

# CS 7337 – Natural Language Processing

## Midterm Exam

### By Ben Goodwin

**Instructions:** Clarity of answers is more important than length of answers. Although not required (unless indicated otherwise), feel free to use graphs, charts, visuals, et al in your answers if you feel these artifacts can help support your answers. There are no bonus points for using these artifacts.

**Due date:** See instructors note; submission should be in PDF or Word DOCX file format.

---

#### **Q1. a. [5 pts] Define homonymy and polysemy and give an example of each.**

Homonymy – Refers to the existence of two or more words having the same spelling or pronunciation but different meanings and origins.

Examples: Stalk

Examples: The main stem of a herbaceous plant

Examples: Pursue or approach stealthily

Polysemy – Refers to the coexistence of many possible meanings for a word or phrase. In the examples below we can see that sometimes a polysemous word can be a noun or verb.

Examples: He drank a glass of milk. He forgot to milk the cow.

Examples: He fixed his hair. They fixed a date for the wedding.

#### **b. [5 pts] Define NLU and NLG and give an example of each.**

Natural Language Understanding: Subset of natural language processing, which uses syntactic and semantic analysis of text and speech to determine the meaning of a sentence. Syntax refers to the grammatical structure of a sentence and semantics alludes to its intended meaning.

In plain English: Aim to get a machine to produce useful representation of inputted natural language.

Examples of NLU: Text annotation, corpus analytics, search applications, machine translation

Natural Language Generation: Subset of natural language processing. The process of producing a human language text response based on some data input.

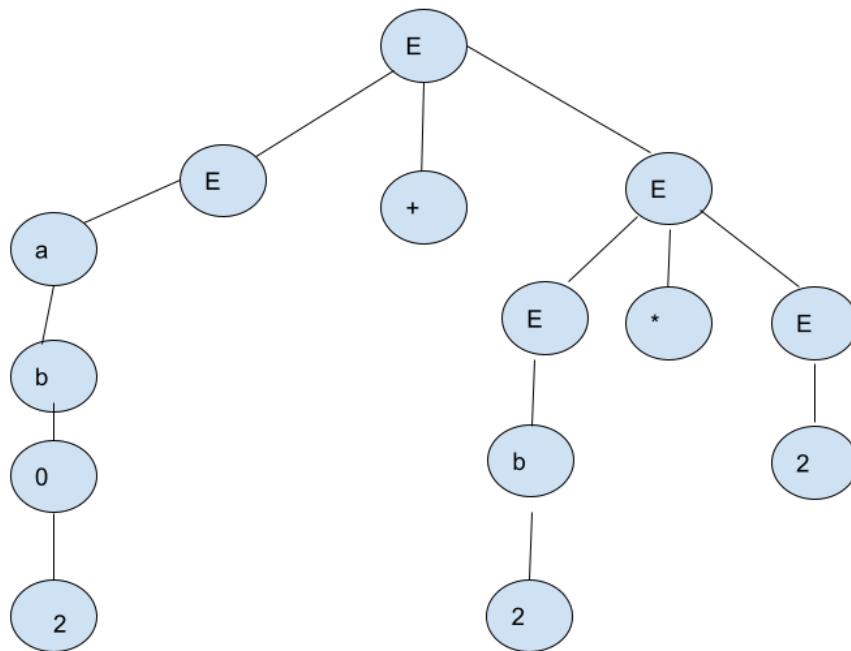
In plain English: Aim to get a machine to produce usable, natural language output.

NLP = Natural Language Processing = Natural Language Understanding + Natural Language Generation.

**Q2. You are given the following grammar for expressions:**

$E \rightarrow I$	$I \rightarrow a$
$E \rightarrow E + E$	$I \rightarrow b$
$E \rightarrow E * E$	$I \rightarrow 0$
$E \rightarrow (E)$	$I \rightarrow 2$

**a. [10 pts] Show parse tree(s) for the expression  $2 + 2 * 2$**



**b. [10 pts] Describe any interesting observations in your answer to a.**

This expression is interesting because it reminds me more of a computer language. The interesting thing about computer languages is that they are generally context free. The symbols are tokens without any semantic meaning, so all numbers are the same and can sometimes be regarded equally. Additionally, all variables are regarded equally, and this is nice because we will have a finite symbol set.

