

## **Об использовании синтаксического анализатора Cognitive Dwarf 2.0**

А. А. Антонова, А. В. Мисюрёв

Описывается новая версия программы синтаксического анализа Cognitive Dwarf 2.0, предназначенной для автоматической морфологической и синтаксической разметки текста на русском или английском языке. Изменения в новой версии синтаксического анализатора упрощают интерпретацию и обработку результатов разбора.

В частности, упрощается его применение при написании дипломных и курсовых работ студентами технических и гуманитарных специальностей. Изменения в формате выходного файла позволяют импортировать и обрабатывать результаты разбора стандартными средствами (базами данных, электронными таблицами, системой Mathworks MATLAB и т. д.) без использования таких языков программирования, как Си++ или Java.

### **Введение**

В статье описываются особенности использования синтаксического анализатора Cognitive Dwarf 2.0. Эта программа предназначена для автоматической морфологической и синтаксической разметки текстов на русском или английском языке и ориентирована на универсальную обработку текста независимо от жанра и предметной области. Предыдущая версия Cognitive Dwarf описана в [2].

Результатом работы программы является текст, разбитый на слова и предложения, где для слов определяются нормальные формы, грамматические характеристики и синтаксические отношения, в которые они вступают внутри предложения. Используемое синтаксическое описание основано на отношениях зависимости между словами.

Хотя изначально синтаксический анализ разрабатывался для нужд машинного перевода, он зависит только от входного языка и не подстраивается ни под какую языковую пару. Благодаря этому увеличивается область его применения в других приложениях, в том числе не связанных с переводом. В частности, синтаксический анализатор используется при написании курсовых и дипломных работ студентами механико-математического факультета МГУ, ОТиПЛ филологического факультета МГУ.

По сравнению с предыдущей версией программы можно выделить следующие основные отличия.

- Используется другой алгоритм синтаксического разбора и другой набор правил, описывающих синтаксис языка. Есть отличия в наборе выделяемых синтаксических связей и способе построения результирующего дерева разбора.
- Учен опыт использования предыдущей версии программы. Изменения в формате выходного файла позволяют импортировать и обрабатывать результаты разбора стандартными средствами (базами данных, электронными таблицами, системой Mathworks MATLAB и т. д.), без использования таких языков программирования, как Си++ или Java. Это, как предполагается, может сделать программу привлекательной для более широкого круга пользователей.

## 1. Синтаксический разбор в Cognitive Dwarf 2.0

Синтаксический анализатор Cognitive Dwarf осуществляет *частичный синтаксический разбор* предложений, решая практическую задачу построения наилучшей синтаксической интерпретации для *произвольной* входной цепочки слов. Различают два способа представления предложений в виде иерархической структуры грамматических конструкций: дерево непосредственных составляющих и дерево зависимостей.

В дереве непосредственных составляющих составные лингвистические конструкции образуются из набора более простых линейно непересекающихся отрезков, которые называются непосредственными составляющими этой конструкции. При этом в дереве различают терминальные узлы, соответствующие отдельным словам, и нетерминальные — соответствующие конструкциям. Существуют общепринятые наименования нетерминальных узлов (например, S — предложение, NP — именная группа, VP — глагольная группа).

Дерево зависимостей строится из бинарных отношений непосредственного подчинения зависимого слова главному. При этом не вводятся

нетерминальные узлы, а представителем конструкции всегда является ее главное слово. Дерево непосредственных составляющих можно преобразовать в дерево зависимостей, и наоборот.

### 1.1. Дерево синтаксического разбора

В основе лингвистического описания системы Cognitive Dwarf лежит грамматика зависимостей [4–7]. Дерево разбора представляется в виде набора бинарных направленных отношений зависимости между словами. В каждом отношении задаются слово-мама (главное слово) и слово-дочка (зависимое слово).

Для того чтобы результирующий граф являлся деревом синтаксического разбора, соблюдаются следующие требования.

- 1) Древесная структура. Каждое слово может иметь произвольное количество слов-дочек, но только одно слово-маму.
- 2) Целостность дерева. Алгоритм разбора допускает разбиение исходного предложения на два и более независимых поддеревя. Поэтому, чтобы обеспечить целостность результирующего дерева, перед предложением вводится дополнительное пустое слово **\*Тор\***, которое играет роль слова-мамы для слов, являющихся вершинами не связанных между собой поддеревьев.
- 3) Включение в дерево всех слов без исключений. Каждое слово в предложении должно иметь слово-мату.

