P&P 3

Compositional semantics and discourse

Compositional semantics is concerned with modeling the meaning of phrases and sentences. It relies on the principle of compositionality, which states that the meaning of a phrase or a sentence can be derived from the meanings of its parts, guided by syntactic structure.

In the following examples, identify and explain the issue compositional semantics models might face in each of the examples. Additionally, explain how these issues might complicate compositional semantics.

1a) *it snows*

The word *it* in *it snows* is a pleonastic pronoun that does not refer to a specific entity, unlike the word *it* in a sentence like *it barks* which may refer to a specific referent, like a dog. Compositional semantics assumes the meaning of a sentence can be derived from the meaning of its parts, however, the pronoun *it* in this case does not contribute any referential meaning in this case.

1b) rock the boat

The phrase *rock the boat* is an idiom, and its meaning of "causing disruption to the status quo" cannot be derived from the meanings of the individual words (rock, the, boat) with compositional semantics. Idioms violate the principle of compositionally because their meanings are not a function of their parts or syntactic structure.

1c) enjoy a cigarette/ enjoy a book

These phrases involve logical metonymy, where the verb *enjoy* implies an associated activity that many be different depending on the objects involved. In this example, *enjoy a cigarette* means smoking a cigarette, and *enjoy a book* means to read a book. The activity associated is inferred based on context and world knowledge, rather than the meaning of the word *enjoy*. This poses a challenge to compositional semantics as the meaning of the phrase is not fully contained within the meanings of the phrase's parts.

1d) inject life/ inject hydrogen

Similar to the previous example, the example phrases demonstrate logical metonymy and polysemy of the word *inject*. *Inject hydrogen* may mean the physical action of literally injecting hydrogen into something, while the phrase *inject life* metaphorically means to bring energy or vitality to something. This is an complication for compositional semantics as the literal and metaphorical meanings of *inject* may not be distinguishable from the composition of meanings of the lexical items in the phrase.

Distributional representations of phrases and sentences can be derived by combining, or *composing*, the distributional representations of individual words in them.

1e) Describe two limitations of vector mixture models.

1. Vector mixture models do not take word order or syntax in to account: the additive and multiplicative models means the model symmetric, thus "John hit the ball" and "the ball hit John" receives the same representation.
2. By collapsing multiple vectors into one, vector mixture models fail to retain information about how words interact in the phrase or sentence, and function words like conjunctions or prepositions (*some* dogs, lice *on* dogs) are lost in the representations.

1f) *Briefly* describe how do lexical function models address these limitations.

Lexical function models address these limitations by incorporating word specific transformation functions, in the form of lexical functions, that capture the contextual and relational information between words. By using lexical functions of specific words like verbs or modifiers to transform the representation of their arguments like objects and subjects, the model can capture word order and relational dependencies.

1g) Describe how adjectives can be represented as lexical functions and how such lexical functions can be learned from a text corpus.

Adjectives like "old" or "big" are modeled as lexical functions that can be represented with parameter matrices, A_{old} and A_{big}. The lexical functions $f_{adj}(\vec{dog})$ can then be expressed as linear transformations of the vector representation of the noun $\vec{dog}$:

$$\text{big dog} = A_{big} \cdot \vec{dog}$$
$$\text{old dog} = A_{old} \cdot \vec{dog}$$

The adjective matrices can be learned from a text corpus with the following steps:

1. **Build word representations**: construct distributional vectors using methods like Word2Vec , GloVe for all words in the vocabulary, including vectors for each noun $\vec{n_j}$ and adjective-noun pair $\vec{p_{ij}}$, for each adjective noun pairs (a_i, n_j).
2. **Extract training data**: identify the set of tuples $\{(n_j, p_{ij})\}_j$ that represents the dataset $D(a_i)$ for the adjective a_i.
3. **Optimize matrices**: learn matrix A_i from $D(a_i)$ using linear regression, where we minimize the squared error loss, $L(A_i) = \sum_{j \in D(a_i)} ||p_{ij} - A_i n_j||^2$

Suggest a way in which the above lexical function model of adjective meaning can be used to:

1h) Paraphrase adjective-noun phrases, obtaining paraphrase pairs such as smart students → clever students

This can be achieved with the lexical function model assuming synonyms like "smart" and "clever" will produce similar transformations on the noun vector "student", then the paraphrasing can be obtained by

identifying adjective-noun representations that are close to one another in the vector space. For example, the cosine similarity between "smart student" and all other "adjective-student" vectors in the vocabulary can be computed to identify pairs of high similarity.

1i) Disambiguate the meaning of the adjective, for instance, the meaning of *warm* in warm tea vs. warm person.

The lexical function f_{warm} can be applied to the noun vectors of tea and person to encode context specific meanings of the adjective "warm". Then the resulting vector representations for the adjective-noun pairs can then be compared with vectors of potential meanings of the phrases. For instance, computing similarities of representations between the adjective-noun pairs with the representations of words like "temperature" and "affection" would then reveal that the phrase "warm tea" is more associated with the temperature sense of the adjective "warm" and "warm person" is more associated with the "affection" sense of the adjective "warm". This method allows lexical function models to disambiguate the meanings of adjective depending on the noun the adjective modifies.

