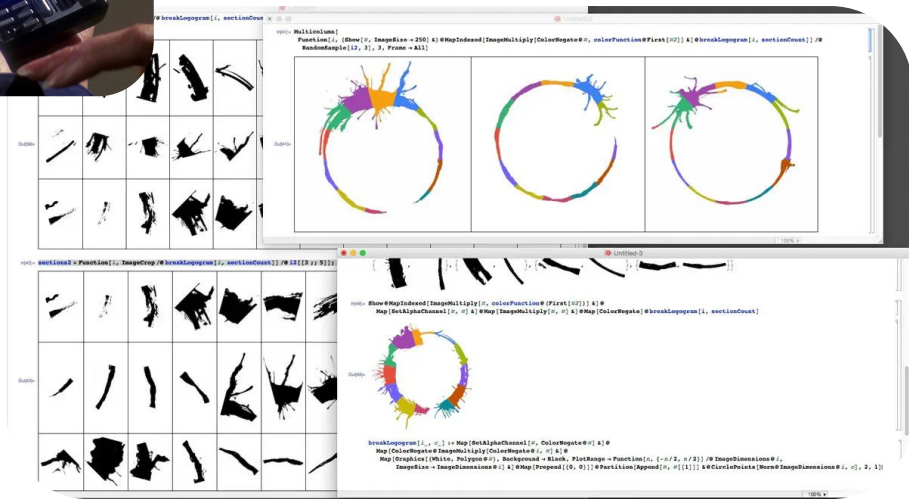


Welcome to **Applications of Natural Language Processing** (CS 490A)!

Prof. Thompson & Prof. O'Connor, Fall 2021

[https://people.cs.umass.edu/~brenocon/cs490a\\_f21/](https://people.cs.umass.edu/~brenocon/cs490a_f21/)



# Language is Ambiguous – Why NLP is hard

- What is your purpose on Earth?
- Red Tape Holds Up New Bridges
- March Planned For Next August
- Two Sisters Reunited After 18 Years in Checkout Counter
- Lawyers Give Poor Free Legal Advice
- At the drop of a hat
- As fit as a fiddle

# Levels of Linguistic Structure

This is a simple sentence.

# Levels of Linguistic Structure: Characters

T	h	i	s		i	s		a		s	i	m	p	l	e		s	e	n	t	e	n	c	e	.
---	---	---	---	--	---	---	--	---	--	---	---	---	---	---	---	--	---	---	---	---	---	---	---	---	---