Требование проективности<sup>1</sup> (как сильной так и слабой), обязательное в некоторых синтаксических теориях, не является таковым в системе Cognitive Dwarf. Однако непроективность в результатах разбора возможна только внутри единого поддеревя, что значит — в рамках распознанной конструкции. Непроективное присоединение независимых поддеревьев к **\*Тор\***у невозможно.

---

<sup>1</sup>Требование проективности налагает следующие условия:

— если два слова **a** и **b**, расположенные в тексте в соответствующем порядке, связаны подчинительным отношением, то слово, находящееся в тексте между ними, не может вступать в отношения ни с каким словом, расположенным в тексте левее, чем **a**, и правее, чем **b** (слабая проективность);

— слово-дочка, расположенное в тексте слева от мамы, не может иметь своих дочек, расположенных правее мамы, а слово-дочка, расположенное в тексте справа от мамы, соответственно, не может иметь своих дочек, расположенных левее мамы (сильная проективность).

## 1.2. Отношения зависимости

В результатах разбора каждому слову приписывается то отношение зависимости, в котором данное слово является словом-дочкой — по определению такое отношение только одно для данного слова. Участие слова в некотором отношении зависимости в качестве слова-дочки определяет его синтаксическую роль в предложении, и, соответственно, синтаксическую роль всего поддерева, возглавляемого данным словом. Отношение зависимости имеет имя, обозначающее синтаксическую роль слова-дочки. Дерево разбора для предложения «*Читать книгу*»:

\*Top\*

inf *читать*  
           acc *книга / книгу*

Особо следует сказать про отношения зависимости, привязывающие вершины независимых поддеревьев к пустому слову \*Top\*. В этом случае имя связи не указывает на синтаксическую роль поддерева (так как синтаксическая роль раскрывается только при реальном подчинении), а отражает принадлежность поддерева к определенному синтаксическому классу<sup>2</sup> (предложение, финитный глагол, краткое прилагательное, именная группа, предложная группа и т. д.).

Список возможных синтаксических ролей находится в приложении (Таблица 1, Таблица 2).

## 2. Особенности анализа конструкций

Синтаксический анализатор Cognitive Dwarf придерживается синтаксической теории, при которой вершиной предложения является предикат, а структура предложения во многом определяется возможным набором актантов данного предиката (субъектом, объектом и пр.). Имеются особенности в разборе конструкций, где предикат управляет другими глагольными формами (например, инфинитивом), в разборе предложных групп, а также в разборе конструкций с однородными членами.

### 2.1. Предложные группы

В версии Cognitive Dwarf 2.0. в предложной группе предлог считается дочкой имени, несмотря на то, что с синтаксической точки зрения

---

<sup>2</sup> В предыдущей версии Cognitive Dwarf каждое имя отношения состояло из трех частей: идентификатор синтаксической роли слова-дочки, стрелка, указывающая, слева или справа находится слово-мама, и идентификатор синтаксического класса слова-мамы. \*Впоследствии этот идентификатор был сочтен избыточным и удален из результатов разбора.

предлог управляет именем. Дерево разбора для предложения «*Смотреть в телескоп*»:

**\*Топ\***

```

inf  смотреть
     ргерпр телескоп
           ргер  в
  
```

Данный прием не влияет на результативность разбора, так как имя после присоединения предлога выступает в качестве предложной группы, а для дальнейшей работы с деревом (при сборе статистики, при оценке дерева, и т. п.) часто представляется более удобным, чтобы слова были связаны напрямую, а не через предлог.

## 2.2. Однородные члены

Одно из отличий версии Cognitive Dwarf 2.0. от предыдущего Cognitive Dwarf состоит в том, что в дереве больше нет специального промежуточного узла, к которому присоединялись однородные члены. Вместо этого каждый последующий однородный член присоединяется как дочка к предыдущему с синтаксической ролью **homo**. Запятая или сочинительный союз, соединяющие однородные члены, присоединяются к тому члену, который расположен левее, с синтаксической ролью **conj**. Дерево разбора для предложения «*Пришел, сел и посмотрел в телескоп*»:

**\*Топ\***

```

fin  прийти / пришел
     conj  ,
     homo  сесть / сел
           conj  и
           homo  посмотреть / посмотрел
                 ргерпр телескоп
                       ргер  в
  
```

## 2.3. Служебные глаголы

В версии Cognitive Dwarf 2.0. большим изменениям подвергся анализ конструкций со служебными глаголами. Служебными теперь считаются почти все употребления, где одна глагольная форма управляет другой глагольной формой, например, «*будет делать*», «*хочу пойти*», «*мне нравится плавать*», «*приказав ему явиться*». Словом-мамой в таких конструкциях объявляется смысловой глагол, хотя с синтаксической точки зрения служебный глагол управляет смысловым.