The two sentences in each sentence pair below are linked by a particular rhetorical relation. Which rhetorical relation does each sentence pair exhibit? Focus on the relation of the second sentence with respect to the first. Answer the following questions (j)-(l) by naming the correct relation and giving a 1 sentence explanation. No marks will be awarded for incorrectly identified relations without an explanation. For incorrectly identified relations with *plausible* explanations, partial marks might be awarded.

1j) *The use of diesel in transport has come under increasing scrutiny in recent years. According to WHO, around three million deaths every year are linked to exposure to outdoor air pollution.*

The second sentence provides a statistical support from WHO that explains why diesel use is under increasing scrutiny, serving as an **explanation** for why the claim was made in the first sentence.

1k) *Nitrogen oxides can help form ground-level ozone. This can exacerbate breathing difficulties.*

The second sentence describes the **result** (**cause and effect**) of nitrogen oxides forming ground-level ozone, showing that it leads to exacerbated breathing difficulties.

1l) *Paris has already taken a series of steps to cut the impact of diesel cars and trucks. Vehicles registered before 1997 have already been banned from entering the city.*

The second sentence provides a specific example of the steps that Paris took, which forms a **justification** for the statement that Paris has already taken a series of steps.

Specify which salient rule is used to resolve the highlighted anaphora in the given discourse. Name the relevant rule and give a short 1-2 sentence explanation of how it applies to the anaphora.

1m) *Lee bought his first car in 2005 and his second car, last year. He now drives it to work.*

The pronoun "it" refers to "his second car", which is the more recent suitable antecedent in the discourse. This suggests the **recency** rule can be applied to resolve the anaphora.

1n) *Lara scolded Maria for breaking the glass; she still couldn't contain her anger.*

The pronoun "she" refers to "Lara", which is the subject of the previous sentence. The **grammatical role** rule states that pronouns prefer antecedents that are subjects rather than objects, and thus "she" resolves to the subject of the previous sentence, "Lara".

1o) *Elizabeth first danced with Mr Darcy. Anna danced with him next.*

The **parallelism** rule states that entities which share the same role as the pronoun in the same sort of sentence are preferred. In this case, "him" refers to "Mr Darcy" in the previous sentence.

Consider the following discourse:
Arya went to Hilda's car showroom to check out a Fiat Linea.
She decided to buy it, after hours of inspection.

1p) Create the feature vector for each possible (pronoun, antecedent) pair for anaphora resolution in the discourse with *Cataphoric, Number agreement, Gender agreement*, Same verb, Sentence distance, *Grammatical role*, *Parallel*, and *Linguistic form* features as described in the lecture. Focus on the following pairs: (she, Arya); (she, Hilda); (it, car showroom); (it, Fiat Linea).

Cataphoric: if pronoun appears before antecedent
Number agreement: if pronoun compatible with antecedent
Gender agreement: if gender agreement
Same verb: if the pronoun and the candidate antecedent are arguments of the same verb
Sentence distance: discrete of {0, 1, 2, ...}
Grammatical role: the role of the potential antecedent, discrete of {subject, object, other}
Parallel: if the potential antecedent and the pronoun share the same grammatical role.
Linguistic form: {proper, definite, indefinite, pronoun} of antecedent

---

(she, Arya)

```
[False, True, True, False, 1, subject, True, proper]
```

(she, Hilda)

```
[False, True, True, False, 1, object, False, proper]
```

(it, car showroom)

```
[False, True, True, False, 1, other, False, definite]
```

(it, Fiat Linea)

```
[False, True, True, False, 1, object, True, indefinite]
```

1q) Describe how this feature vector can be used to resolve the anaphora assuming a supervised approach. Make sure you describe how you collect data for your approach and explain clearly how to implement your approach.

The supervised approach can be implemented with:

1. For each pronoun in a large corpus, extract pronoun-antecedent pairs by identifying all potential antecedents within a certain window of text.
2. Label each pronoun-antecedent pair as True for correct antecedent and False for incorrect ones.
3. For each labeled pair, extract feature vector.
4. Use labeled feature vector to train a probabilistic classifier. The model should learn to estimate the probability that a given pronoun-antecedent pair is correct based on the input features.
5. At test time, the pronoun-antecedent pairs of an unseen text can be extracted, and converted into feature vectors.
6. The trained model can then predict the probability that each candidate pair is correct. The candidate pair with the highest predicted probability can be selected as the antecedent of the pronoun.

1r) Mention two problems with this simple classification approach and briefly (1-2 sentences per problem) explain why they arise.

1. This simple classification approach based on the feature vectors may overlook the most salient entities that are repeatedly referred to in the discourse, as it evaluates each pronoun-antecedent pair independently without considering the frequency of mentions. This means the model cannot

leverage the "repeated mention" effect where entities mentioned multiple times are more likely to be the antecedents of pronouns.

2. Earlier anaphoric links in the text may often provide crucial context for resolving subsequent pronouns. Since the model treats each pair in isolation, it is unable to capture continuity throughout the discourse, which leads to potential inconsistencies in anaphora resolution.