# Levels of Linguistic Structure: Morphology

This	is	a	simple	sentence	.
------	----	---	--------	----------	---

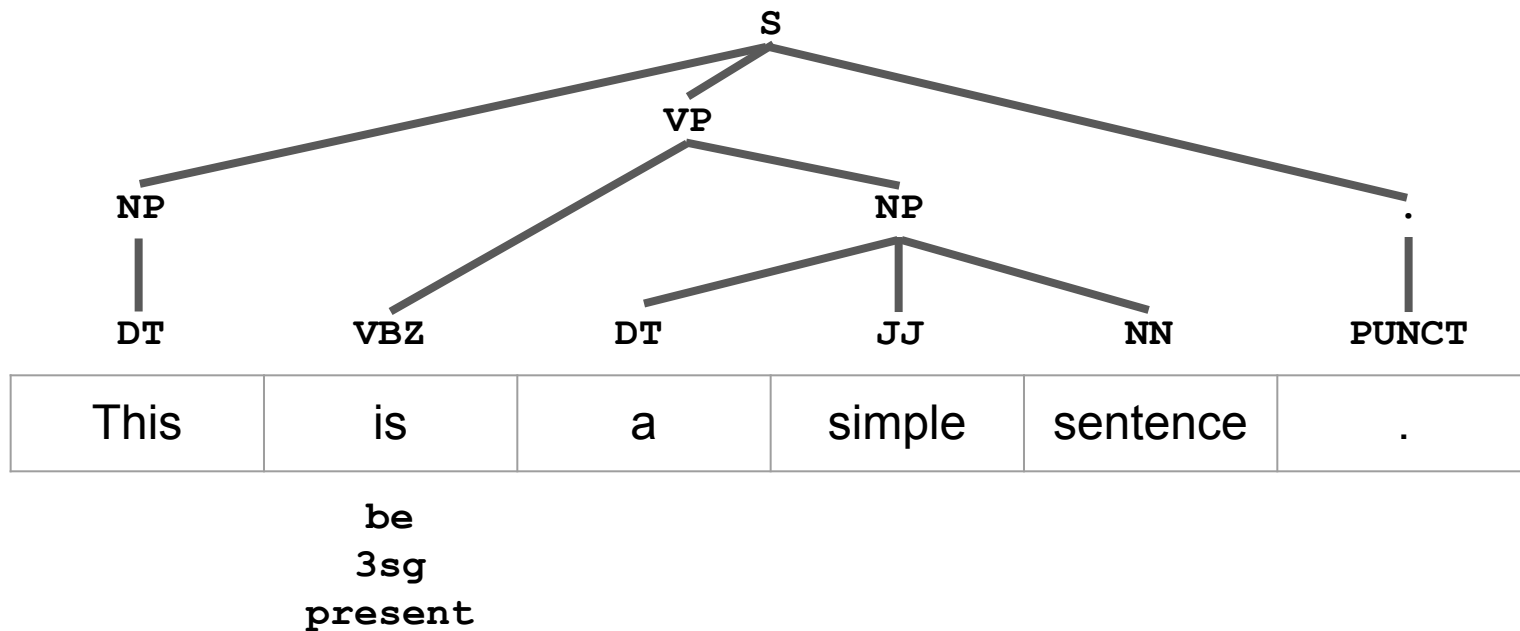
**be**  
**3sg**  
**present**

# Levels of Linguistic Structure: Parts of Speech

DT	VBZ	DT	JJ	NN	PUNCT
This	is	a	simple	sentence	.