Различается два типа служебных глаголов: глаголы, подлежащее которых совпадает с подлежащим смыслового глагола (они получают синтаксическую роль **auxs**), и глаголы, подлежащее которых отличается от подлежащего смыслового глагола (они получают синтаксическую роль **auxd**). Дерево разбора для предложения «Я хочу пойти»:

\*Топ\*

**sent** *пойти*  
**subj** *я*  
**auxs** *хотеть / хочу*

Дерево разбора для предложения «Мне нравится плавать»:

\*Топ\*

**sent** *плавать*  
**subj** *я / мне*  
**auxd** *нравиться / нравится*

В языке встречаются конструкции с более чем одним служебным глаголом, например, «хочу начать рассказывать», «хотел приказать ему явиться», «мне будет хотеться пойти». Дерево разбора для предложения «Хотел приказать ему явиться»:

\*Топ\*

**sent** *явиться*  
**auxd** *приказать*  
**auxs** *хотеть / хотел*  
**subj** *он / ему*

## 2.4. Вложенные предложения

Синтаксический анализатор Cognitive Dwarf 2.0. умеет распознавать часть конструкций с вложенными предложениями (например, изъяснительными и относительными). Вложенное предложение может подчиняться глаголу или именной группе, в таком случае главное слово вложенного предложения получит синтаксическую роль **sent**. Дерево разбора для предложения «Видел девочку, которая ему нравится»:

\*Топ\*

**fin** *видеть / видел*  
**acc** *девочка / девочку*  
**sent** *нравиться / нравится*  
**subj** *который / которая*  
**misc** *,*  
**dat** *он / ему*

Однако, следует заметить, что, в связи с большой комбинаторикой при разборе конструкций с вложенными предложениями, надежность их выделения пока оставляет желать лучшего.

### 3. Результаты синтаксического разбора

Результатом работы программы является текстовый файл (таблица с фиксированным числом столбцов), в котором исходный текст разбит на слова и предложения, для слов определены нормальные формы, грамматические характеристики и отношения зависимости, в которые они вступают внутри предложения.

Пример результатов синтаксического разбора для предложения:

*Тося договорился с Хубертом выкупить свое кольцо, и деньги для этого должен буду дать я.*

1 *Тор*	*Тор*	1	_	/
2 тося	тося	6	subj	/nn/sg/msc/anm/nom/trd/heu/
3 договорился	договориться	6	auxs	/vrb/sg/msc/fin/fst/sec/trd/pst/ind/act/
4 с	с	5	prep	/prp/ins/
5 хубертом	хуберт	6	prepnp	/nn/sg/msc/anm/ins/trd/heu/
6 выкупить	выкупить	1	sent	/vrb/inf/act/
7 свое	свой	8	adj	/prn/sg/neu/acc/trd/
8 кольцо	кольцо	6	acc	/nn/sg/neu/acc/trd/
9 ,	,	10	misc	/pnt/
10 и	и	16	conj	/cnj/
11 деньги	деньги	16	acc	/nn/pl/fem/acc/trd/
12 для	для	13	prep	/prp/gen/
13 этого	этот	16	prepnp	/prn/sg/msc/gen/trd/
14 должен	должен	16	auxs	/adj/sg/msc/fst/sec/trd/sht/
15 буду	быть	14	auxs	/vrb/sg/fem/msc/neu/fin/fst/prs/ind/act/
16 дать	дать	1	sent	/vrb/inf/act/
17 я	я	16	subj	/prn/sg/fem/msc/neu/nom/fst/
18 .	.	1	misc	/pnt/

Значения столбцов:

- 1 — номер текущего слова;
- 2 — текст текущего слова;
- 3 — нормальная форма текущего слова;
- 4 — номер слова-мамы (каждое слово по определению имеет одно слово-маму);
- 5 — тэг отношения зависимости (этим тегом текущее слово связано со своим словом-мамой);
- 6 — грамматические атрибуты текущего слова (порядок следования атрибутов фиксирован).