This is a good representation of how a compiler works at a very, very high level.

**Q3. Consider the following grammar and sentence:**

S → NPVP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP
PP → Preposition NP
Nominal --> Nominal PP

Section 8.1

**Sentence:** *I booked a flight from LA*

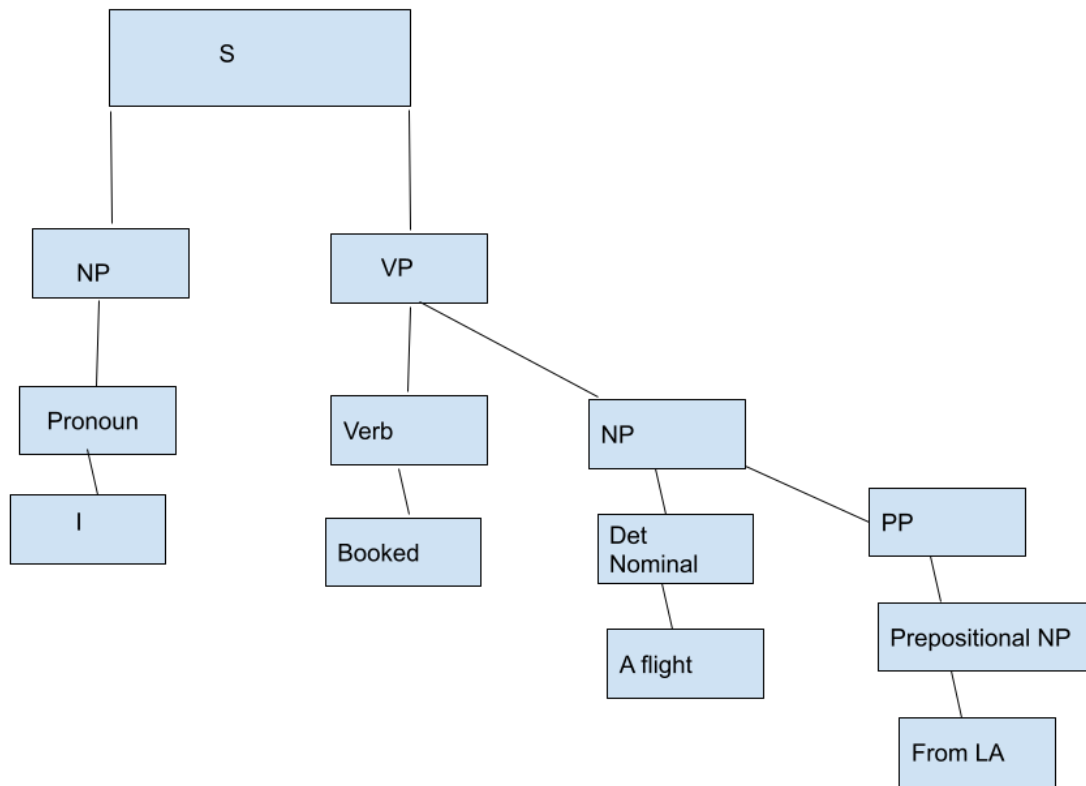
a. [10 pts] In what way is this sentence ambiguous? Describe different interpretations of this sentence.