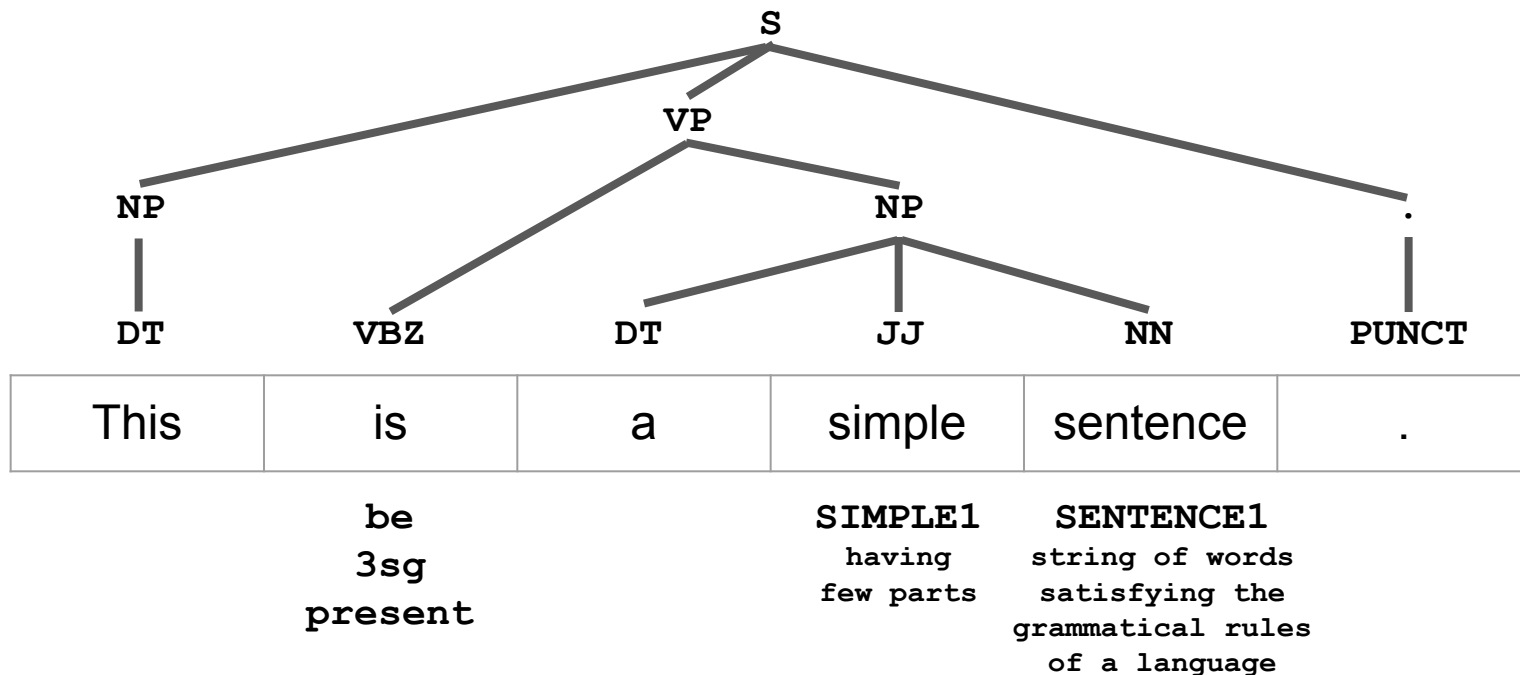
be  
3sg  
present

# Levels of Linguistic Structure: Syntax

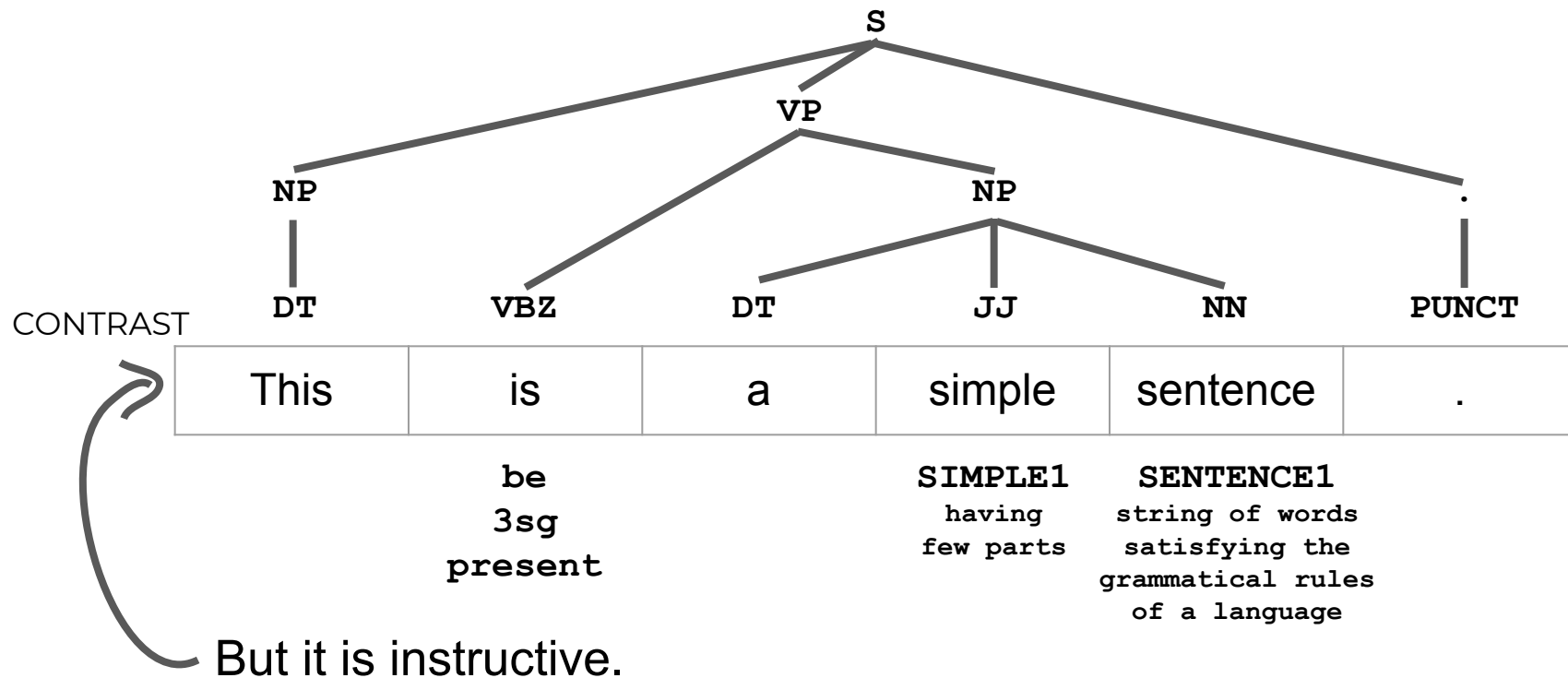




# Levels of Linguistic Structure: Semantics



# Levels of Linguistic Structure: Discourse



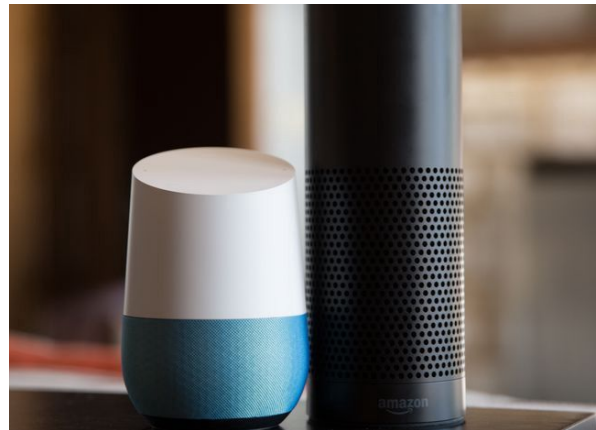
# NLP Today: Speech Interfaces & Voice Assistants



# NLP Today: Speech Interfaces & Voice Assistants



Ophelia, call the  
police



Sure. Playing F\*\* tha  
Police by N.W.A.