Примечания.

1. Слова в результирующем файле расположены в том порядке, как они идут в тексте.
2. Первым словом в каждом предложении является специальное пустое слово **\*Тор\***. Оно выполняет роль слова-мамы для тех слов, которые являются вершинами своих поддеревьев. Если исходный текст состоит из нескольких предложений, в результирующем файле начало каждого нового предложения можно определить по слову **\*Тор\***.
3. Расшифровку грамматических атрибутов для русского и английского языков, а также тэгов отношений зависимости см. Приложение.

Дерево синтаксического разбора однозначно определяется совокупностью всех отношений зависимости, установленных для данного предложения:

**\*Тор\***

```

sent   выкупить
      subj   тося
      auxs   договориться / договорился
      ргепрп3 хуберт / хубертом
      ргеп   с
      acc    кольцо
      adj    свой / свое
sent   дать
      conj   и
      misc   ,
      acc    деньги
      ргепрп   этот / этого
      ргеп    для
      auxs   должен
      auxs   быть / буду
      subj   я
misc   .

```

#### 4. Обработка результатов разбора

Текстовый файл с результатами разбора может быть импортирован электронными таблицами и базами данных. Используя встроенные сред-

<sup>3</sup> Имеет место неверное присоединение предложной группы («выкупить с Хубертом» вместо «договориться с Хубертом»).

ства (конструктор запросов, язык SQL-запросов), можно строить частотные словари, выделять подмножества конструкций, удовлетворяющих заданным условиям, извлекать лексико-грамматические параметры текстов. Например, в [2] приведены примеры сравнения произведений различных авторов при помощи признаков, вычисляемых на основе морфологических атрибутов слов<sup>4</sup>, а также на основе распределения синтаксических связей. И та и другая информация может быть извлечена стандартными средствами из результатов разбора.

Далее приводятся примеры обработки результатов разбора в Microsoft Access. Приводимые названия пунктов меню соответствуют Microsoft Access 2003.

Текстовый файл с результатами разбора можно импортировать в таблицу (разделы меню Файл → Внешние данные → Импорт, результат показан на рис. 1).

Выбрав в столбце «Объекты» окна базы данных кнопку «Запросы», далее кнопку «Создать» на панели инструментов этого окна, далее «Повторяющиеся записи» и указав поле **link** в качестве поля, содержащего

word_id	form	norm	parent_id	link	attr
1	*Тор*	*Тор*	1	_	/
2	тося	тося	6	subj	/nn/sg/msc/anm/nom/trd/heu/
3	договорился	договориться	6	auxs	/vrb/sg/msc/fin/fst/sec/trd/pst/ind/act/
4	с	с	5	prep	/prp/ins/
5	хубертом	хуберт	6	prepnp	/nn/sg/msc/anm/ins/trd/heu/
6	выкупить	выкупить	1	sent	/vrb/inf/act/
7	свое	свой	8	adj	/prn/sg/neu/acc/trd/
8	кольцо	кольцо	6	acc	/nn/sg/neu/acc/trd/
9	,	,	10	misc	/pnt/
10	и	и	16	conj	/cnj/
11	деньги	деньги	16	acc	/nn/pl/fem/acc/trd/
12	для	для	13	prep	/prp/gen/
13	этого	этот	16	prepnp	/prn/sg/msc/gen/trd/
14	должен	должен	16	auxs	/adj/sg/msc/fst/sec/trd/sht/
15	буду	быть	14	auxs	/vrb/sg/fem/msc/neu/fin/fst/prs/ind/act/
16	дать	дать	1	sent	/vrb/inf/act/
17	я	я	16	subj	/prn/sg/fem/msc/neu/nom/fst/
18	.	.	1	misc	/pnt/

Запись: 1 из 18

Рис. 1. Результаты разбора в окне базы данных

<sup>4</sup> Признаки были заимствованы с изменениями из [Галышина Е. И. Основы судебного речеведения. М.: СТЭНСИ, 2003].

link поле	Повторы
acc	2
auxs	3
misc	2
prep	2
prepnp	2
sent	2
subj	2

Запись: 7 из 7

**Рис. 2.** Результаты запроса, позволяющего получать частоты связей

повторяющиеся значения, можно получить частоты для связей, встретившихся в тексте больше одного раза (рис. 2).