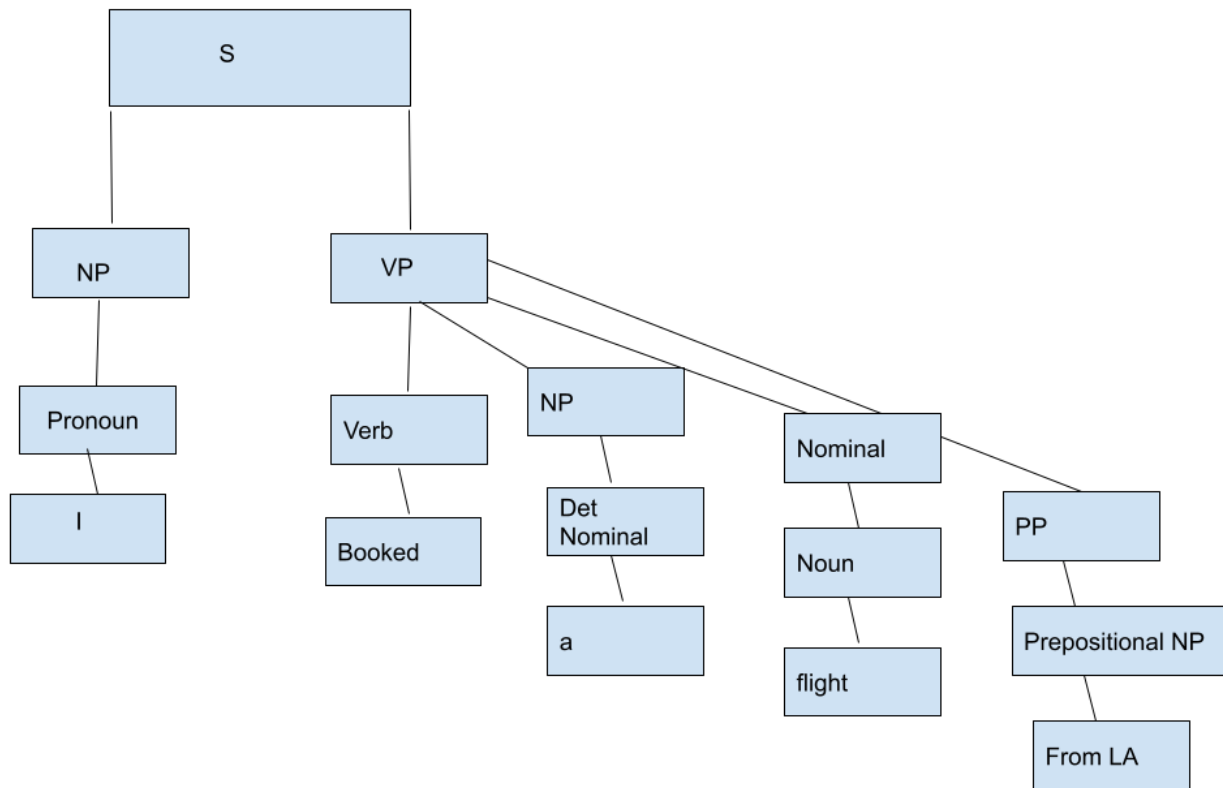
Answer: The base interpretation (using some contextual knowledge) would be to assume that the subject (I) of this sentence purchased a plane ticket from LA (The city in California).

However, this brings up many points as to its level of ambiguity.

“From LA” does indicate that the ticket was purchased in Los Angeles California? Could this also mean that the flight originates from LA? LA has the potential to stand for many things, could this instead refer to a flight from Louisiana? The state of Louisiana is abbreviated LA, or perhaps LA stands for something entirely different all together.

b. [10 pts] Show the parse trees for this sentence and where the ambiguity manifests in the parse trees.





The ambiguity really manifests itself within the verb phrase.

**Q4.** The image below shows Google search results for the query “harry potter”

en.wikipedia.org › wiki › Harry\_Potter\_(film\_series) ▼

### Harry Potter (film series) - Wikipedia

Harry Potter is a British-American film series based on the eponymous novels by author J. K. Rowling. The series is distributed by Warner Bros. and consists of ...

Release date: 2001–2011      Box office: Total (8 films); \$7.7 billion  
 Budget: Total (8 films); \$1.2 billion      Directed by: Chris Columbus (1–2); Alfonso C...

Harry Potter books · Harry Potter and the Prisoner ... · Harry Potter and the Half ...

---

www.warnerbros.com › movies › harry-potter-complete-8-film-collec... ▼

### Harry Potter 8-Film Collection | Movies - WarnerBros.com

The Magic Is All Here in the Complete 8-Film Collection. The collection includes: Harry Potter and the Sorcerer's Stone™ Harry Potter and the Chamber of ...

---

kids.scholastic.com › kids › books › harry-potter ▼

### Harry Potter | Scholastic Kids

A dazzling new edition of J.K. Rowling's Harry Potter and the Sorcerer's Stone, fully illustrated in brilliant color and featuring exclusive interactive paper craft ...