# NLP Today: Machine Translation

## 三体 (小说) [编辑]

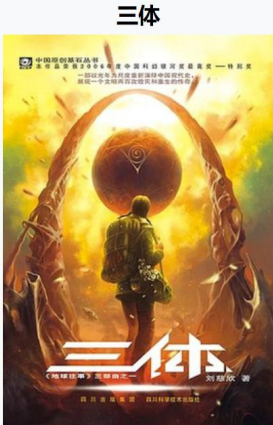
维基百科，自由的百科全书

本條目介紹的是刘慈欣所著之科幻小说。关于天体力学中的基本力学模型，请见「[三体问题](#)」。

《**三体**》是中國大陸作家**刘慈欣**于2006年5月至12月在《**科幻世界**》杂志上连载的一部长篇科幻小说，出版后成为**中国大陆**最畅销的科幻长篇小说之一<sup>[1]</sup>。2008年，该书的单行本由重庆出版社出版。本书是**三体系列**（系列原名为：地球往事三部曲）的第一部，该系列的第二部《**三体II：黑暗森林**》已经于2008年5月出版。2010年11月，第三部《**三体III：死神永生**》出版发行。2011年，“地球往事三部曲”在台湾陆续出版。小说的英文版獲得美國科幻奇幻作家協會2014年度「**星云獎**」提名<sup>[2]</sup>，并荣获2015年**雨果奖**最佳小说奖。<sup>[3]</sup>

### 目录 [隐藏]

- 1 主体脉络
- 2 世界观设定
- 3 内容概述
- 4 出场角色
- 5 用语
- 6 改动



《三体》封面

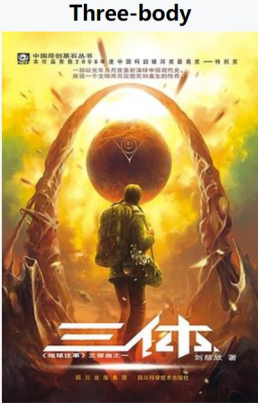
作者 [刘慈欣](#)  
类型 [科幻小说](#)  
系列 [中国科幻基石丛书](#)

## Three Body (Novel) [Edit]

Wikipedia, the free encyclopedia

This article introduces science fiction r celestial mechanics, please see "[Three-Body Problem](#)".

"**Three Body**" is a full-length science fiction novel serialized by the Chinese writer [Liu Cixin](#) in the "[World of Science Fiction](#)" magazine from May to December 2006. After publication, it became one of the best-selling science fiction novels in [Mainland China](#) <sup>[1]</sup>. In 2008, the pamphlet of the book was published by Chongqing Publishing House. This book is the first part of the [three-body series](#) (the series was originally called the Trilogy on Earth). The second part of the series "[Three-body II: Dark Forest](#)" was published in May 2008. In November 2010, the third "[Three-Body III: Death Eternal Life](#)" was published. In 2011, the "Earth Past Trilogy" was successively published in Taiwan. The English version of the novel was nominated for the 2014 "[Nebula Award](#)" by the American Association of Science Fiction and Fantasy Writers <sup>[2]</sup> and won the 2015 [Hugo Award](#) for Best Novel. <sup>[3]</sup>



"Three Body" Cover

author [Liu Cixin](#)  
type [science fiction](#)  
series [China Science Fiction Cornerstone Series](#)

Chinese (Simplified)

English

⋮

✕

☒ Always translate Chinese (Simplified)

Google Translate

# NLP Today: Machine Translation

LATIN - DETECTED

MACEDONIAN

BELARUSIAN

↕

↔

ENGLISH

LATIN

ESTONIAN

↕

Ense petit placidam sub libertate quietem|

×

With his sword he seeks a peaceful repose under freedom

*By the sword we seek peace, but peace  
only under liberty*

# NLP Today: Sentiment Analysis

---

Food was **ok**...*not* the money they charge. I was **unimpressed** and will *not* return. I was **excited** to try this place and was so **disappointed** as my expectations were high. Service *not* **great** and The parking is **awful**.