Выбрав в столбце «Объекты» окна базы данных кнопку «Запросы», далее кнопку «Конструктор» на панели инструментов этого окна и далее в меню «Вид» раздел «Режим SQL», можно вводить текст SQL-запроса непосредственно. Например, если имя таблицы с результатами разбора — **result**, то запрос

```
SELECT COUNT(1)
FROM result
WHERE attr LIKE "/nn*/acc*";
```

позволяет получить количество существительных в тексте в винительном падеже. Запрос

```
SELECT a.word_id, a.norm, c.norm, b.norm
FROM result AS a, result AS b, result AS c
WHERE (a.word_id=b.parent_id) AND
(b.word_id=c.parent_id) AND
(c.attr Like "/prp/*")
ORDER BY a.word_id;
```

позволяет получить все слова в тексте, связанные при участии предлога (рис. 3)

Этот запрос может быть разбит на три.

1. Выделить все предлоги в тексте.

```
SELECT word_id, form, parent_id
INTO a
FROM result
WHERE attr LIKE "/prp*";
```

2. Выделить все слова, которые могут быть дочками в связях с предлогом.

```
SELECT result.word_id, result.form,  
result.parent_id, a.form AS prp  
INTO b  
FROM result, a  
WHERE result.word_id=a.parent_id;
```

3. Построить таблицу, содержащую слова, связанные при участии предлога.

```
SELECT result.word_id, result.form, b.prp, b.form as obj  
INTO c  
FROM result, b  
WHERE result.word_id=b.parent_id;
```

## Заключение

Изменения в описанной версии синтаксического анализатора упрощают интерпретацию и обработку результатов разбора и делают возможным его применение для более широкого круга задач. Представляется, что материалы, полученные при помощи данного синтаксического анализатора, могут использоваться при написании дипломных и курсовых работ студентами технических и гуманитарных специальностей. Темы исследований могут быть связаны с изучением и сравнением лексико-грамматических параметров текстов (поиском авторского инварианта, изучением жанровых и тематических различий, сравнением параллельных текстов на русском и английском языке).

Возможны также применения, связанные не только с анализом отдельных текстов, но и с исследованиями языка в целом. Синтаксический анализатор использовался для сбора статистических данных для русского и английского языков. Объем разобранный текст составлял приблизительно 1.5 Гбайт (около 200 млн словоупотреблений). Такого рода данные



word_id	a.norm	c.norm	b.norm
6	выкупить	с	хуберт
16	дать	для	этот

Рис. 3. Результаты запроса, выбирающего конструкции с предложной группой

могут быть полезны при построении частотных словарей или, например, статистической модели языка.

Дополнительные сведения о синтаксическом анализаторе можно найти в сети Интернет по адресу <http://cs.isa.ru:10000/dwarf>.

## Литература

1. Антонова А. А., Мисюрев А. В. Реализация синтаксического разбора для русского и английского языков // Первая Международная конференция «Системный анализ и информационные технологии» 12–16 сентября 2005 г.: сб. ст. Переславль-Залесский, 2005. С. 245–249.
2. Антонова А. А., Мисюрев А. В. Синтаксический анализатор для русского и английского языков // Сб. трудов ИСА РАН. Т. 29. Информационно-аналитические аспекты в задачах управления. Под ред. В. Л. Арлазарова и Н. Е. Емельянова. М.: Издательство ЛКИ/URSS, 2007.
3. Гершензон Л. М., Ножов И. М., Панкратов Д. В., Сокирко А. В. Синтаксический анализ в системе РМЛ. [Электронный ресурс] <http://www.aot.ru/docs/synan.html>
4. Ермаков А. Е. Неполный синтаксический анализ текста в информационно-поисковых системах // Компьютерная лингвистика и интеллектуальные технологии: труды Международного семинара Диалог'2002: В 2 т. Т. 2. Прикладные проблемы. М.: Наука, 2002.
5. David Hays G. 1964. Dependency theory: A formalism and some observations. *Language*, 40: P. 511–525.
6. Hudson Richard *English Word Grammar*. Basil Blackwell, Cambridge, MA, 1991.
7. Mel'čuk Igor A. *Dependency Syntax: Theory and Practice*. State University of New York Press, Albany, 1987.

