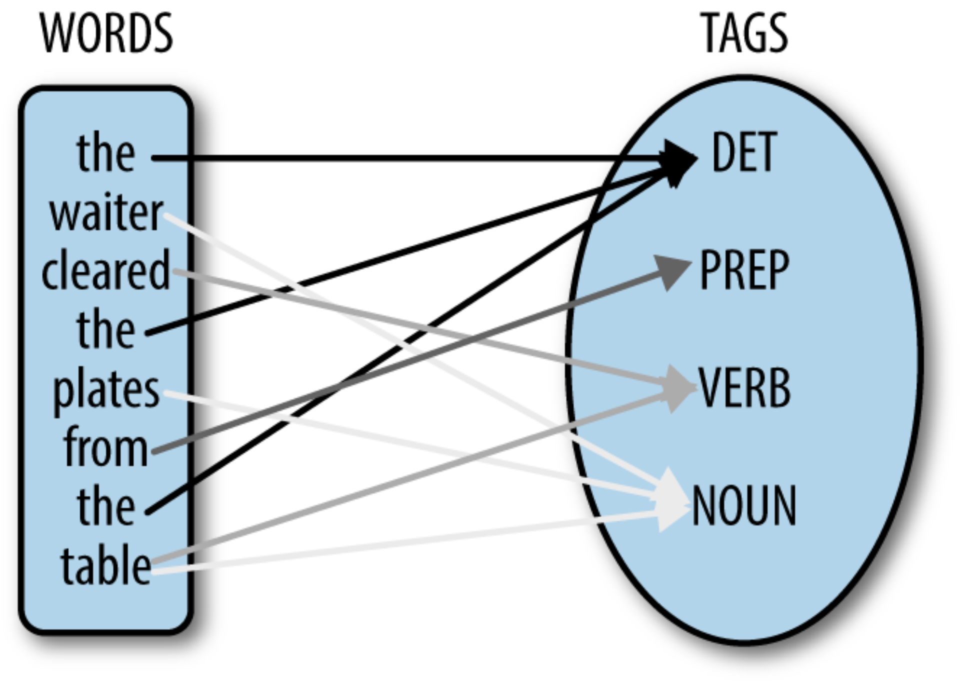


http://recognize-speech.com/language-model/n-gram-model/comparison

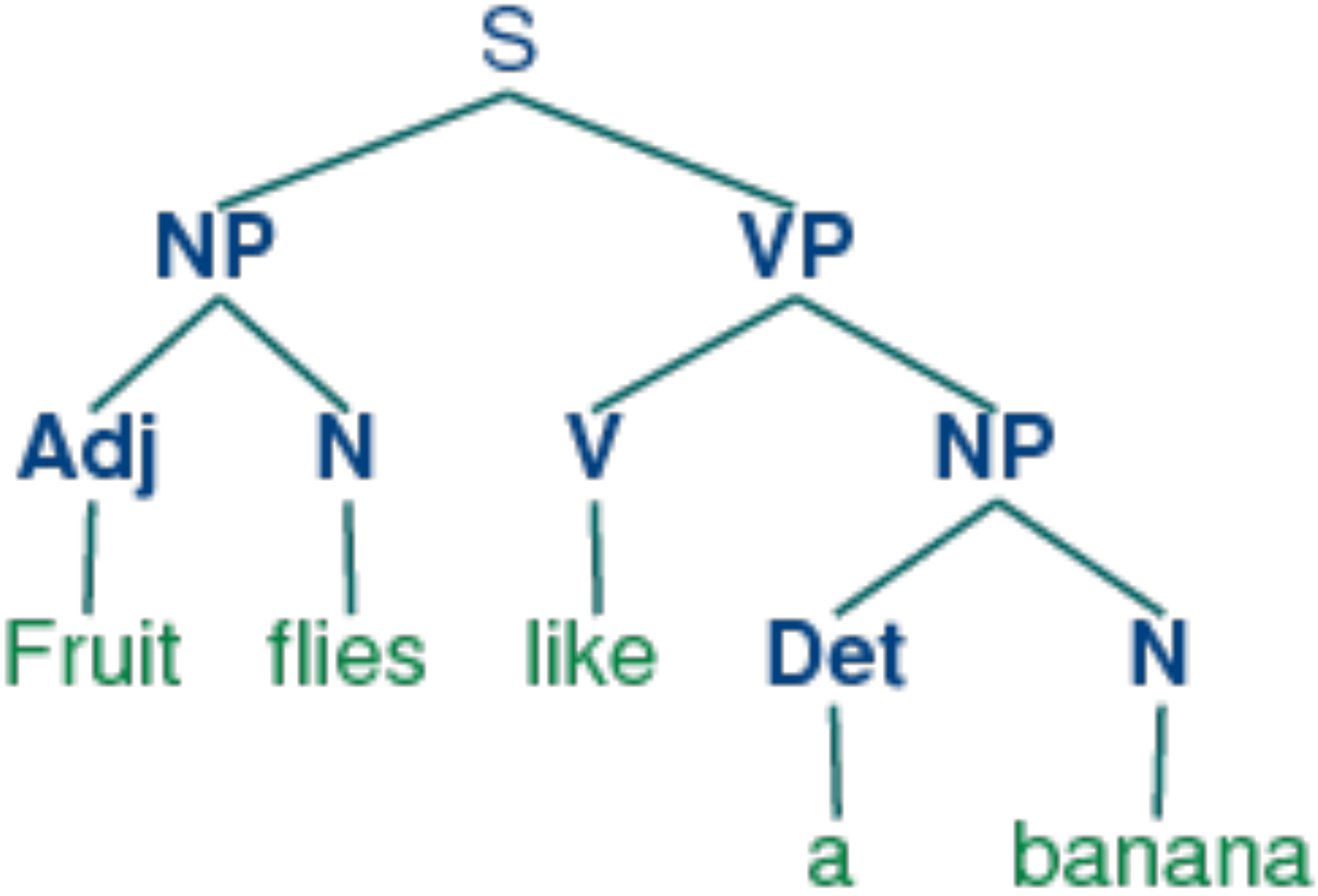
N-(N-grams -1)

Part of Speech Tagging

part-of-speech tagging (POS tagging or PoS tagging or POST is the process of marking up a word in a text based on both its definition and its context—i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. A simplified form of this is commonly taught to school-age children, in the identification of words as nouns, verbs, adjectives, adverbs, etc.



Syntactic parsing


S
NP
VP
Adj
N
V
NP
Fruit
flies
like
Det
N
a
banana