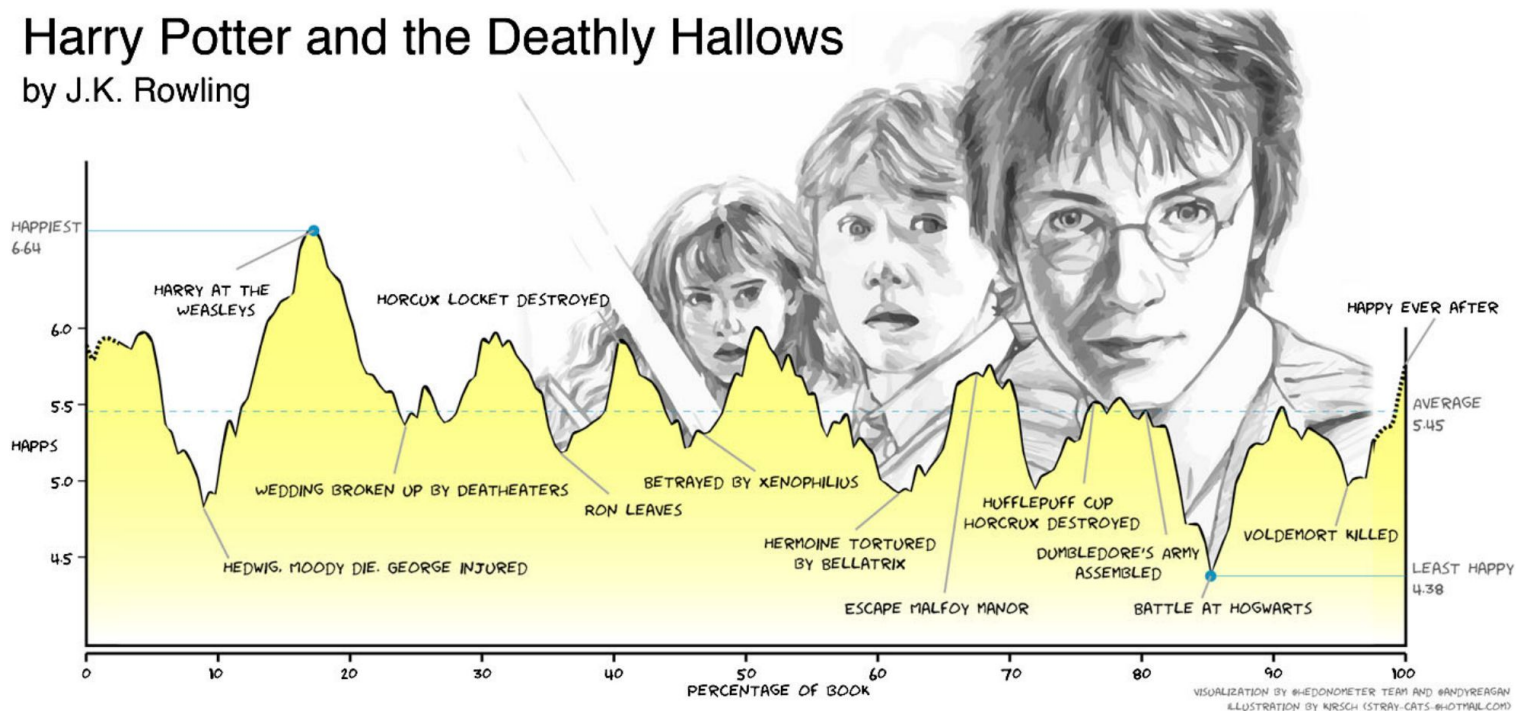
---



# NLP Today: Sentiment Analysis

## Harry Potter and the Deathly Hallows

by J.K. Rowling

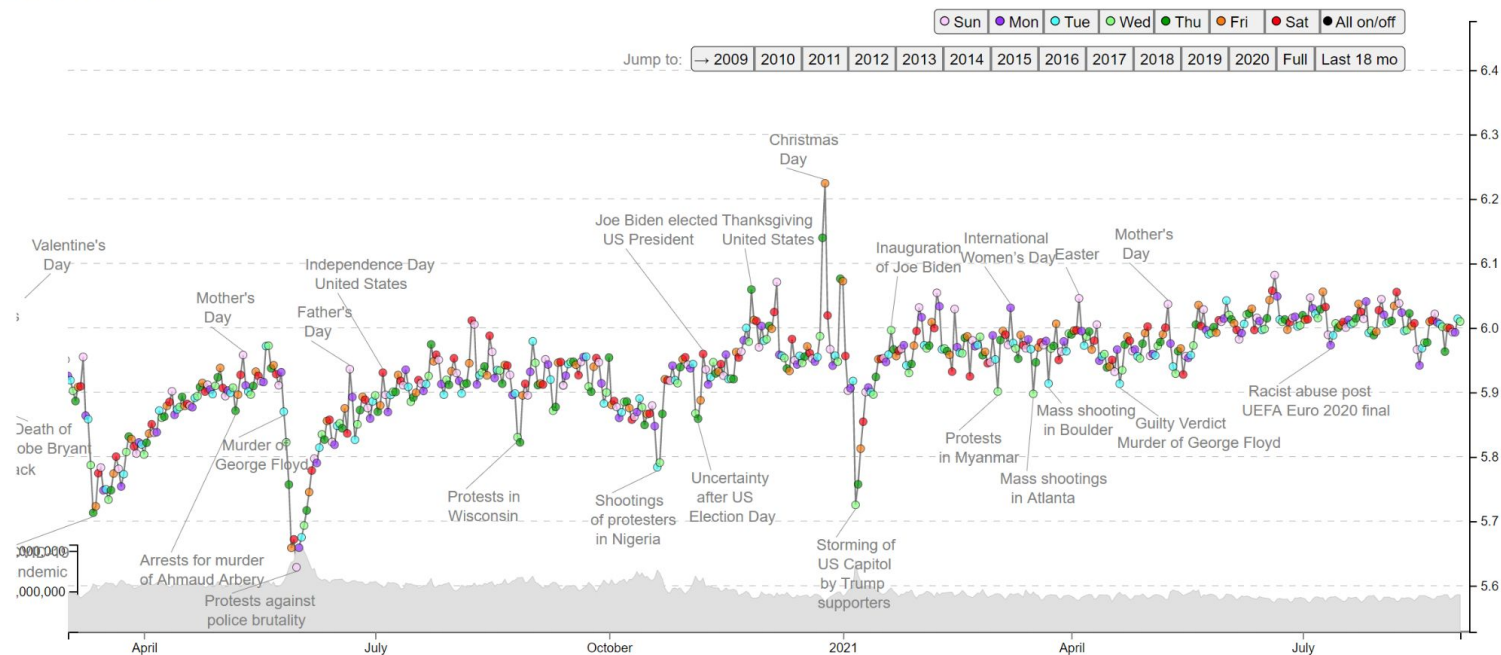




# NLP Today: Sentiment Analysis

Average Happiness for Twitter

