

Natural Language Processing: Assignment 3

Shifei Chen

1 Dependency Annotation

Here are three typical errors I have made during my manual annotation.

1.1 `nmod` vs `compound`

"If a Turkish employee quits, then the Turkish work councils come."

The ambiguation between a single `nmod` and a `compound` noun is confusing to me. In the phrase "work councils" I have annotated "work" as a nominal modifier to "councils" while the golden standard says they should be compounded. A compound word usually consists of nouns (or verbs, adjectives, etc.) being writted closely to each other. In fact they are so close that in some Language like Swedish people usually write them together, without whitespaces. Meanwhile a nominal modifier is usually used to describe an attribute or genitive complement[1]. In this case, "work" can be argued to be the attribute of "councils" but since this is an idiom[2] so I prefer the golden standard.

1.2 `mark`

"National reaction to the events in Kansas demonstrated how deeply divided the country had become."

Here I have marked the word "how" as a `mark` to the head word "become" rather than an `advmod` to the head word "deeply". It is clear that in this sentence, "how deeply divided the country had become" is a complement clause, acting as the object to the verb "demonstrated". The word "become" is a `ccomp` to "demonstrated"[3]. The clause does clearly indicate that its object is "divided", not "how", hence we should consider "how" as an adverbial modifier to the other adverb "deeply", even though sometimes "how" can be the mark word of a complement clause[4].

1.3 Subjects in Passive Voice

"The RHS collected comments sent in by schoolchildren and teachers involved in the experiment."

I annotated "sent" to be the clause modifier of noun (`acl`) to "collected" and "schoolchildren" to be the nominal subject of "comments". The error in the former situation is pretty obvious since "collected" is a verb[5]. I should have annotated "sent" to be the `acl` of "comments". The latter is also wrong. Judging by the meaning, it may be intriguing to annotate "schoolchildren" as a `nsubj` because those children are the one who sent comments. But there is an example from Universal Dependency which explicly says "(`obl` is used) for the agent of a passive verb (with the optional subtype `obl:agent`):"[6]. We should mark "schoolchildren" as the `obl` to the verb "sent".

2 Dependency Parsing

Below are two examples of errors made by the parser in Lab 9.

A witness told police that the victim had attacked the suspect in April.

In this sentence my parser annotated "police" as the `obj` of the verb "told", which should be an `iobj`. "Tell" can have a direct object like in the verb phrase "tell a lie" but here "police" is not the subject. In fact, the clause beginning with "that" is the real subject. This might be caused by our corpus containing too many "tell"s having direct objects, or by the fact the our parser, as a transition based parser, always has a limited scope of

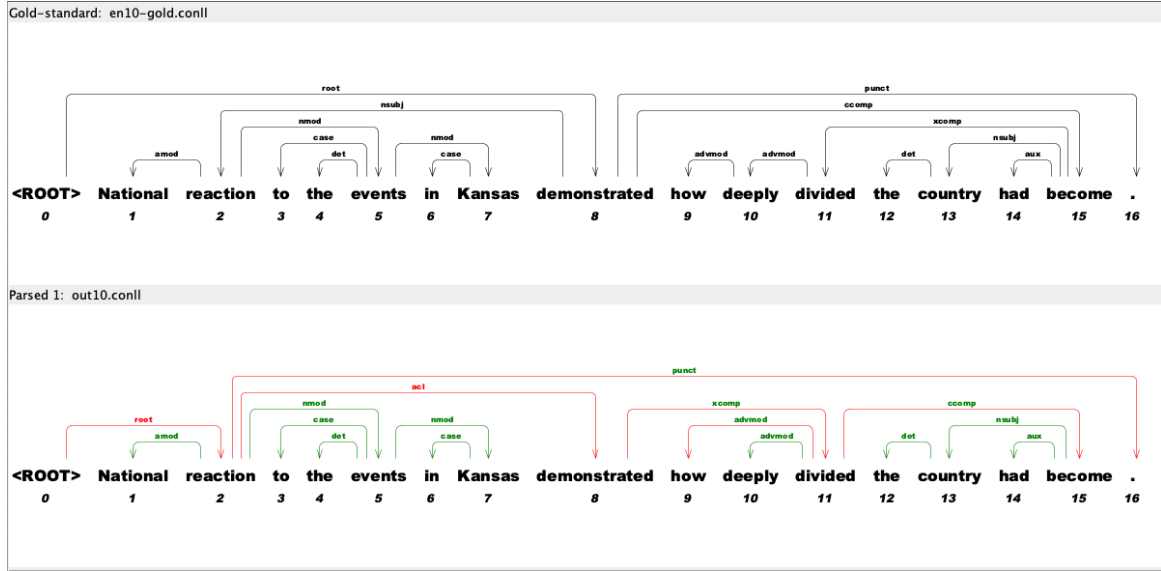


Figure 1: The Problem of Limited Scope

words—it can’t foresee the words behind ”police” therefore it prefers ”tell” to have a direct object. Nevertheless UD has a specific label `iobj` for this kind of indirect subject[7] and we should stick to that.