---

www.imdb.com › title ▼

### Harry Potter and the Sorcerer's Stone (2001) - IMDb

★ ★ ★ ★ ★ Rating: 7.6/10 - 595,398 votes

Daniel Radcliffe in Harry Potter and the Sorcerer's Stone (2001) Alan Rickman in Harry Potter and the Sorcerer's Stone (2001) Daniel Radcliffe in Harry Potter ...

---

https://twitter.com/HarryPotterFilm

### Harry Potter Film (@HarryPotterFilm) · Twitter

Filmmakers were able to bring Quidditch to life through the magic of VFX.

🌟 Love is in the air, along with some exciting news from the Wizarding World.

#TBT to one of the most stressful, but rewarding classes of Defence Against the Dark Arts.

As the results show, the query could represent any of the seven books in the harry potter franchise, any of the film adaptations of the books, a theme park, or a ride, an audiobook, cartoons, et al.

**a. [10 pts] Discuss why google shows a mix of such results and what factors can influence the search results for this query that will be presented to you.**

Primarily, Google shows a mix of results due to users' location, personalization, and subtle variations in the algorithm.

Location is important because many times a user is interested in queries that are relevant to their current location. For example, a user searching Google in Scotland may be interested in visiting some of the filming locations or areas that inspired various characters or events in the series. A user in the United States may be more interested in where to purchase one of the books or where to see one of the movies.

Personalization is important because google uses your prior history to learn and make assumptions based on websites you've clicked through. In future searches Google will prioritize these sites upon future visits. With the results above, the user may have visited the Wikipedia page concerning "Harry Potter" or viewed the Warner Brother's website to learn more about the movies.

Variations in the algorithm: Google is constantly changing its algorithm to optimize delivery of information to users. Google has recently implemented a new tool titled “BERT” which attempts to deduce the meaning of each individual word in a body of text which is based on the concept of “transformers,” models that process words in relation to all the other words in a sentence.

Alternatively, if no user signed in, Google is using the query “Harry Potter” and tries to disambiguate at the search bar level, however it looks like this query is just “Harry Potter.” Some of the results below are Google’s attempt at delivering the top webpages that other users visit when searching “Harry Potter.”

**b. [15 pts] Consider the following sentence:**

*The **bank** can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.*

The word **bank** has multiple senses. Use Wordnet to show the top two sense, glossaries and examples for **bank** and describe (at a high level) how you can use this information to find the proper sense for this word in a sentence.

Wordnet link: <http://wordnetweb.princeton.edu/perl/webwn>

Top two senses:

Bank^1: Bank (sloping land (especially the slope beside a body of water))

Bank^1: Depository financial institution, bank, banking concern, banking company. (A financial institution that accepts deposits and channels the money into lending activities)

Glossaries:

Bank^1: Direct hyponym/ full hyponym

Bank^1: Direct hypernym/ inherited hypernym / sister term

Examples:

Bank^1: “They pulled the canoe up on the bank”, “He sat on the bank of the river and watched the currents”

Bank^2: “He cashed a check at the bank”, “that bank holds the mortgage on my home”

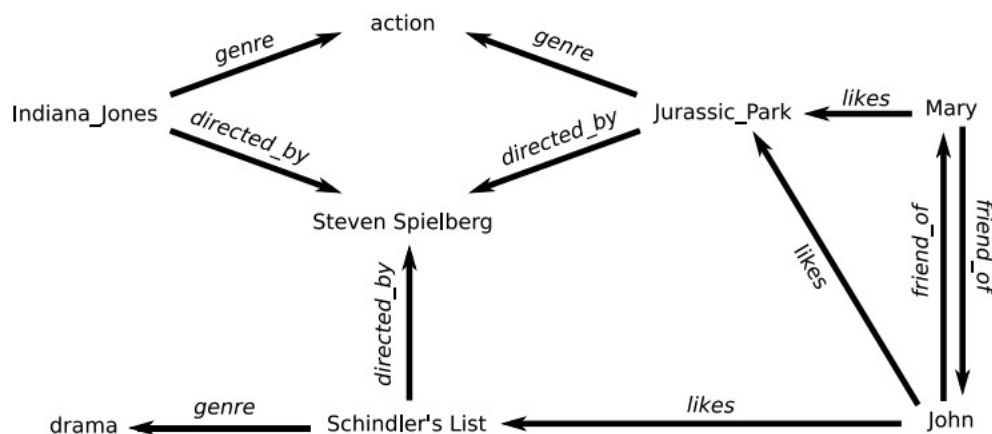
Senses are discrete representations of one aspect of the meaning of a word. Knowing the relation between two senses can play an important role in tasks involving meaning. Using the example of bank, we can understand how an embedding model like word2vec would be confused by a word like bank. Word sense disambiguation is also important here to determine