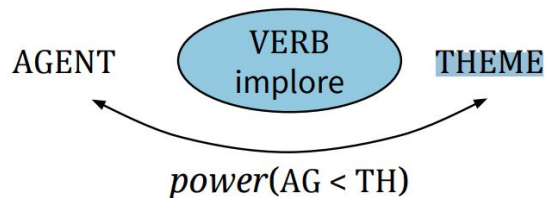
All Tweets in English.



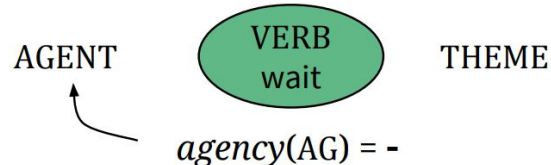
# NLP Today: Studying Social & Power Dynamics



He **implored** the tribunal to show mercy.

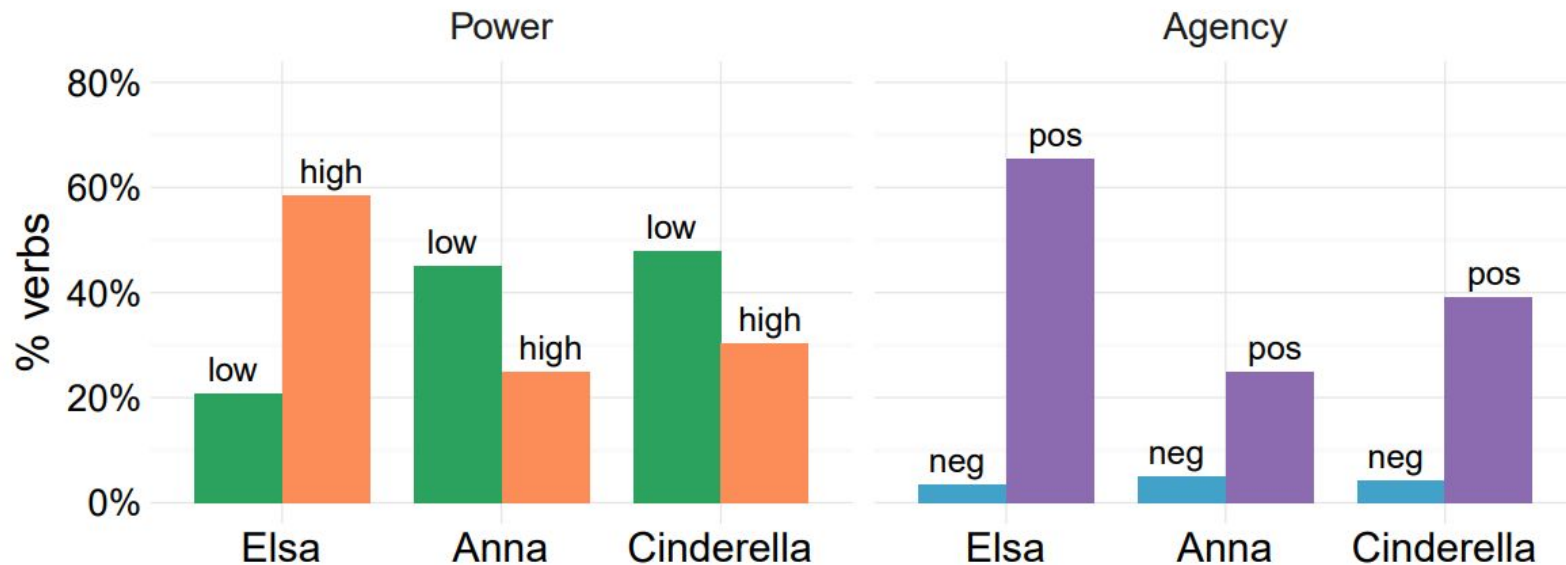


The princess **waited** for her prince.



