1. (14 pts) For each sentence below, label the head noun of each noun phrase (NP) with the thematic role that is most appropriate based on its semantic relationship with the main verb.

(a) The boy walked with his dog to the park.

*The boy/AGENT walked with his dog/CO-AGENT to the park/DESTINATION.*

(b) Sylvia paid the plumber 100 dollars with a personal check.

*Sylvia/AGENT paid the plumber/RECIPIENT 100 dollars/THEME with a personal check/INSTRUMENT.*

(c) Jim studied hard for his mother.

*Jim/AGENT studied hard for his mother/BENEFICIARY.*

(d) The Picasso painting was donated to the Smithsonian museum.

*The Picasso painting/THEME was donated to the Smithsonian museum/RECIPIENT.*

(e) The Utah man did his next door neighbor a favor.

*The Utah man/AGENT did his next door neighbor/BENEFICIARY a favor/THEME.*
2. (12 pts) For each sentence below, name the conceptual dependency primitive ACT that best represents the meaning of the activity described in the sentence. (You do not need to draw a picture – just name the primitive ACT.)

(a) Theresa began to drool from a tooth infection.

   EXPEL

(b) Bobby smiled during the exam.

   MOVE

(c) Sam felt the snow on his nose.

   ATTEND

(d) Kathy solved the challenging riddle.

   MBUILD

(e) William shared his secret with Amy.

   MTRANS

(f) Donald Trump gave his daughter a mansion for her birthday.

   ATRANS

(g) Keith took cough medicine in the morning.

   INGEST

(h) Tom smelled the roses in his garden.

   ATTEND

(i) The spider crawled up the wall.

   PTRANS
(j) George snored all night long.

SPEAK

(k) Martin received an insulin injection for his diabetes.

INGEST

(l) Susan sent Happy Birthday wishes to Sandeep.

MTRANS
3. (8 pts) Suppose that a visitor in your home asks the question:

   *Do you know where the bathroom is?*

   (a) What is the *locutionary speech act* for this question?

   *The act of uttering the question: “Do you know where the bathroom is?”*

   (b) What is the *illocutionary speech act* for this question?

   *A request for you to tell the visitor where your bathroom is.*

   (c) What is the *perlocutionary speech act* for this question?

   *That the visitor successfully finds and uses the bathroom in your house.*

   (d) Does this statement represent a *direct* or *indirect* speech act?

   *Indirect, because the literal interpretation of the question is not the intent (i.e., the visitor is not asking whether you know where your own bathroom is, but asking you for directions to the bathroom.)*
4. (16 pts) Answer each question below as TRUE or FALSE. No explanation is necessary.

(a) Relevance feedback techniques use N-gram language models to improve information retrieval performance.

FALSE

(b) “Why” questions are particularly difficult for question answering systems because question typing does not help much.

TRUE

(c) The Hobbs algorithm uses statistics to do pronoun resolution.

FALSE

(d) A named entity recognizer could be useful for recognizing compatibility with selectional restrictions in case frames.

TRUE

(e) For bootstrapping algorithms, it is important to choose seed words or seed rules that will be frequent and match many instances in the unlabeled corpus.

TRUE

(f) Yarowsky’s word sense disambiguation algorithm with Roget’s thesaurus is an example of a method that exploits distributional similarity.

TRUE

(g) The Basilisk semantic lexicon induction algorithm identifies one extraction pattern that occurs with many seed words for a semantic category and then assumes that all of that pattern’s extractions belong to the same semantic category.

FALSE

(h) A script can be useful for pronoun resolution.

TRUE
5. (8 pts) Consider the following short story, which consists of 9 sentences, S1 to S9:

S1: Jack and Jill bought a house in Alaska.
S2: Jack wanted to fly to Alaska and Jill wanted to as well, so they bought plane
tickets to Juneau.
S3: Jill hired a moving company, United Van Lines, to move their furniture to Alaska.
S4: But United Van Lines complained when they discovered that Jack and Jill owned
a grand piano.
S5: So Jack sold the piano to a musician, who was happy to have it.
S6: Jack felt that it was important for the piano to be played.
S7: On the trip, the moving van caught fire and their furniture was destroyed.
S8: Mary was upset with herself for choosing United Van Lines.
S9: She vowed that she would never use that company again!

Identify one example of each phenomenon below in this story, or answer NONE if no
descriptions of the phenomenon exist in the story. IMPORTANT: Please give both the
sentence number and the corresponding word or phrase in your answer. For example,
if you were asked to find cities in Alaska, the answer would be “S2: Juneau”

- appositive
  
  S3: “moving company, United Van Lines”

- metonymy
  
  S4: “United Van Lines complained”

- verb phrase ellipsis
  
  S2: “wanted to [fly] as well”

- relative pronoun
  
  S5: “who”

- reflexive pronoun
  
  S8: “herself”

- possessive pronoun
  
  S3 & S7: “their”
• gendered pronoun

S8: “herself”; S9: “she”

• pleonastic pronoun

S6: “it”
6. (12 pts) Assume that an imaginary thesaurus contains the following word lists for the categories MAMMALS and SPORTS.