The problem of limited scope of words can also be demonstrated by the other example in Figure 1.

My parser mistook the main structure of the sentence: it believed ”reaction” is the root therefore it continued to make further mistakes, which I believe is the case of ”error propagation”. It had to do some greedy ”reduce” step in order to release some space in its stack, especially for sentences which the root word is located in the further back.

A transition based dependency parser has three elements: a stack S , a buffer B for unprocessed words and a list A for arcs it has built. Starting with the root word in S , it scans the sentence from left to right and makes predictions learned from machine learning about whether to build a left arc or a right arc, to pop the top word in the stack (reduce) or to push a word from the buffer to the stack (move). The termination point is when the parser empties its buffer and at that time A is the complete list of parsed arcs.

We shall look at this sentence as an example to show how the parser works.

$$[\text{ROOT}]_s[\text{He worked for the BBC for a decade .}]_B \quad (1)$$

We start with root being our W_0 .

$$[\text{ROOT He}]_s[\text{worked for the BBC for a decade .}]_B \quad (2)$$

Shift ”He” into S

$$[\text{ROOT}]_s[\text{worked for the BBC for a decade .}]_B \quad (3)$$

Build a left arc from ”worked” to ”He”, label it `nsubj`.

$$[\text{ROOT worked}]_s[\text{for the BBC for a decade .}]_B \quad (4)$$

Build a right arc from root to ”worked”, label it `ROOT`.

$$[\text{ROOT worked for the}]_s[\text{the BBC for a decade .}]_B \quad (5)$$

Shift ”for” and the.

$$[\text{ROOT worked for}]_s[\text{BBC for a decade .}]_B \quad (6)$$

Build a left arc from ”BBC” to ”the”, label it `det`.

$$[\text{ROOT worked}]_s[\text{BBC for a decade .}]_B \quad (7)$$

Build a left arc from ”BBC” to ”for”, label it `case`.

$$[\text{ROOT worked BBC}]_s[\text{for a decade .}]_B \quad (8)$$

Build a right arc from "worked" to "BBC", label it *obl*.

$$[\text{ROOT worked BBC for a}]_s[\text{decade .}]_B \quad (9)$$

Shift "for" and "a".

$$[\text{ROOT worked BBC for}]_s[\text{decade .}]_B \quad (10)$$

Build a left arc from "decade" to "a", label it *det*.

$$[\text{ROOT worked BBC}]_s[\text{decade .}]_B \quad (11)$$

Build a left arc from "decade" to "for", label it *case*.

$$[\text{ROOT worked}]_s[\text{decade .}]_B \quad (12)$$

Reduce "BBC".

$$[\text{ROOT worked decade}]_s[.]_B \quad (13)$$

Build a right arc from "worked" to "decade", label it *obl*.

$$[\text{ROOT worked}]_s[.]_B \quad (14)$$

Reduce "decade".

$$[\text{ROOT worked .}]_s[]_B \quad (15)$$

Build a right arc from "worked" to ".", label it *punc*. The empty buffer B marks our end of parsing.

References

- [1] *nmod*. [Online; accessed 5. Dec. 2018]. 2017. URL: <http://universaldependencies.org/u/dep/nmod.html>.
- [2] *Works council* - *Wikipedia*. [Online; accessed 5. Dec. 2018]. 2018. URL: https://en.wikipedia.org/wiki/Works_council.
- [3] *ccomp*. [Online; accessed 5. Dec. 2018]. 2017. URL: <http://universaldependencies.org/u/dep/ccomp.html>.
- [4] *mark*. [Online; accessed 5. Dec. 2018]. 2017. URL: <http://universaldependencies.org/u/dep/mark.html>.
- [5] *acl*. [Online; accessed 5. Dec. 2018]. 2017. URL: <http://universaldependencies.org/u/dep/acl.html>.
- [6] *obl*. [Online; accessed 5. Dec. 2018]. 2017. URL: <http://universaldependencies.org/u/dep/obl.html>.
- [7] *iobj*. [Online; accessed 5. Dec. 2018]. 2017. URL: <http://universaldependencies.org/u/dep/iobj.html>.