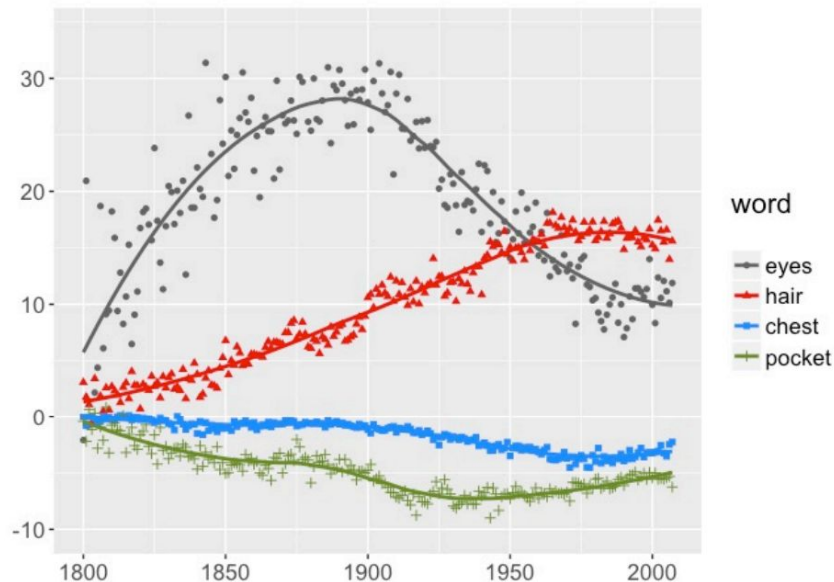
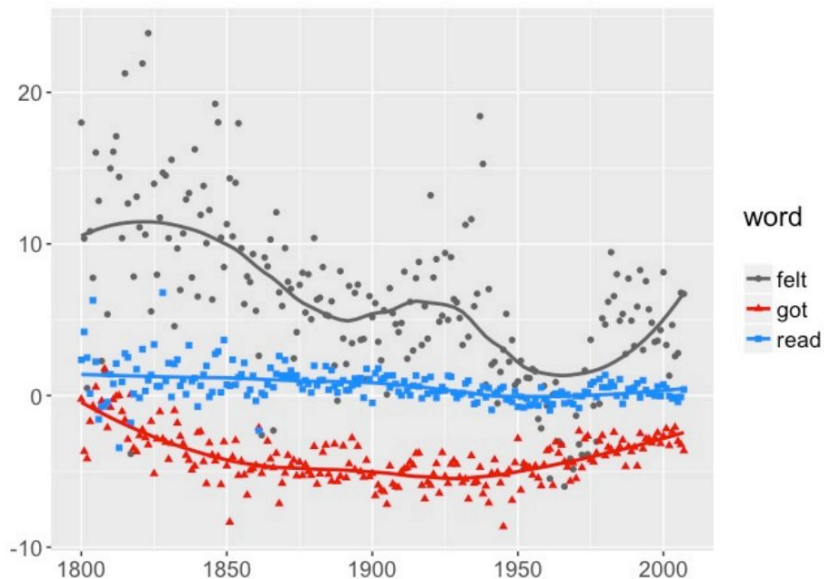
# NLP Today: Studying Social & Power Dynamics

