

грамматика зависимостей

от чего?

Саша Сергиенко, 08.12,
формлаба

Самая база

Грамматика зависимостей - стрелочки, соединяющие отдельные слова

Грамматика составляющих - деревья, соединяющие составляющие

Dependency grammar

Phrase structure grammar

Почему нам это важно?

Казалось бы, какая разница, как представлять языковой материал. Это всего лишь схема, которую мы используем.

Однако теоретические последствия выбора DG или PSG гораздо шире, чем кажется сначала

Сегодняшняя встреча: разбор методологических (=> теоретических?) отличий грамматик зависимостей от привычных нам формализмов

Немного истории

Почему мы используем PSG, а не DG?

Ответ даже не “Хомский”, а “Блумфилд et al.”

Bloomfield 1933, Wells 1947, ...

Американские дескриптивисты любили PSG

А кто делал DG?

Панини - его караки можно сравнить с различными видами отношений

Люсьен Теньер (Lucien Tesnière) - Основы структурного синтаксиса (Éléments de syntaxe structurale)

1959 г., после смерти автора

David Hays в 1962 г. написал первый парсер

Позднее - И.А. Мельчук, а также Dependency-based systems are increasingly being used to parse natural language and generate tree banks.

Описательный формализм?

‘Dependency formalism is a tool proposed for representing linguistic reality, and, like any tool, it may or may not prove sufficiently useful, flexible or appropriate for the task for which it has been devised; but it cannot be true or false.’ [Mel’čuk 1987:12]

“Universal Dependencies (UD) is a project that is developing cross-linguistically consistent treebank annotation for many languages, with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective. [[website](#)]

Чуть-чуть про UD

Проект по синтаксической аннотации корпусов

Stanford Dependencies, Google universal part-of-speech tags, and the Interset interlingua

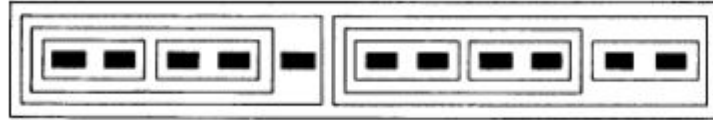
Однослойная теория + сильно привязанная к поверхностной структуре

Отличия от PSG

1. “содержание” vs. отношения
2. группа слов vs. слово
3. категоризация vs. отсутствие таковой
4. возможность терминальных узлов vs. невозможность
5. линейная соположенность узлов vs. отсутствие ограничения
6. отсутствие “ролей” vs. наличие оных
7. функциональные слова управляют vs. подчиняются
8. равноправие предиката и субъекта vs. главенство предиката

constituency vs. relation

1. Множества слов, где одно последовательно вложено в другое
2. Отношения между отдельными словами



37 ВОЗМОЖНЫХ ОТНОШЕНИЙ В UD!

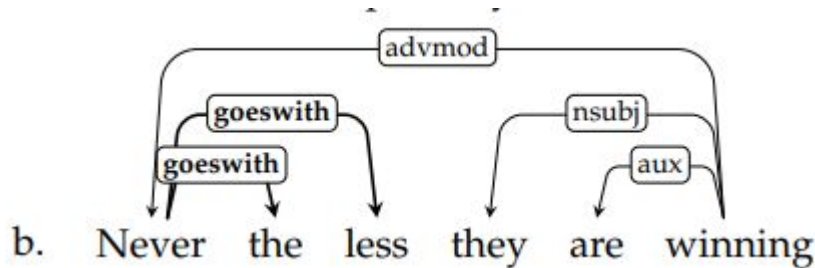
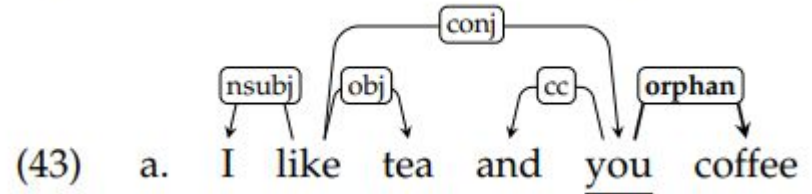
adcl adnominal clause - finite or non-finite clause modifying a nominal; **advcl** -adverbial clause modifying a predicate or modifier word **advmod** adverb or adverbial phrase modifying a predicate or modifier word; **amod** adjectival modifier of a nominal; **appos** appositional modifier; a nominal used to define, name, or describe the referent of a preceding nominal; **aux** auxiliary; links a function word expressing tense, mood, aspect, voice, or evidentiality to a predicate... (de Marneffe 2021: 266)