which sense of a word is being used a particular context. We can use all this information to generalize which senses are among the top and then based on the type of the word (noun, adjective, verb) we can get a good picture of the sense in this specific context.

We can use the other words in the sentence to match to the senses, the example sentence has words like deposit, which is included in the second sense example. The rest of the context of the sentence also contains words that are heavily related to the sentence in the question prompt.

**Q5. You are building an online moving streaming service which enables looking up information on movies, genres, directors, actors and customer movie preferences.**



\*\*(Indiana Jones refers to Raiders of the Lost Ark (1981))

**a. [10 pts] What is the customers intent (i.e. what are they looking for) with the following queries? (these are individual queries, not queries entered in succession)**

*“Drama”, “Jurassic Park”, “Indiana Jones: Raiders of the lost ark”, “Steven Spielberg”*

**Drama:** In this query the customer is interested in searching for a genre rather than a specific movie, in the case of this service a query for drama would return the result of Schindler’s List. They could also be interested in a drama movie that John liked. They could also be interested in a movie that Steven Spielberg directed that falls into the drama category

**Jurassic Park:** In this query the customer is interested in a specific movie and because the streaming service offers this movie, “Jurassic Park” can be directly returned. They could be interested in a movie that Mary likes. Additionally, they could be interested in an action movie directed by Steven Spielberg.

Indiana Jones: Raiders of the lost ark: In this query the customer is interested in a specific movie and because the streaming service offers this movie, "Indiana Jones: Raiders of the lost ark" can be directly returned. It is also possible that the customer is interested in action movies directed by Steven Spielberg.

Steven Spielberg: This is an interesting query, that could return a handful of results. Obviously we could return the Steven Spielberg movies available on this service, but we could also offer insight into the different genres he directs and direct the customer in that way.

**b. [5 pts] A customer searches for "Indiana Jones" but clicks on and watches "Jurassic Park" – what insights can you get from this customer action?**

If a customer searches for "Indiana Jones" but clicks and watches "Jurassic Park" we can get the following insights:

The customer was interested in a movie from the action genre and used Indiana Jones as a starting point

The customer knew that "Indiana Jones" was directed by Steven Spielberg and selected Indiana Jones with the knowledge that the service would recommend similar movies and a closely related title was "Jurassic Park."

It is possible the customer selected "Jurassic Park" after noticing that a friend watched the film in close succession to Indiana Jones.

**c. [10 pts] The customer searches for "Indiana Jones: Raiders of the lost Ark" but it's not available in their region (US, EU, Asia). What search results would you show the customer? Discuss how you would build that experience from a technical design perspective.**

From a design perspective we have a handful of artifacts to use here. In the first approach we could examine historical data and see what prior customers have watched in response to "Indiana Jones: Raiders of the lost Ark" not being available in their region. Perhaps there is a closely associated action movie available in their region.

In the second approach we could examine returning other action movies that are directed by Steven Spielberg such as "Jurassic Park."

In the third approach we could return other Steven Spielberg movies, and not just ones specific to the action genre.

In the final approach we could try a departure from director and genre, and instead we could examine movies that a user's friends were interested in.

I think from a design perspective, I'd go with either the first or second approach, these methods seem to offer a good level of trade-off between not being able to deliver the requested title and returning a suitable alternative. The worst thing for a streaming service is for a visitor to not watch a title on that service and go elsewhere. The best design would be to deeply study prior

customers behavior at this decision point and go with the most probable choice for a given customer and make the distinction if they searched for that title because they wanted an action movie or because they wanted to watch a movie directed by Spielberg.