Power and Agency levels of Disney princesses



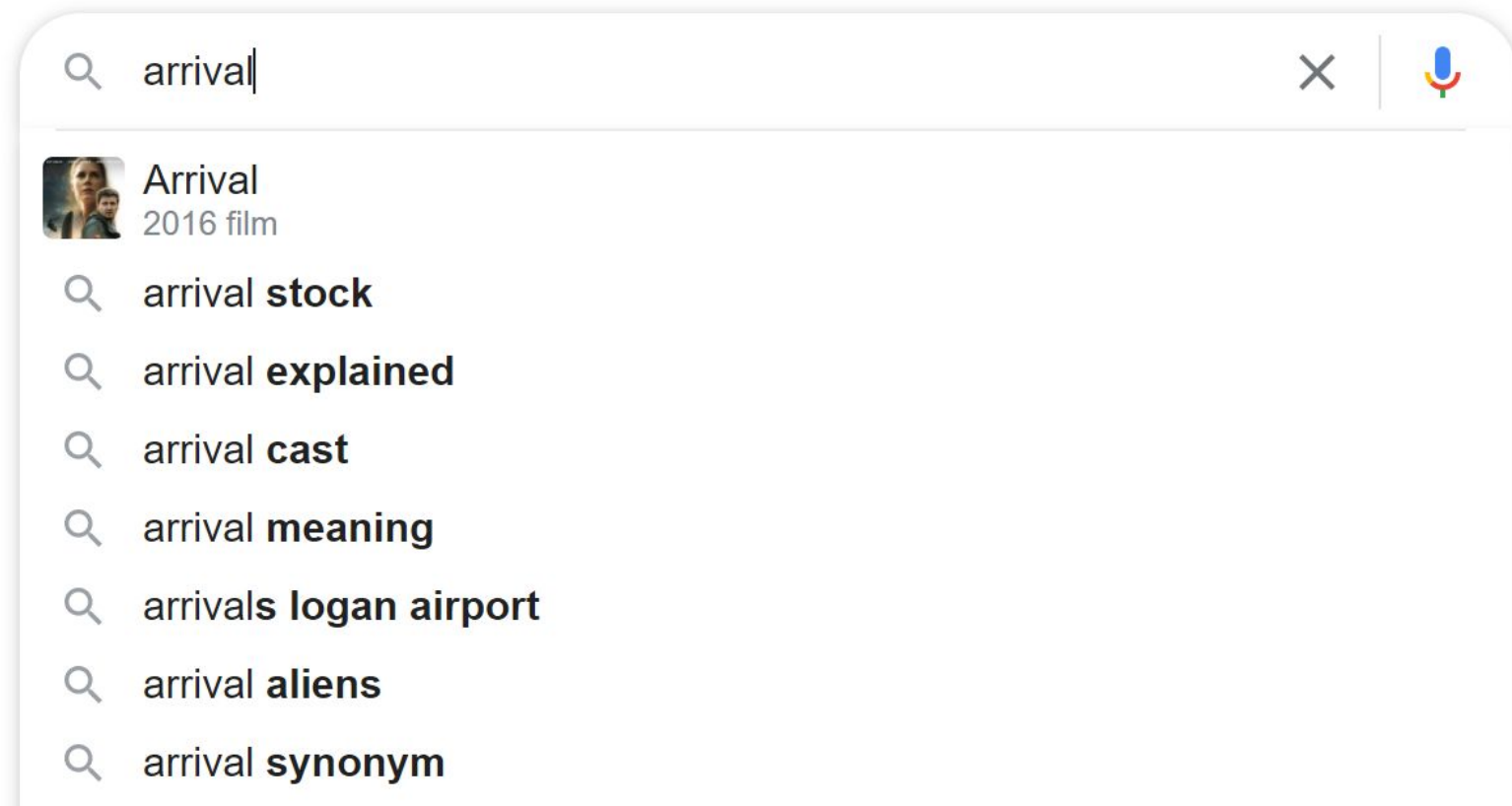
# NLP Today: Studying Social & Power Dynamics

## Uses for Women - Uses for Men in Fiction Novels over Time

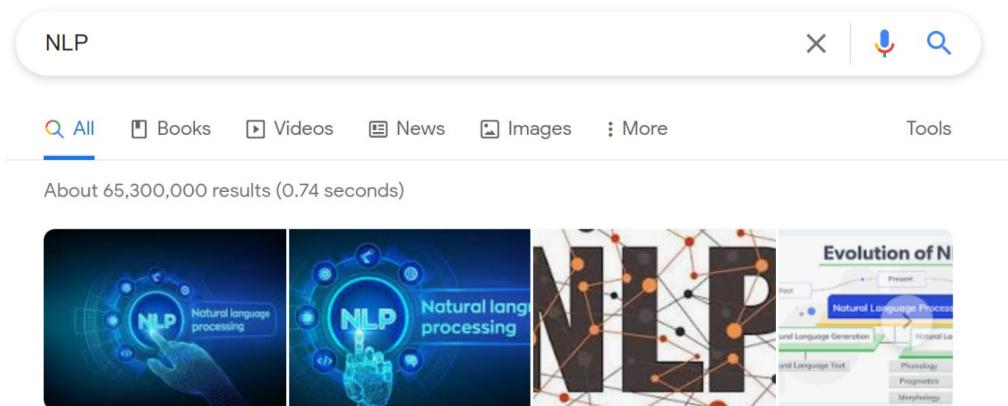


Underwood et al. 2018

# NLP Today: Search & Summarization



# NLP Today: Search & Summarization



**Natural language processing (NLP)** refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

Jul 2, 2020


<https://www.ibm.com> › Cloud › Cloud Learn

[natural language processing \(NLP\) - IBM](#)

# NLP Today: Search & Summarization

TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

## Entity Extraction



- Have technology (thanks to R6) – for English, Arabic and Chinese
- Allow queries like:
- Show me all the word documents with references to IAEO
- Show me all documents that reference Osama Bin Laden
- Will allow a 'show me more like this' capability

TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

- See Exercise #1 on Gradescope – class survey!
- HWo, a quick math review, is coming ~tomorrow; will be due next ~Thursday
- Long and Chloe are teaching an optional Python tutorial next week. Please respond to Long's scheduling request (Piazza)!