Приколы

orphan - links orphaned dependents of an elided predicate

goeswith - links parts of a word that are separated but should go together according to standard

orthography or linguistic wordhood



Ещё одно разделение

nominals vs. clauses vs. modifiers

Nominals: the primary means for referring to entities

Clauses: the primary means for referring to events

Modifiers: the canonical attributive modifiers of nominals, clauses, and other modifiers (de Marneffe 2021: 257)

НЕ ЗАБЫВАЕМ что clauses здесь просто “глаголы”, а не “глагольная группа”

Табличка

Table 4

Typology of the syntactic relations.

Head \ Dependent	Nominals	Clauses	Modifier words	Function words
Clausal core arguments	nsubj obj iobj	csubj ccomp xcomp		
Clausal non-core arguments	obl vocative expl dislocated	advcl	advmod discourse	aux cop mark
Nominal dependents	nmod appos nummod	acl	amod	det clf case
Coordination	MWE	Loose	Special	Other
conj cc	fixed flat compound	list parataxis	orphan goeswith reparandum	punct root dep

группа слов vs. слово

The primacy of words can be understood as a commitment to the lexical integrity principle (Chomsky 1970; Bresnan and Mchombo 1995; Aronoff 2007)

Это называется лексикализм! Не декомпозируем меньше слова

Ещё на практике этот принцип нарушается, когда мы присоединяем относительные клаузы например: считается, что мы присоединяем группу предиката

Отсутствие категоризации

В ГГ у нас специфицированы нетерминальные узлы (NP, VP)

the symbols representing the syntactic class membership and other syntactic properties of an item are not admitted as immediate elements of syntactic structure. All such information is relegated to the dictionary entry of the item in question [Mel'čuk 1987:14]

Но при этом у нас используются категоризированные синтаксические отношения: nsubj = **nominal** subject, advmod = **adverbial** modifier

Невозможность терминальных узлов

Каждый узел - это слово.

Исключение - синтаксические нули вроде нулевой копулы в русском

UD: нулей нет



А как же эллипсис? (в UD)

Проблема: мы удаляем вершину, а кто будет приписывать отношения зависимым?

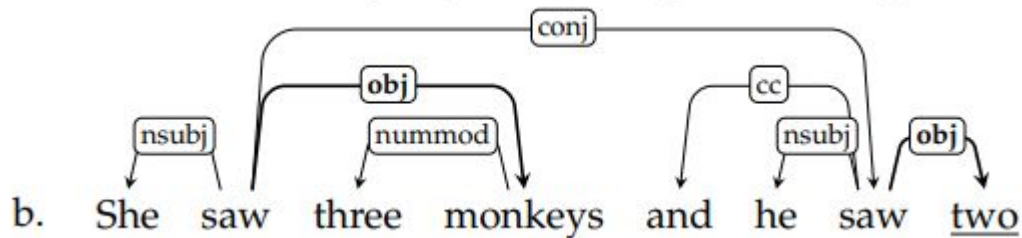
- If the elided element has no overt dependents, nothing is done.
- If the elided element has overt dependents, one of these is promoted to the role of the head.
- If the elided element is a predicate and the promoted element is one of its arguments or phrasal modifiers, the special orphan relation is used when attaching other non-functional dependents to the promoted head.

его идеи - полная фигня

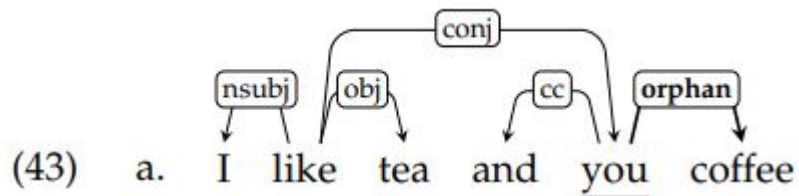
А как же эллипсис? (в UD)

Почему так делать плохо?