## Приложение

### Грамматические атрибуты для русского языка

*часть речи*

- nn** — существительное
- vrб** — глагол
- adj** — прилагательное
- adv** — наречие
- num** — числительное
- prn** — местоимение
- ptp** — причастие
- dee** — деепричастие
- inj** — междометие
- pt** — частица
- cnj** — союз
- prp** — предлог

**pth** — вводное слово  
**inv** — инвариант (например, *кз*)  
**pdv** — предикатив (например, *можно, жаль*)  
**frn** — слово другого языка (написанное буквами другого языка)  
**pnt** — знак препинания  
**dgt** — цифры

*число*

**sg** — единственное  
**pl** — множественное

*род*

**fem** — женский  
**msc** — мужской  
**neu** — средний

*одушевленность*

**anm** — одушевленное

*падеж*

**nom** — именительный  
**gen** — родительный  
**dat** — дательный  
**acc** — винительный  
**ins** — творительный  
**loc** — предложный

*финитность*

**fin** — финитная форма  
**inf** — инфинитив

*лицо*

**fst** — первое  
**sec** — второе  
**trd** — третье

*время*

**pst** — прошедшее  
**prs** — настоящее (у глаголов совершенного вида соответствует будущему времени)

*наклонение***imp** — повелительное**ind** — изъявительное*залог***act** — активный**psv** — пассивный*краткая форма***sht** — краткое (у прилагательных и причастий, например, *красив, пойманы*)*степень***cmp** — сравнительная степень (у прилагательных и наречий)*эвристическая нормализация***heu** — неизвестное слово — нормализовано эвристически**Грамматические атрибуты для английского языка***часть речи [part of speech]***nn** — существительное [noun]**vrб** — глагол [verb]**adj** — прилагательное [adjective]**adv** — наречие [adverb]**num** — числительное [numeral]**prn** — местоимение [pronoun]**inj** — междометие [interjection]**pt** — частица [particle]**cnj** — союз [conjunction]**prp** — предлог [preposition]**pth** — вводное слово [parenthetic]**det** — [determiner]**inv** — инвариант [invariant]**dtn** — [determiner/pronoun]**frn** — слово другого языка [foreign]**ctr** — сокращенная форма подлежащего и служебного глагола (*he'd, we've*)**dgt** — цифры [digit]**pnt** — знак препинания [punctuation]*число [number]***sg** — единственное [singular]**pl** — множественное [plural]

*притяжательный падеж [possessive case]*

**gen** — слово в притяжательном падеже (например, father's)

*финитность*

**fin** — финитная форма [finite]

**inf** — инфинитив [infinitive]

**ptp** — причастие [participle]

*время [tense]*

**pst** — прошедшее [past]

**prs** — настоящее [present]

*лицо [person]*

**fst** — первое [first]

**sec** — второе [second]

**trd** — третье [third]

*степень [degree]*

**cmp** — сравнительная [comparative]

**sup** — превосходная [superlative]

*эвристическая нормализация*

**heu** — неизвестное слово — нормализовано эвристически

**Таблица 1**

Расшифровка тэгов отношений зависимости для разбора русского языка

Тэг отношения зависимости	Синтаксическая роль слова-дочки
<b>subj</b>	подлежащее (зависит от сказуемого)
<b>acc</b>	прямое дополнение
<b>dat</b>	дополнение в дательном падеже
<b>ins</b>	дополнение в творительном падеже
<b>gen</b>	генитивное дополнение (зависит от имени) или дополнение в родительном падеже (зависит от глагола)
<b>prenp</b>	предложная группа
<b>adj</b>	простое определение (зависит от имени), не имеющее собственных дополнений

Продолжение таблицы 1

Тэг отношения зависимости	Синтаксическая роль слова-дочки
<b>ptp</b>	сложное определение (причастный оборот)
<b>adv</b>	наречие
<b>prep</b>	предлог (зависит от имени), превращающий именную группу в предложную
<b>conj</b>	союз
<b>digit</b>	число из цифр
<b>card</b>	количественное числительное (зависит от имени или от другого числительного)
<b>auxs</b>	служебный глагол (зависит от другого глагола), подлежащее которого совпадает с подлежащим глагола-мамы
<b>auxd</b>	служебный глагол (зависит от другого глагола), подлежащее которого отличается от подлежащего глагола-мамы
<b>inf</b>	инфинитив (зависит от глагола)
<b>sent</b>	придаточное предложение (изъяснительное или относительное)
<b>by</b>	сослагательная частица «бы»
<b>li</b>	частица «ли»
<b>pt</b>	частица
<b>emph</b>	эмфатический союз «и»
<b>homo</b>	однородный член (зависит от предыдущего однородного члена)
<b>sharp</b>	часть составного слова (зависит от предыдущей части составного слова)
<b>hyph</b>	часть слова с дефисом (зависит от другой части, считающейся главной)
<b>quoml</b>	левая кавычка (зависит от главного слова в кавычках)
<b>quomr</b>	правая кавычка (зависит от главного слова в кавычках)
<b>misc</b>	разное (то, что не входит в вышеприведенные обозначения)

