

CS 421 – Natural Language Processing – Spring 2015

Term Project (Part 2)

1 General Information

This is the second part of the project. We will deal with syntactic well-formedness in general, and with rudimentary notions of coherence. Due dates are repeated here for your convenience (meant as 11:59pm):

	Due	Points
Competition	4/27 (Mon)	10
Part 2	5/5 (Tue)	140

2 Syntactic well-formedness

You will use one of the three parsers (Stanford, OpenNLP, NLTK) to complete the evaluation of the grammatical well-formedness of the essay. In particular, we have not dealt with evaluation criterion 1d, *sentence formation* (please see p. 2 in the handout for the first part of the project). A criterion for complete sentences could include the following subcriteria:

1. are main sentences formed properly? i.e. they should begin and end properly; the constituents should be formed properly: are there missing words or constituents (prepositions, subject, object etc.)?
2. If subordinating conjunctions are used (e.g., *when*, *although*, *if*), is there a main verb, or a gerund, to go with them? For many subordinating conjunctions, the corresponding clause can be finite, i.e., it has a main verb: *when I travel alone*, ...); or it's not finite, and includes a gerund: *when traveling alone*, But e.g. *because* cannot. You can say *Because I travel alone*, ..., but not **Because traveling alone*, ... – you could say *Because of traveling alone*, ...

You can use POS tagging to evaluate some features contributing to (1d), for example correct word order, e.g. in English no main verb should be in first position in a sentence; or no subordinating conjunction should be in a sentence with only one main verb, like in *Because I think the science and technology are developping* (this is the whole sentence).

In general however, you are asked to use the results the parser returns to judge criterion (1d). Note that (1d) applies to the whole essay, like the other criteria; so it's not that a single wrong sentence should drive the score down a lot, if the other sentences are correct.

All these parsers will return at least some parse trees for most (?all) sentences. For example, consider “My dog with a broken leg I not want” (an S but not a correct one).

The Stanford parser returns:

```
(ROOT
  (FRAG
    (NP (PRP$ My) (NN dog))
    (PP (IN with)
      (NP (NP (DT a) (JJ broken) (NN leg))
        (SBAR
          (S
            (NP (PRP I))
            (RB not)
            (VP (VBP want))))))))))
```

Here “FRAG” is an indication that the sentence is not complete; additionally, SBAR should be a daughter of another S or a VP, as in “I told them that I would come”. Compare the result above with the tree returned for “I do not want my dog with a broken leg” – no FRAG and no embedded SBAR:

```
(ROOT
  (S
    (NP (PRP I))
    (VP (VBP do) (RB not)
      (VP (VB want)
        (NP (PRP$ my) (NN dog))
        (PP (IN with)
          (NP (DT a) (JJ broken) (NN leg)))))))
```

Similar considerations apply to the OpenNLP parser. For “My dog with a broken leg I not want” it returns:

```
(TOP
  (NP (NP (NP (PRP$ My) (NN dog))
    (PP (IN with)
      (NP (DT a) (JJ broken) (NN leg)))))
  (SBAR
    (S
      (NP (PRP I))
      (RB not)))
  (. want)))
```

Compare it with its output for the grammatical version of this sentence, “I do not want my dog with a broken leg”:

```
(TOP
  (S
    (NP (PRP I))
    (VP (VBP do) (RB not) (VP (VB want)
      (NP (PRP$ my)
        (NN dog))
      (PP (IN with) (NP (DT a) (JJ broken) (NN leg)))))))
```

For the correct sentence, there’s an S as immediate daughter of TOP, no embedded SBAR, and *want* is assigned its correct POS tag.

Embedded SBAR’s are fine when eg the sentence contains a “because” clause, as in “I came because he was sick”. However, the SBAR is a daughter of a VP, and has an immediate daughter which is the conjunction in question (POS tag = IN). Stanford parser output:

```
(ROOT
  (S
    (NP (PRP I))
    (VP (VBD came)
      (SBAR (IN because)
        (S
          (NP (PRP he))
          (VP (VBD was)
            (ADJP (JJ sick))))))))
```

OpenNLP:

```
(TOP
  (S
    (NP (PRP I))
    (VP (VBD came)
      (SBAR (IN because)
        (S
          (NP (PRP he))
          (VP (VBD was)
            (ADJP (JJ sick)))))))
```

One approach may be to write some patterns that correspond to common mistakes, as noted above (e.g. Stanford: there’s FRAG; OpenNLP: there’s TOP without an S as immediate daughter), and check whether the output of the parser includes those patterns. In theory, another approach would be to retrain these parsers with the new data, but the corpus of essays is too small; besides, the essays would have to be annotated with syntactic trees, at least for the Stanford parser.

These are some *possible* ideas – meaning, once you think about it, you may come up with other ways of using the parsers. **Be creative!** In addition, note that the work you did in part 1 to

evaluate criteria 1b and 1c can also be used to seed the work done by the parser; and conversely, the results of parsing may be useful to better evaluate 1b and / or 1c. You are not asked to redo the work you did in part 1 for these three criteria, but to integrate the two approaches together if they provide complementary information.

3 Text coherence

The grading criteria we listed in part 1 include *2a/2b*, which pertain to semantics and pragmatics, and are repeated here:

2a Is the essay coherent? Does it make sense?

2b Does the essay address the topic?

