Ольга Рудина, БКЛ-152. АОТ, дз №5. Базовые алгоритмы зависимостного парсинга.

Моя сестра положит новую книгу о цветах на стол.

На вход мы получаем обычное синтаксическое дерево (КС). Переход к НС-грамматикам совершается посредством поднятия всех слов-вершин на уровень корня.

1. Примеры ограничений.

1.1. Полное прилагательное (новую).

word(pos(x)=Adj) => (word(mod(x)) = Noun, label(x) = NMOD, pos(x) < mod(x))
Прилагательное модифицирует существительное справа и становится NMOD — pos слева от mod.

1.2. Местоимение (моя).

word(pos(x)=Pr), word(mod(x) = Noun, pos(x) < mod(x)) => label(x) = NMOD выполняет те же функции, что и прилагательное выше, также модифицирует существительное справа от себя; однако мы можем определить, что у местоимения функция именно модификатора имени, только по тому, с какой стороны оно стоит от существительного (а значит, нужно различать типы местоимений)

1.3. Существительное (сестра, книгу, цветах, стол) word(pos(x)=Noun) => word(label(x) = Subj, mod(x) = Verb, pos(x) < mod(x)) Существительное в роли подлежащего модифицирует глагол справа

word(pos(x)=Noun) => (word(mod(x)) ∈ {Verp, Prep}, label(X) = Obj, mod(x) < pos(x)) Существительное в роли дополнения модифицирует глагол и предлог слева. В презентации не менялись знаки, но элементы (pos/mod) местами менялись. Я придерживаюсь такой же логики, потому что так нагляднее. Что правильно, я не знаю.

1.4. Предложная группа (о, на)

word(pos(x)=Prep, word(mod(x)) = Verb, mod(x) < pos(x)) => label(x)=loc Здесь логично поставить другое условие: по названию части речи не понятно, в какой функции она употребляется. Однако если мы говорим, что у нас есть предлог, модифицирующий глагол, то очевидно, что он модифицирует глагол слева от себя, при этом называясь локативом.

word(pos(x)=Prep, word(mod(x)) = Noun, mod(x) < pos(x)) => label(x)=pmod В каком-то смысле близко к предыдущему: если предложная группа модифицирует существительное (а не наоборот), то является предложным модификатором, находясь справа от имени.

2. Nivre

Алгоритм начинается с того, что у нас есть два типа данных – стек и очередь. Стек условно слева, очередь – справа.

В очереди последовательно хранятся слова предложения, стек условно пустой (считаем, что там сразу хранится неосязаемый root)

[root] [моя, сестра, положит, новую, книгу, о, цветах, на, стол]

Сначала движемся по т.н. левым стрелочкам – из очереди в стек, т.е. влево, передвигаем элементы.

'моя'

[root, моя] [сестра, положит, новую, книгу, о, цветах, на, стол]

Ольга Рудина, БКЛ-152. АОТ, дз №5. Базовые алгоритмы зависимостного парсинга.

Добавляем 'cecmpy' в стек и сравниваем с «левым соседом» - кто из них кому вершина? Вершина – 'cecmpa', поэтому «моя» из стека удаляем. Никуда более не добавляем, просто удаляем и всё. И, видимо, оставляем ссылку на то, что слева от имени был какой-то член. Наглядно удалённые элементы я отображу в стеке светлым шрифтом, чтобы показать ссылки на него (стрелки не могут пересекаться)

```
[root, моя, сестра] [положит, новую, книгу, о, цветах, на, стол] [root, моя, сестра, положит][новую, книгу, о, цветах, на, стол] [root, моя, сестра, положит, новую][ книгу, о, цветах, на, стол] [root, моя, сестра, положит, новую, книгу][о, цветах, на, стол] [root, моя, сестра, положит, новую, книгу, о][цветах, на, стол] [root, моя, сестра, положит, новую, книгу, о, цветах][на, стол] [root, моя, сестра, положит, новую, книгу, о, цветах, на][стол] [root, моя, сестра, положит, новую, книгу, о, цветах, на, стол][] Получились следующие вершинные элементы:
```

Теперь наша задача — определиться с правыми стрелочками. Берем выделенные корневые элементы и так же линейно, только справа налево, перекладываем элементы в очередь и сравниваем между собой — что выше.

[root, положит, книгу, цветах, стол][]
[root, положит, книгу, цветах][стол]
[root, положит, книгу][цветах, стол]
[root, положит][книгу, цветах, стол]
[root][положит, книгу, цветах, стол]
[][гооt, положит, книгу, цветах, стол]

[root, положит, книгу, цветах, стол]