Окончание таблицы 1

Тэг отношения зависимости	Синтаксическая роль слова-дочки
<b>Тэги для независимых поддеревьев, присоединяющихся к пустому слову *Top*</b>	
<b>sent</b>	предложение (чаще всего, требуется наличие подлежащего и сказуемого)
<b>np</b>	именная группа
<b>prenp</b>	предложная группа
<b>prep</b>	предлог
<b>adj</b>	группа прилагательного или причастия (в полной форме)
<b>adv</b>	группа наречия
<b>fin</b>	группа глагола в финитной форме
<b>krat</b>	группа краткого прилагательного или причастия
<b>inf</b>	группа инфинитива
<b>dee</b>	группа деепричастия
<b>imper</b>	группа глагола в повелительном наклонении
<b>misc</b>	разное (то, что не входит в вышеприведенные обозначения)

Таблица 2

Расшифровка тэгов отношений зависимости для разбора английского языка

Тэг отношения зависимости	Синтаксическая роль слова-дочки
<b>subj</b>	подлежащее (зависит от сказуемого)
<b>sm</b>	первое именное дополнение глагола
<b>sm2</b>	второе именное дополнение глагола
<b>prenp</b>	предложная группа
<b>det</b>	артикль или другой определитель (determiner) при имени
<b>gen</b>	генитивная группа (зависит от имени)

Продолжение таблицы 2

Тэг отношения зависимости	Синтаксическая роль слова-дочки
<b>poss</b>	поссессивная группа (зависит от имени)
<b>adj</b>	определение, стоящее перед именем, прилагательное/причастие, зависящее от глагола
<b>ptp</b>	причастный оборот, стоящий после имени
<b>adv</b>	наречие
<b>prep</b>	предлог (зависит от имени), превращающий именную группу в предложную
<b>conj</b>	союз
<b>digit</b>	число из цифр
<b>card</b>	количественное числительное (зависит от имени или от другого числительного)
<b>auxs</b>	служебный глагол (зависит от другого глагола), подлежащее которого совпадает с подлежащим глагола-мамы
<b>auxd</b>	служебный глагол (зависит от другого глагола), подлежащее которого отличается от подлежащего глагола-мамы
<b>inf</b>	инфинитив с «to»
<b>bare</b>	инфинитив без «to»
<b>ing</b>	ing-форма (зависит от глагола)
<b>sent</b>	придаточное предложение (изъяснительное или относительное)
<b>to</b>	частица «to» при инфинитиве
<b>not</b>	частица «not», стоящая после глагола
<b>pt</b>	частица
<b>homo</b>	однородный член (зависит от предыдущего однородного члена)
<b>sharp</b>	часть составного слова (зависит от предыдущей части составного слова)

Окончание таблицы 2

Тэг отношения зависимости	Синтаксическая роль слова-дочки
<b>hyph</b>	часть слова с дефисом (зависит от другой части, считающейся главной)
<b>quoml</b>	левая кавычка (зависит от главного слова в кавычках)
<b>quomr</b>	правая кавычка (зависит от главного слова в кавычках)
<b>misc</b>	разное (то, что не входит в вышеприведенные обозначения)
<b>Тэги для независимых поддеревьев, присоединяющихся к пустому слову *Top*</b>	
<b>sent</b>	предложение (чаще всего, требуется наличие подлежащего и сказуемого)
<b>np</b>	именная группа
<b>prenp</b>	предложная группа
<b>prep</b>	предлог
<b>adj</b>	группа прилагательного или причастия прошедшего времени
<b>adv</b>	группа наречия
<b>fin</b>	группа глагола в финитной форме
<b>inf</b>	группа инфинитива с to
<b>bare</b>	группа инфинитива без to
<b>ing</b>	группа ing-формы глагола
<b>misc</b>	разное (то, что не входит в вышеприведенные обозначения)