MAMMALS: bears, deer, coyotes
SPORTS: ball, game, glove

The box below shows seven sentences, which you should use as the text corpus for this question. This text corpus contains exactly 80 words.

S1: The population of bears and wolves has decreased this season.
S2: Most species of bears hibernate in dens all winter.
S3: Game hunting is popular where deer and wolves are plentiful in the fall season.
S4: Kirby Puckett won the golden glove award last season.
S5: The Chicago Bears cheered when the ball was caught by a fan.
S6: The ball landed in center field and Puckett scored a run.
S7: The Arizona Coyotes scored a quick goal to win the last game of the season.

Compute the following salience values as defined by Yarowsky’s word sense disambiguation algorithm. You should assume that the context window for a word spans the entire sentence containing the word (but do not cross sentence boundaries). Please show all your work! You may leave your answers in fractional form.

(a) salience(species, MAMMAL)

\[
P(species \mid MAMMAL) = \frac{1}{5} \\
P(species) = \frac{1}{80} \\
\text{Salience} = \frac{\frac{1}{80}}{\frac{1}{5}} = 16
\]

(b) salience(Puckett, MAMMAL)

\[
P(Puckett \mid MAMMAL) = \frac{0}{5} \\
P(Puckett) = \frac{2}{80} \\
\text{Salience} = \frac{0}{\frac{2}{80}} = 0
\]

(c) salience(season, MAMMAL)

\[
P(season \mid MAMMAL) = \frac{3}{5} \\
P(season) = \frac{1}{80} \\
\text{Salience} = \frac{\frac{3}{80}}{\frac{1}{80}} = 12
\]
(d) salience(species, SPORTS)

\[ P(species \mid SPORTS) = \frac{0}{5} \]
\[ P(species) = \frac{1}{80} \]
\[ \text{Salience} = \frac{0}{\frac{1}{80}} = 0 \]

(e) salience(Puckett, SPORTS)

\[ P(Puckett \mid SPORTS) = \frac{2}{5} \]
\[ P(Puckett) = \frac{2}{80} \]
\[ \text{Salience} = \frac{\frac{2}{5}}{\frac{2}{80}} = 16 \]

(f) salience(season, SPORTS)

\[ P(season \mid SPORTS) = \frac{3}{5} \]
\[ P(season) = \frac{4}{80} \]
\[ \text{Salience} = \frac{\frac{3}{5}}{\frac{4}{80}} = 12 \]
7. (10 pts) For each (short) story below, state whether a script or plan would be the appropriate knowledge structure to understand how the sentences in the story relate to one another. You do not need to give an explanation.

(a) “Tom went to the doctor for his annual check-up. He checked in with the receptionist and waited for 10 minutes until the doctor was ready to see him. The doctor measured his blood pressure and weight, took a blood sample, and gave him a physical exam. Tom was happy that the doctor found no problems.”

SCRIPT

(b) “Carol could not find her car keys and was late for work. She remembered that there was a bus stop a few blocks away so she took the bus.”

PLAN

(c) “Jim wanted to learn how to ski so he asked his father to take him to Solitude and teach him.”

PLAN

(d) “Martha purchased a new cell phone on Monday. Just 24 hours later, she dropped the phone down an entire flight of stairs and it hit the ground hard, cracking the glass. She remembered that she had purchased a replacement warranty for the phone, though, so took it back to the shop to be replaced.”

PLAN

(e) “10-year-old Tony walked to school and took a spelling test in his 1st period English class. He was starving by the time it was lunch period, so he ate his sandwich and bought some french fries. He fell asleep in his 7th period class, but was awakened by the bell and happy to finally go home.”

SCRIPT
8. (14 pts) Consider the following query, documents, and vocabulary:

QUERY: natural language

DOCUMENT #1: natural language is natural and is fun
DOCUMENT #2: natural language processing is fun
DOCUMENT #3: utah is fun and natural
DOCUMENT #4: the fun is natural language

VOCABULARY: and fun is language natural processing the utah

Suppose you want to create feature vectors for the query and for document #1, for an information retrieval system. Fill in the tables below with the values that would be produced using three different types of feature values (weights): binary, term frequency, or TF-IDF. For logarithms, please leave them as logarithms, (e.g., log(1)) and do not reduce them!