2a corresponds to an open research problem in NLP. Part of coherence of text depends on referring expressions. In these essays, many referring expressions are definite NPs which are too difficult to deal with within a project like this. However, we can get some mileage out of assessing whether pronouns are used felicitously.

here are some observations on pronoun usage in these essays.

1. First person singular pronouns and possessive adjectives –*I, me, my, mine*– refer to the speaker / writer, are solved based on who the speaker is, and are not ambiguous. Same for first person plural pronouns, *we, our*, although they are harder to interpret since they refer to a “group” that includes the speaker. Second person pronouns (*you, your*) can be used as well, in an impersonal sense – as in the following example from the excerpt of the “high” essay included in the first part of project: ... *going to the places you choose to go to and discovering everything on your own.*
2. Third person singular pronouns are hardly used in these essays. Doublecheck if they do. If you find a *he* or *she* you can quickly assess whether it is used properly: any third person pronoun should have a possible antecedent. If *she* is used and no feminine entity has been introduced, then *she* is wrong (see below a note on where to find the information about gender and number); likewise for *he* and male antecedents.
3. Third person plural pronouns (*they*) instead are often used. For these,
 - (a) First, you should check if there are potential correct antecedents: either plural nouns, or nouns with compatible number (see sec 21.6.4 in the book), but used properly. I.e, *someone, group, family* can be used as antecedents for *they/them*, but it often doesn’t sound felicitous when the antecedent is in a prepositional phrase:
 - *A group travelling together can be fun. You will get to know them* is felicitous
 - *I don’t agree that the best way to travel is in a group. They will have many problems* is not as felicitous
 - (b) Second, the antecedent to *they/them* should not be too far: so, a pronoun should have an appropriate referent in the previous one-two sentences, which could be another pronoun referring to the same entity. This is called a *chain*.

- (c) Finally, if there is more than one possible antecedent, one of them should be more prominent than the other. In our simplified scenario, *more prominent* means it has been mentioned more recently; the further apart the various possible antecedents are, the better the referent is.

3.1 One possible algorithm, more in detail

Your algorithm then should

- Collect all pronouns and possessive adjectives
- Eliminate all pronouns/adjectives that are not third person
- For singular third person pronouns and possessives *he/she/his/her*, quickly check the existence of appropriate male/female antecedents as mentioned earlier.
- For plural third person pronouns and possessives *they/them/their*, check if there are possible antecedents:
 - no plural antecedent, or no singular antecedent with compatible number: mistake. If antecedent is singular, try to assess if it's used correctly (given its syntactic position)
 - only one possible antecedent: correct / felicitous
 - more than one possible antecedent: evaluate how ambiguous that pronoun is, based on how many antecedents there are, and how recent they are. The more ambiguous, the less felicitous.

For this last step, you can use the results that the coreference modules in the Stanford / OpenNLP / NLTK packages provide. However, what these packages return is one possible interpretation for pronouns (and possibly other NPs) – e.g. they don't tell you how many possible antecedents are there for a single pronoun. If you use the results of these coreference modules, you will have to augment them, as just discussed.

3.1.1 Notes

1. You can exploit POS tags for number.
2. Note that *they* sometimes does not have an explicit previous plural referent, but rather, refers to collections of items, as in
I have a son and a daughter. They play together a lot.
You should think of some appropriate heuristics to deal with these cases (if they arise).
3. As usual, we are greatly oversimplifying the problem. For example, it is not true that if there's no antecedent within the previous 1-2 sentences the pronoun is infelicitous, it really depends on the global structure of discourse.

4 Topic Coherence

For criterion 2b (Does the essay address the topic?), we want to understand whether the student has written about the stated topic. For example, suppose our topic was “will people have fewer kids in the future?”: you could see how many of the common nouns they use are related to family, but also to potential relevant topics, such as work, economics, etc.

To assess this, you could check the common nouns in the essay by exploiting Wikipedia, or electronic dictionaries, eg Wordnet (<http://wordnet.princeton.edu>) or ontologies, e.g. SUMO (<http://www.ontologyportal.org/>).

In Wordnet, words are organized in terms of hypernyms (is-a relation), meronym (part-of) and many other relations. You can exploit that information to find related words to the ones in the essay topic. The problem that you may encounter is that there are many senses for a word and one would need to disambiguate the correct sense. But for your project you can assume that we know the sense and just use the correct sense.

In SUMO, one subhierarchy is on “familyRelation”. For example, if you browse *aunt* in SUMO, you get

```
109823502 the sister of your father or mother; the wife of your uncle.
```

SUMO Mappings: familyRelation (subsuming mapping)

Once you have an estimate for how many common nouns refer to family, work, economics, and other potential relevant topics, you can assess the relevance of the essay in its entirety – for example, by computing the percentage of all the nouns that relate to the relevant subtopics with respect to all the nouns in the essay.

This is a fairly naive method. To start with it doesn’t take verbs into account. Second, in the best essays, the prompt is addressed from a variety of points of view, that are difficult to predict. You don’t want to penalize writers who are more creative.

If you have more creative ideas, feel free to experiment. Creativity will be rewarded when grading your work (as long as it makes sense).

5 Report and Submission

When submitting your work, please follow the guidelines already given for the first part of the project. Specifically, your submission will now contain the complete project and updated files.

You are also required to write a project report. In the report you have to comment on your project, what worked, what did not work, what you learned from the assignment, what you think would be necessary to take your software to a higher level of performance, etc. The report should be at least one page long – ie, not just a single paragraph.