Эта штука разрешает предложение типа he saw two



В предложении с orphan они просто сдаются

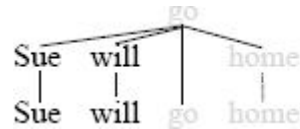


VP-эллипсис: всё совсем плохо

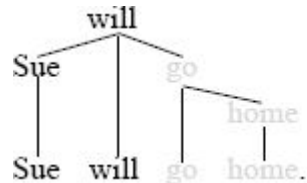
У нас вспомогательные глаголы - зависимые лексических глаголов (will ← do)

Поэтому получается вот так:

Fred won't go home, but Sue will.



– Expected UD analysis



– Purely syntactic analysis

Ну и никаких функциональных проекций

А если очень хочется?

Вводить “глубинную структуру” (см. Смысл \Leftrightarrow Текст). Deep UD - (Droganova, Zeman 2019), Enhanced UD

Deep UD: давайте введём нули для разных видов эллипсиса, PRO, сделаем плоскую структуру конъюнктов

Линеаризация

Отсутствие ограничения на соположенность вершины и зависимого!

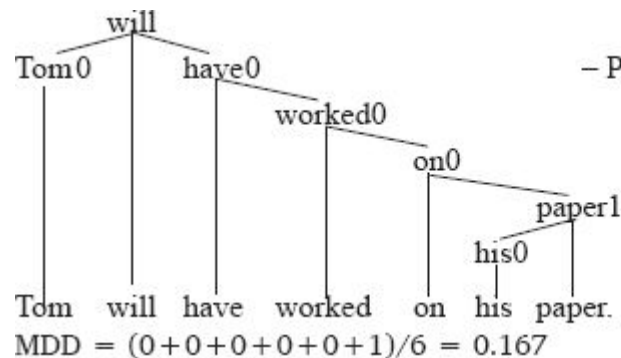
В дереве зависимостей прописаны лишь иерархические отношения, из них не следуют линейные

Нам не нужно думать “вот основной порядок слов, а как из него вышел вот такой?”

Тем не менее: в (Теньер 1959) очень интересная дискуссия линеаризации

Зато можем измерять Dependency distance

Mean dependency distance: сколько слов в среднем линейно отделяет слово от его главы?



– Purely syntactic analysis

Синтаксические роли

ГГ: не субъект, а спецификатор vP (по месту в структуре)

UD: субъект это зависимое от предиката nsubj (по роли nominal subject)

Круто?

Функциональные слова - зависимые?

“фишка” UD: предлоги зависимые имён, will - зависимое глагола,
подчинительные союзы - зависимые глагола,

От этого происходит просто куча проблем

Эллипсис уже был

Фразовые глаголы вроде [laugh at], [focus on] и т.д.

В UD была бы структура focus → it → on!

Предикат - главное слово

Интересный ход, благодаря которому мы можем “отсчитывать” все отношения от предиката

Теньер: во всём виноваты логики (Аристотель, Пор-роялевцы). Это относится к логической организации предложения.

Double dependency?

Прилагательное получает падеж от глагола, а фи-признаки - от существительного. Мы нашли зал **пустым**.

Что грамматики зависимостей дали нам?

Head-directionality parameter. Nichols 1986 (head-final vs. head-initial языки)
переработала термины Теньера “центростремительные” и “центробежные”.

Nichols: ‘but it is less obvious to constituency grammar’

Синтаксический парсинг

F-структуру в LFG!

Проективность vs. непроективность

Рестриктивность UD и похожих?

UD разрешает слишком много строк

Почему так происходит - потому что UD настроен не на синтез, а на анализ.
Нам нужно построить графы для всех предложений из корпуса.

UD занимается даже неграмматичными предложениями!

goeswith - ошибки токенизации, reparandum - оговорки

Для этой теории важно не “как бы нам не разрешить неправильное предложение”, а “как бы нам не ошибиться в парсинге правильного”

В заключение: милое дерево от Теньера

A mes enfants, Michel, Bernard et Yveline Tesnière, je dédie ce livre dont leurs curiosités d'élèves de sixième A ont hâté la maturation.

'To my children, Michel, Bernard and Yveline Tesnière, students at the sixth grade level, I dedicate this book, the development of which has been promoted by their curiosities.'