**Feature Vector for the Query**

<table>
<thead>
<tr>
<th>and</th>
<th>fun</th>
<th>is</th>
<th>language</th>
<th>natural</th>
<th>processing</th>
<th>the</th>
<th>utah</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Term Freq</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TF-IDF</td>
<td>0*log(4/2)</td>
<td>0*log(4/4)</td>
<td>0*log(4/4)</td>
<td>1*log(4/3)</td>
<td>1*log(4/4)</td>
<td>0*log(4/1)</td>
<td>0*log(4/1)</td>
</tr>
</tbody>
</table>

**Feature Vector for Document #1**

<table>
<thead>
<tr>
<th>and</th>
<th>fun</th>
<th>is</th>
<th>language</th>
<th>natural</th>
<th>processing</th>
<th>the</th>
<th>utah</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Term Freq</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TF-IDF</td>
<td>1*log(4/2)</td>
<td>1*log(4/4)</td>
<td>2*log(4/4)</td>
<td>1*log(4/3)</td>
<td>2*log(4/4)</td>
<td>0*log(4/1)</td>
<td>0*log(4/1)</td>
</tr>
</tbody>
</table>
9. (6 pts) Suppose a word sense disambiguation (WSD) system is trying to distinguish between two senses of the word “bug”: $sense_A$ is the meaning of bug as an insect, and $sense_B$ is the meaning of bug as a software error. Assume that a WSD system has disambiguated some instances of bug in the two documents below, labeling them as either sense A or sense B. (Note that these labels may not be correct!)

DOC #1
S1: The computer bug/B is probably in the print function.
S2: The bug is very subtle and may be hard to fix.
S3: We will have the bug/B fixed by the end of the day.
S4: I hope this bug didn’t cause too many problems!

DOC #2
S5: The bug/A flew around the room and landed on a window sill.
S6: John hit the bug/A with a fly swatter.
S7: But the bug survived!
S8: The bug/B is harmless, but John was intent on killing it.
S9: But the bug/A was too fast and eventually flew out the window.
S10: The bug escaped!

(a) Can Yarowsky’s “one sense per discourse” heuristic be used to disambiguate any unlabeled instances of bug in Doc #1? If so, for each instance of “bug” that can be disambiguated by this heuristic, give its sentence number and its new sense label (e.g. “S6: A”) If not, answer NONE.

Yes – S2: B and S4: B.

(b) Can Yarowsky’s “one sense per discourse” heuristic be used to disambiguate any unlabeled instances of “bug” in Doc #2? If so, for each instance of “bug” that can be disambiguated by this heuristic, give its sentence number and its sense label (e.g. “S6: A”) If not, answer NONE.

Yes – S7: A, S10: A
(c) Can Yarowsky’s “one sense per discourse” heuristic be used to change any of the original sense labels in Doc #1? If so, for each instance of “bug” whose label would be changed by this heuristic, give its sentence number and its new sense label (e.g. “S6: A”) If not, answer NONE.

No – none.

(d) Can Yarowsky’s “one sense per discourse” heuristic be used to change any of the original sense labels in Doc #2? If so, for each instance of “bug” whose label would be changed by this heuristic, give its sentence number and its new sense label (e.g. “S6: A”) If not, answer NONE.

Yes – S8: A
10. (10 pts) Consider the following two (short) stories:

(1) A twister hit Kansas on Monday and it caused massive damage. Three people were killed and 20 people were injured. A tree later fell on three of the injured men. A dog was also hit by flying debris and injured. A historic farm took a direct hit from the twister. The tornado occurred at midnight.

(2) A group of children played twister on Monday. A bizarre accident occurred and 1 child was injured and another child was killed while they were playing the game. A large chandelier hit them and killed a boy. An ambulance took the injured kids to a hospital. The children often played in that room. The incident occurred at noon. 10 people have been killed while playing twister this year.

Assume that Story 1 is a relevant text, and Story 2 is an irrelevant text. For each of the information extraction patterns \( \langle p_i \rangle \) below, compute \( P(\text{relevant} \mid p_i) \). PassiveVP(verb) means the verb appears in a passive voice verb phrase construction. ActiveVP(verb) means that the verb appears in an active voice verb phrase construction. Leave your answers in fractional form!

(a) \( \langle \text{subject} \rangle \) PassiveVP(killed)

\[
\frac{1}{3}
\]

(b) \( \langle \text{subject} \rangle \) ActiveVP(hit)

\[
\frac{1}{2}
\]

(c) \( \langle \text{subject} \rangle \) PassiveVP(injured)

\[
\frac{2}{3}
\]

(d) ActiveVP(played) in \( \langle \text{np} \rangle \)

\[
0/1
\]

(e) ActiveVP(took) \( \langle \text{direct-object} \rangle \)

\[
\frac{1}{2}
\]